Identifying Responders
to
Melphalan and Dexamethasone
for
Newly Diagnosed Multiple Myeloma Patients

by
Abbas Esmaeili

A thesis submitted to the Department of Community Health and Epidemiology in conformity with requirements for the degree of
Master of Science

Queen’s University
Kingston, Ontario, Canada
July 2008

Copyright © Abbas Esmaeili, 2008
Abstract

**Background:** MY7 clinical trial compared dexamethasone plus melphalan (MD) vs. prednisone plus melphalan (MP) in multiple myeloma treatment and found no statistically significant difference in overall survival (OS) between the two groups. But, patients reacted to treatment differently. We aimed to identify patients who might have benefited from dexamethasone, and characterize them by their baseline demographic and clinical factors.

**Methods:** First, the prognostic model for OS was developed on the MP arm. The estimated coefficients and baseline hazard were applied to the MD arm to derive martingale residuals (MR). Classification and regression tree analysis was done to identify independent predictive factors for OS and MR was used as response variable. All covariates in categorical shape were used as independent variables to develop the predictive model in MD arm. MP arm was divided accordingly. Subgroups with negative mean MR (survived > expected) were candidates for positive responders while those with positive mean MR (survived < expected) were candidates of negative responders. Mean MR in each subgroup and p values from comparison of OS (log rank test stratified by subgroups) were used to combine the appropriate subgroups as the positive responders or negative responders.

**Results:** A total of 97 patients (42%) in MD arm were identified as positive responders and their OS (median of 44.5 months) was significantly longer than that (median of 33 months) in the corresponding subgroups in MP arm (HR =
0.56, 95% CI 0.4-0.8; p = 0.0014). All positive responders had three common baseline characteristics: aged ≤75 years, calcium concentration ≤2.6 mmol/L and Durie-Salmon stages 2 or 3. Among patients with ECOG performance status <2 those with either HGB ≥100 mg/dl or HGB <100 mg/dl and WBC ≥4,000 and <4 lytic bone lesions were categorized as positive responders. Also, among the patients with ECOG performance status ≥2, males with >3 lytic bone lesions were positive responders. Negative responders (HR = 1.56, 95% confidence interval 1.1 – 2.2; p = 0.006) included patients aged >75 or aged ≤75 with calcium concentration >2.6 mmol/L or aged ≤75 with calcium concentration ≤2.6 mmol/L but had Durie-Salmon stage 1.

**Conclusions:** Evaluation of the hypotheses validity warrants further studies.
Acknowledgements

This thesis arose in part after years of research that was done before I came to Queen’s University. By this time, I have worked with different people whose direct or indirect contribution to the research deserves special mention. It is my pleasure to convey my deep gratitude to them all from the bottom of my heart.

I would like to record my gratitude to Keyue Ding for his supervision, advice, and guidance from the very early stage of this research as well as giving me extraordinary experiences throughout the work. Above all and the most needed, he provided me with persistent encouragement and support in various ways. I gratefully acknowledge Ralph Meyer for his advice, supervision, and crucial contribution, which made him a backbone of this research and so to this thesis. His involvement with his originality has triggered and nourished my academic maturity that I will benefit from, for a long time to come.

To the role model for hard workers in the department, Kristan Aronson, I would like to thank for being the first person who taught me how to go forward step by step during my two-year master graduate study here at Queen’s University. I am proud to record that I had several opportunities to talk with an exceptionally experienced scientist like her. Her office door was always open to students and she never turned me down for taking her precious time to answer wonderfully my endless questions. Kristan, I am grateful in every possible way and hope to keep up our collaboration in the future.

Many thanks go in particular to Linda Levesque and Paul Peng because of their precious time to read the outline of proposal and giving me the constructive
comments on this thesis. Thanks to Linda for her willingness to share her bright thoughts with me, which were very fruitful for shaping up my ideas and research. She was the internal reviewer of the current thesis and her critical comments helped me beautifully shape the final draft of my thesis. I have also benefited by advice and guidance from Ray Viola who was the external reviewer of my thesis and kindly granted his time for carefully reading the pre-final draft of the thesis and asked me some important questions about the project during the thesis defense.

I am deeply grateful to all professors who patiently taught me the science of biostatistics and clinical epidemiology. Of those, my special thanks to Yuk-Miu Lam and William Mackillop who provided the real concepts of biostatistics and clinical epidemiology, respectively. It is a pleasure to pay tribute also to other colleagues: Tina Dyer, Suzanna Keller and Brenda Bass. I am thankful that in the midst of all their activities, they accepted to attend in my practice presentations and offered their constructive comments on this thesis.

It is a pleasure to express my gratitude to Andrew Day and Jina Zhang-Salomons for their great teaching in SAS lab and answering my questions. Thanks to Weidong Kong for letting me use his great knowledge and experience in SAS. He was always ready to lend me a hand with a smiling face. Many thanks to Lee Watkins, administrative assistant of Community Health and Epidemiology department for arranging and informing us all the necessary steps towards graduation during the last two years. I acknowledge Katherine Cook, graduate assistant of the department, who patiently set thesis examination
committee, time, date and place of thesis defense. Collective and individual acknowledgments are also owed to my classmates for the science discussion and exhilarating time we spent together as classmates. Of those, my special thanks go to Jing Jin, Vikki Ho and Anne Grundy who were very supportive in my graduate pathway.

Where would I be without my family? My parents deserve special mention for their support and prayers. My Father, Ali, in the first place was the person who taught by example the fundamentals of my learning character, showing me the joy of intellectual pursuit ever since I was a child. My Mother, Touran, was the one who sincerely raised me with her caring and gentle love. Farzaneh, Fereshteh, Firoozeh, Forouz, Abdollah, Aref and Erfan thanks for being supportive and caring siblings.

Words fail me to express my appreciation to my wife, Mojgan, whose dedication, love and persistent confidence in me has taken the load off my shoulder. She is the person who has spent most of her time in raising our two beautiful flowers, Pegah and Parsa. I love you all.

Abbas Esmaeili Tamandegani

Rajab 13, 1429 / Tir 27, 2008 / July 17, 2008
Table of Contents

Abstract ........................................................................................................ ii
Acknowledgement ................................................................................ iv
Table of Contents .................................................................................. vii
List of Figures ....................................................................................... x
List of Tables ......................................................................................... xii
List of Abbreviations ........................................................................... xiii

Chapter 1  INTRODUCTION ................................................................. 1

  Purpose ................................................................................................. 1

  Literature Review .............................................................................. 2

    Multiple Myeloma ............................................................................ 2

    Prognostic Factors In Multiple Myeloma .................................... 4

    Multiple Myeloma Treatment ....................................................... 7

    Roles of Dexamethasone ............................................................... 7

    NCIC CTG MY7 ............................................................................. 9

    Subgroup Analysis, Why? .............................................................. 14

    Subgroup Analysis in Multiple Myeloma .................................... 16

    Subgroup Analysis, How? .............................................................. 17

    Classification and Regression Trees .......................................... 18

  Rationale ............................................................................................. 21
Chapter 2 METHODS ................................................. 23

Ethical Considerations ............................................ 24
Descriptive Statistics .............................................. 24
Univariate analysis .................................................. 25
Prognostic model .................................................... 25
Predictive model .................................................... 28
Stratified log rank test ............................................. 30
Statistical softwares ............................................... 31

Chapter 3 RESULTS .................................................. 32
Study group .......................................................... 32
Univariate Analysis ................................................ 35
Prognostic model .................................................... 35
Predictive Model ...................................................... 45
Positive responders .............................................. 52
Negative responders .............................................. 58
Non-responders ................................................... 63

Chapter 4 DISCUSSION ............................................. 64
Interpretation of the Results .................................... 64
Strengths ............................................................. 67
Limitations .......................................................... 68
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Development of plasma cells from stem cells.</td>
<td>3</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Design and conduct of the MY-7 Trial.</td>
<td>10</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Schematic summary of statistical analysis of the current study.</td>
<td>27</td>
</tr>
<tr>
<td>Figure 4</td>
<td>Smoothing spline of martingale residuals for age.</td>
<td>36</td>
</tr>
<tr>
<td>Figure 5</td>
<td>Smoothing spline of martingale residuals for serum calcium concentration.</td>
<td>37</td>
</tr>
<tr>
<td>Figure 6</td>
<td>Smoothing spline of martingale residuals for albumin concentration.</td>
<td>38</td>
</tr>
<tr>
<td>Figure 7</td>
<td>Smoothing spline of martingale residuals for serum B2microglobulin.</td>
<td>39</td>
</tr>
<tr>
<td>Figure 8</td>
<td>Smoothing spline of martingale residuals for log-transformed serum B2microglobulin.</td>
<td>40</td>
</tr>
<tr>
<td>Figure 9</td>
<td>Smoothing spline of martingale residuals for blood glucose level.</td>
<td>41</td>
</tr>
<tr>
<td>Figure 10</td>
<td></td>
<td>42</td>
</tr>
</tbody>
</table>
Smoothing spline of martingale residuals for log-transformed blood glucose level.

**Figure 11** ........................................................................................................ 43

Smoothing spline of martingale residuals in MP arm (red) and MD arm (black).

**Figure 12** ........................................................................................................ 49

Ten predictive subgroups of MD arm.

**Figure 13** ........................................................................................................ 51

Corresponding subgroups of MD arm in MP arm.

**Figure 14** ........................................................................................................ 56

Determination of positive responders.

**Figure 15** ........................................................................................................ 57

Overall survival of positive responders in MD arm vs. that in MP arm.

**Figure 16** ........................................................................................................ 60

Determination of negative responders.

**Figure 17** ........................................................................................................ 62

Overall survival of negative responders in MD arm vs. that in MP arm.

**Figure 18** ........................................................................................................ 63

Overall survivals of non-responders (53 patients) in MD arm vs. that of corresponding subgroups in MP arm.
List of Tables

Table 1  ........................................................................................................ 5
Assessment of myeloma mass (Durie-Salmon).

Table 2  ........................................................................................................ 33
Baseline characteristics of the patients in both arms.

Table 3  ........................................................................................................ 34
Frequency distribution of categorical variables used in Cox regression analysis.

Table 4  ........................................................................................................ 44
Hazard ratios (HR) and p values for covariates in the Cox models for overall survival in patients of MP arm.

Table 5  ........................................................................................................ 48
The cutoff point used for each continuous variable along with the frequency distribution of the categories.

Table 6  ........................................................................................................ 70
Guidelines for deciding whether apparent differences in effects between subgroups are real.

Table 7  ........................................................................................................ 73
Bradford-Hill Criteria of Causality.

Table 8  ........................................................................................................ 90
Eastern Cooperative Oncology Group (ECOG) scaling, a simple measure of quality of life.
List of Abbreviations

**AIC**: Akaike Information Criterion

**CART**: Classification and Regression Tree

**CRP**: C-Reactive Protein

**ECOG**: Eastern Cooperative Oncology Group

**HGB**: Hemoglobin

**HR**: Hazard Ratios

**IL-6**: Interleukin 6

**LDH**: Lactate Dehydrogenase

**MD**: Melphalan and Dexamethasone

**MS**: Median Survival

**MMR**: Mean Martingale Residuals

**MP**: Melphalan and Prednisone

**NCIC CTG**: National Cancer Institute of Canada Clinical Trials Group

**PBSC**: Peripheral Blood Stem Cells

**PFS**: Progression Free Survival

**VAD**: Vincristine, Doxorubicin, and Dexamethasone

**WBC**: White Blood Cells
Chapter 1

INTRODUCTION

Purpose

MY7 was a multi-centre, non-blinded randomized phase III trial conducted by the National Cancer Institute of Canada Clinical Trials Group (NCIC CTG). Newly diagnosed patients with multiple myeloma were randomized to receive either melphalan plus prednisone (MP) or melphalan plus dexamethasone (MD) as induction therapy. Dexamethasone is more potent and longer-acting corticosteroid compared to prednisone. Overall survivals of the two groups didn’t show any a statistically significant difference at the end of trial. A
preliminary post-hoc subgroup analysis showed that patients with no intention to use bisphosphonates responded better to MD compared to MP. But, there were no guidelines regarding the use of bisphosphonates in the patients’ population, so it might not best characterize those patients. The current study was motivated by the fact, with intention to find baseline demographic and disease characteristics that could be used to identify patients who were positive responders to MD. As a by-product, we could also find factors that identified patients who benefited from MP.

**Literature Review**

**Multiple Myeloma**

Multiple myeloma (MM) is a malignancy of late-stage B cells that mature into neoplastic plasma cells rather than normal plasma cells (figure 1). These myeloma cells are present at multiple foci and produce a complete and/or partial (light-chain) monoclonal immunoglobulin protein. Symptoms are caused by tumor mass effects, by cytokines released directly by tumor cells or indirectly by host cells (marrow stroma and bone cells) in response to adhesion of tumor cells, and by the myeloma protein leading to deposition diseases, notably AL (amyloid-light chain) amyloidosis and light-chain deposition disease \(^1\). Canadian Cancer Society reported 2100 new cases of multiple myeloma per year, which accounts for 1.2% and 1.3% of all new cancer cases in females and males, respectively \(^2\). About 1% and 1.8% of mortalities due to cancers per year in women and men respectively, are
secondary to multiple myeloma. It has the second highest increase in cancer mortality burden (total number of deaths) and risk of death (age-standardized mortality rates) in men among all cancers in Canada, over the decade of 1992-2001. Although autologous stem cell transplantation is the standard of care for multiple myeloma, older patients with important co-morbidities will be excluded. For these patients, several chemotherapeutic regimens have been applied to treat multiple myeloma but melphalan plus prednisone (MP) has been the most widely used treatment strategy.

Figure 1: Development of plasma cells from stem cells. In multiple myeloma, late-stage B cells mature into neoplastic plasma cells rather than normal plasma cells.
Prognostic Factors In Multiple Myeloma

Once the diagnosis of multiple myeloma has been established, prognostic factors should be determined. The serum concentrations of beta\textsubscript{2} microglobulin plus albumin currently provide the most reliable prognostic markers for survival of patients with plasma cell myeloma \textsuperscript{7-10}. The newly developed International Staging System (ISS) consists of the following stages: stage I, serum beta\textsubscript{2} microglobulin <3.5 mg/L plus serum albumin >3.5 g/dL (median survival, 62 months); stage II, neither stage I nor stage III (median survival, 44 months); and stage III, serum beta\textsubscript{2} microglobulin >5.5 mg/L (median survival, 29 months) \textsuperscript{7}. On the other hand, Durie-Salmon staging system was in use for more than 30 years (table 1). It has been derived using standard laboratory measurements, including hemoglobin concentration, protein levels in serum and urine, presence of hypercalcemia, and extent of bone disease \textsuperscript{11}.

However, because of the problems with interpretation, especially of lytic bone lesions, other variables that are more quantitative have been used for tumor staging. Additional independent factors include platelet count, age \textsuperscript{7}, the plasma cell labeling index \textsuperscript{12,13} and C-reactive protein (CRP) levels, reflecting in vivo interleukin 6 (IL-6) activity \textsuperscript{14}. Increased IL-6 activity mediates many of the abnormalities encountered in multiple myeloma, including hypoalbuminemia, anemia, and lytic bone disease \textsuperscript{15-17}. Serum concentrations of syndecan-1, the hallmark of terminal plasma cell differentiation, reflect tumor burden and have been linked to outcome \textsuperscript{18-22}. The degree of marrow plasmacytosis, as assessed by flow cytometry of
DNA and cytoplasmic immunoglobulin, obviously reflects tumor burden and hence has prognostic utility\textsuperscript{23}. However, this evaluation is compromised by the patchy marrow involvement often observed in this malignancy. Hypodiploidy identifies marked resistance to standard drug regimens and, as a result, is associated with inferior survival\textsuperscript{24}.

### Table 1: Assessment of myeloma mass (Durie-Salmon).

1. **High tumor mass (stage III) (>1.2 x 10\textsuperscript{12}/m\textsuperscript{2})**
   One of the following abnormalities must be present:
   - A. Hemoglobin <8.5 g/dl, hematocrit <25%
   - B. Serum calcium >12 mg/dl
   - C. Very high serum or urine myeloma protein production rates:
     - I. IgG peak >7 g/dl
     - II. IgA peak >5 g/dl
     - III. Bence Jones protein >12 g/24 h
   - D. >3 lytic bone lesions on skeletal survey (bone scan not acceptable).

2. **Low tumor mass (stage I) (<0.6x10\textsuperscript{12}/m\textsuperscript{2})**
   All of the following must be present:
   - A. Hemoglobin >10.5 g/dl or hematocrit >32%
   - B. Serum calcium normal
   - C. Low serum myeloma protein production rates:
     - I. IgG peak <5 g/dl
     - II. IgA peak <3 g/dl
     - III. Bence Jones protein <4 g/24 h
   - D. No bone lesions or osteoporosis

3. **Intermediate tumor mass (stage II) (0.6-1.2 x 10\textsuperscript{12}/m\textsuperscript{2})**
   All patients who do not qualify for high or low tumor mass categories are considered to have intermediate tumor mass.
   - A. No renal failure (creatinine <2 mg/dl)
   - B. Renal failure (creatinine >2 mg/dl)
Cytologically plasmablastic myeloma, present in less than 10 percent of newly diagnosed patients, is an adverse parameter frequently associated with a high plasma cell labeling index\textsuperscript{25,26}. A high incidence of extramedullary disease, an elevated serum LDH level\textsuperscript{27,28}, and a high incidence of karyotypic anomalies, are all recognized to confer poor prognosis independently. t(4;14) is probably the most relevant adverse prognostic factor\textsuperscript{29} and deletion of the short arm of chromosome 1 (del Ip) is a strong predictor of poor outcome in myeloma patients undergoing an autotransplant\textsuperscript{30}. In the setting of high-dose therapy, histologic evaluation of marrow biopsy sections identifies short event-free and overall survival in the 20 percent of patients presenting with immature morphology (Bartl grade >1) and increased mitotic activity (>1 per high-power field), regardless of serum beta\textsubscript{2} microglobulin or CRP level or cytogenetics\textsuperscript{31}. In the absence of karyotype anomalies, patients who received tandem melphalan-based autotransplants, survived beyond 10 years. Therefore, some suggest that metaphase karyotyping should be performed before starting initial therapy or therapy for refractory disease\textsuperscript{32}. Increased marrow microvessel density has been associated with poor prognosis\textsuperscript{31,33}. The major angiogenic factors include vascular endothelial growth factor and basic fibroblast growth factor produced by myeloma cells themselves and by stromal cells in the marrow microenvironment\textsuperscript{34}. 
Multiple Myeloma Treatment

Several clinical trials have shown the superiority of autologous stem cell transplantation over conventional dose therapy for patients with multiple myeloma \(^{35-40}\). Also, thalidomide is effective as a first-line therapy for the treatment of multiple myeloma, but its use is limited by peripheral neurotoxicity \(^{41-46}\). Moreover, two randomized clinical trials \(^{4,5}\) and two controlled studies \(^{47,48}\) indicated that high-dose melphalan 200 mg/m\(^2\) supported by autologous peripheral blood stem cells (PBSC; mobilized with cyclophosphamide or etoposide plus filgrastim or filgrastim alone) can be effective in achieving marked tumor cytoreduction in patients up to age 70 years. To prevent hematopoietic stem cell damage that could compromise PBSC collection, induction therapy should not include hematopoietic stem cell-toxic regimens (melphalan, nitrosoureas, radiation to marrow-containing bone sites, such as pelvis and spine) \(^{49,50}\). Commonly used regimens include vincristine, doxorubicin, dexamethasone (VAD) \(^{51-53}\), dexamethasone alone \(^{54,55}\) or in combination with thalidomide \(^{56,57}\) and DT-PACE (dexamethasone, thalidomide, cisplatin, doxorubicin, cyclophosphamide, etoposide) \(^{58}\).

Roles of Dexamethasone

One of the current issues with regard to autologous PBSC-supported high dose therapy pertain to benefit from nonmyelotoxic peritransplantation therapy such as dexamethasone administered during the hematopoietic recovery phase to dampen the cytokine storm that may facilitate myeloma cell survival \(^{59}\). For
example, mucositis and other extramedullary toxicities commonly encountered in elderly patients receiving the standard high-dose melphalan regimen of 200 mg/m$^2$ are seldom seen in patients treated with melphalan 140 mg/m$^2$\textsuperscript{5}. In such cases, prior cytoreduction can be achieved with relatively noncytotoxic regimens, such as regimens using high-dose dexamethasone either alone or in combination with thalidomide \textsuperscript{59}. Although not conclusively studied, posttransplant maintenance strategies seem necessary to sustain disease control and survival. In the setting of standard therapy, higher glucocorticoid doses (prednisone 50 mg versus 10 mg on alternating days) have been shown, in a prospective randomized trial, to extend disease control and to improve event-free and overall survival \textsuperscript{60}, which may be further improved by the addition of thalidomide \textsuperscript{59}. Other applications of dexamethasone are in high risk and resistant or relapsing multiple myeloma patients \textsuperscript{61}, in smoldering myeloma \textsuperscript{62}, in renal failure in the course of multiple myeloma \textsuperscript{63}, in the acute management of hypercalcemia and myeloma-related bone pain \textsuperscript{64} and in spinal cord compression \textsuperscript{65}. Local radiation, typically administered with high-dose dexamethasone, has a palliative role in the treatment of focal lesions as part of disseminated myeloma. However, especially in newly diagnosed patients, primary systemic therapy controls focal problems, including cord compression, as quickly and effectively as does local radiation in the large majority of patients \textsuperscript{65,66}. A retrospective study suggested that the dose and length of treatment with steroids might influence the outcomes of multiple myeloma patients \textsuperscript{53}. A phase II clinical trial demonstrated that pulsed high-dose
dexamethasone may produce responses comparable with those seen with MP. These data led to conduction of NCIC CTG MY7 study.

**NCIC CTG MY7**

National Cancer Institute of Canada Clinical Trials Group (NCIC CTG) conducted MY7 as a multi-centre, non-blinded randomized phase III trial in 37 Canadian centers. MY7 was started in June 1995 and ended in Aug 2004. A total of 585 patients with newly diagnosed, histologically proven, untreated, symptomatic stage I or stage II or stage III multiple myeloma were enrolled in the study; 466 patients were randomized to receive either MP or MD as induction therapy and at the same time, patients of each group were randomized to get maintenance management with observation or dexamethasone, as shown in figure 2. Among them, 234 patients were assigned to MP arm and 232 patients were assigned to MD arm. The objectives of the trial were to compare the overall survival of previously untreated patients, with symptomatic myeloma, between those treated with MP and those treated with MD as induction therapy. Those who did not develop progressive disease on induction study, were further included in the maintenance study, which was designed to test whether dexamethasone as maintenance therapy would prolong the overall survival in comparison to those who received no further treatment. Exclusion criteria, inclusion criteria and stratification factors are shown in figure 2. All medications were administered orally. Each cycle was 28 days. The dose of melphalan was 9 mg/m²/4 days/cycle. The prednisone dose was 100 mg/day for 4 days every cycle. Induction doses of
dexamethasone were 40 mg/day for 4 days every 14 days for the first 84 days (3 treatment cycles) and then every four weeks with remaining cycles. Patients were to receive twelve cycles of therapy; doses of melphalan and corticosteroid were tapered or deleted according to a predetermined schedule based on the observed treatment-related toxicities. Patients who did not show disease progression after completing induction therapy were, as per their initial randomization, either observed or received dexamethasone 40 mg/day for 4 days/cycle until experiencing toxicity or progressive myeloma.

Figure 2: Design and conduct of the National Cancer Institute of Canada Clinical Trials Group MY-7 Trial. A total of 585 eligible patients were evaluated. An interim analysis evaluating induction therapy was conducted in November 2000; 408 patients were assessed and no differences in outcomes between induction therapies were detected. Randomization to induction therapy that included dexamethasone was therefore closed in January 2001, after entry of 466 patients; the final analysis of induction therapy included these patients and was based on the database locked in August 2004. Of the 585 patients randomised to maintenance therapy at the time of trial registration, 292 did not have disease progression during induction therapy and were included in the primary analysis evaluating maintenance therapy.
MY7 trial, Accrual June 1995 - July 2003

N = 595

Inclusion criteria

- histological confirmation;
- or
- bone marrow plasmacytosis > 10% and a measurable serum monoclonal paraprotein of IgA, IgG, IgD or IgE type;
- or
- urinary excretion > 1 g/24 h of monoclonal light chain protein;
- or
- marrow plasmacytosis of < 10% plus a measurable serum or urine paraprotein and associated with at least one osteolytic bone lesion.

Exclusion criteria

- a comorbid condition that precluded the use of protocol therapy;
- or a history of cancer other than for adequately treated squamous or basal cell carcinoma of the skin, carcinoma in situ of the cervix, or cancer that was treated more than 5 years before study entry and considered cured (1 patient);
- or a confirmed peptic ulcer disease within 2 years of study entry;
- or failure to satisfy criteria for MM (8 patients);
- or proxy consent (1 patient).

Stratification

- Center
- Durie Salmon stage (I/II vs. III)
- Creatinine < vs. > 1.75 mg/dl
- Bisphosphonate use

119 patients were randomized for maintenance study and treated with MP

MP arm
N = 234

466 patients were randomized

MD arm
N = 232
Clinical assessments, measurements of routine hematological and biochemical parameters and treatment-related toxicity assessments were done monthly. The primary outcome for both the induction and the maintenance questions, overall survival, was measured from the time of randomization until death from any cause. Secondary outcomes were response to treatment, progression free survival (PFS) and treatment-related toxicity. Patients must have received a minimum of four cycles of induction treatment to be evaluable for a response assessment. Criteria for response included a reduction in the serum monoclonal paraprotein of at least 50% and a reduction in the 24-hour urine excretion of monoclonal light chain of at least 90%. Each criterion required that the reduction be observed on two separate measurements taken at least 28 days apart. Criteria for progressive disease included an increase in the serum monoclonal paraprotein to >50% above the baseline value, or 10 g/L above the nadir value, an increase in the 24-hour urinary monoclonal light chain excretion to >100% above baseline or by more than 2 g/day above the nadir value, the development of hypercalcemia despite chemotherapy, the unequivocal development of a new lytic bone lesion, or a progressive cytopenia in conjunction with increasing marrow plasmacytosis. Patients who did not meet the criteria for either a response or progressive disease were classified as stable non-responders. Those patients who were responders or stable non-responders at the end of cycle 12 were eligible for maintenance management. PFS was measured from the time of randomization until meeting one of the above-mentioned criteria of progressive disease.
Treatment-related toxicity was graded by using the National Cancer Institute Common Toxicity Criteria version 2.0 (http://ctep.info.nih.gov/reporting/ctc.html).

MY-7 was aimed to detect an improvement of 50% in overall survival with MD induction compared to that of MP induction. With a power of 80% and a two-sided 5% level tests at least 196 observed deaths and accrual of 450 patients over 4.5 years with one additional year of follow-up were required. To detect a 50% improvement in overall survival with maintenance therapy of dexamethasone, and with a power of 80% and a one-sided 5% level test, a minimum of 162 deaths and accrual of 600 patients were required over 6 years with two additional years of follow-up. Interim analyses were planned for both the induction and maintenance therapies when around 50% of the required events had occurred. Distributions of overall survival and PFS between treatment arms were calculated with the life-table method of Kaplan and Meier and compared by the log rank test stratified by the variables used in the randomization process. The Cox regression model was used to study the treatment effect when adjusted for baseline factors. Primary analysis was performed using the modified intention-to-treat principle. For the maintenance therapy after induction therapy (MP or MD), each group was subdivided to observation or dexamethasone therapy. Secondary analysis was performed by comparing all four treatment groups to identify the best treatment sequence resulting from the induction and maintenance therapies.
In the final database for the updated induction analysis, 366 of the 466 patients randomized to MD or MP arms died (189 and 177 in MP and MD arms, respectively). Median follow-up time was 62.4 months. No statistically significant difference in survival time was detected between patients who received either MP or MD. The median overall survival was 30 months in those randomised to MP arm vs. 32.4 months in the patients of MD arm (HR =0.91, 95% confidence interval 0.74–1.11; P=0.3) 68.

**Subgroup Analysis, Why?**

In many clinical trials, we may want to examine the consistency of treatment benefit or non-benefit across two or more subgroups of the individuals studied. For example, we may want to know if the observed difference or non-difference is the same for young and old patients. This is recognized as subgroup analysis, which is the investigation of the influence of factors and treatment combination effects on the response variables in clinical trials. In clinical trials, treatment effects are reported based on the difference on population average response between treatment groups 68. But, patients do not share identical characteristics and equivalent risks. Some of the patients' pathologic and/or biologic factors may affect their response to a special type of treatment; some benefit significantly from a treatment, some get less benefit and some could be even harmed by the drug. The latter case may even happen for a drug that showed effectiveness for the whole trial population. An example of this is the application of melphalan and dexamethasone in
multiple myeloma treatment\textsuperscript{69}. Although the response rates at 6 months were significantly higher among patients receiving MD compared to those receiving the standard treatment (because of the higher frequency of severe pyogenic infections in the elderly patients of the MD arm), it is not recommended as the first-line treatment in this subgroup of patients\textsuperscript{70}. The reverse could also be true; i.e., treatment for a subset of trial patients may suggest a benefit while not being associated with meaningful differences in broader populations. Many examples of the latter situation are seen in the literature but the question is the validity of the estimates. For individual patients, subgroup analyses and secondary end points can provide the best guide for treatment selection. For health policy purposes the most important question is whether the primary outcome measure was clinically meaningful and statistically significant\textsuperscript{70}. If the answer to that question is no, as in our case, then the results of the subgroup analysis should be used only for generating hypotheses\textsuperscript{71,72}. The credibility of the estimates of the subgroup analysis are improved if the analysis is confined to the primary outcome and to a few predefined proper subgroups, on the basis of biologically plausible priori hypotheses\textsuperscript{72-75}, if the experiment is executed according to its protocol\textsuperscript{76} and if it has appropriately low levels of type I and type II error to produce a confirmatory evaluation of the intervention's effect in the subgroup\textsuperscript{77}. The findings of any subgroup analysis should not be over-emphasized or over-interpreted\textsuperscript{73}. Unless there is a strong prior hypothesis for a given
differential effect, any findings might be best viewed in the context of a hypothesis-generating exercise.

Subgroup Analysis in Multiple Myeloma

A meta-analysis of 6,633 patients from 27 randomized trials, found no difference, either overall or within any subgroup, in mortality between combination chemotherapy and melphalan plus prednisone. In terms of survival, these therapeutic options, as tested in the trials considered, were approximately equivalent. Subgroup analyses by type of combination chemotherapy or by dose-intensities of combination chemotherapy, of melphalan, or of prednisone did not identify any particular forms of therapy that were either clearly beneficial or clearly adverse. Similarly, analysis of the presentation features of the patients did not find any categories in which combination chemotherapy differed significantly from MP in its effects on mortality; in particular, there was no evidence that poor-risk patients benefited more from combination chemotherapy. On the other hand, the effects of high-dose glucocorticoid therapy on multiple myeloma have been evaluated by some studies. In the Alexanian et al study, the overall response rate of multiple myeloma patients to dexamethasone was similar to those with MP and patients who appeared most likely to benefit from dexamethasone were those with hypercalcemia or pancytopenia or who required simultaneous radiotherapy for a pathologic fracture. On the other hand, subgroup analysis in two other studies demonstrated no significant differences in PFS.
and overall survival between patients given high-dose therapy and those who were continued on standard dose \(^{69,85}\).

**Subgroup Analysis, How?**

Usual techniques for analysis to summarize treatment effect for the trial population are the Mantel-Haenszel test or stratified test \(^{86,87}\). The stratified test synthesizes treatment effects in various subgroups to an "overall" effect (i.e., assuming no interaction). This is meaningful if all subgroups have similar treatment effects. In other words, it is imperative to check whether treatment effects are homogenous across the levels of important factors. In general, however, if there are interactions between treatment and subgroups, the Mantel-Haenszel estimates will be misleading \(^{87}\). In order to recognize a combination of factors as predictive in regression models such as the Cox proportional hazard model, the particular combination has to be present in the model as an interaction term \(^{88}\). In the case assuming interactions among the factors, the order of the interaction term is unknown. In the latter case, all possible interaction terms of up to third order should be considered in order to give a chance for a covariate selection procedure to choose the statistically significant combination \(^{90}\). If there are many subgroup factors, the number of product terms necessary for an adequate modeling of the interactions may be higher than the number of observations and an analysis of the interactions is impossible. Moreover, as the number of variables and subsequently variable combinations increase, the number of statistical tests needed will
also be more than double that required to test for an interaction. All of these reduce the study's power to detect treatment differences. Then, since the trial is not large enough to detect subgroup findings, interaction tests commonly lack enough statistical power for this purpose (type II error). Sometimes one is left in doubt as to whether a suggested subgroup analysis is simply due to chance (type I error) or is intrinsically worth for further investigation.

To overcome these difficulties as an alternative approach the "classification tree" technique is proposed. Two different studies compared the two versions of the responder identification methods on 200 data sets (Cox proportional hazard vs. regression trees) and found that the Cox proportional hazard model with interactions is not a very sensitive method for this purpose when the effect of factors and factor combinations on treatment is weak and there is large percentage of censoring in the data. Regression trees showed acceptable power of identification for data in which the subgroups to be identified had much larger (or much smaller) hazard than the entire data set. Moreover, Kehl V et al showed that as a response factor in regression trees, it is preferable to use martingale residuals rather than deviance residuals. Given the limitations in Cox proportional hazard and the moderate percentage of censoring in our data (about 21.5%) we use the more appropriate method, the regression tree model, to better address the purpose of the project.

Classification and Regression Trees
The classic classification and regression trees (CART) algorithm was popularized by Breiman et al.\textsuperscript{89} It is a nonparametric technique that can select from among a large number of variables, those and their interactions that are most statistically significant in determining the outcome variable to be explained. In general terms, the purpose of the analyses via tree-building algorithms is to determine a set of if-then split conditions that permits accurate classification of subjects. Tree classification techniques, have a number of advantages over many alternative techniques. In most cases, the interpretation of results summarized in a tree is very simple for non-statisticians to interpret. This simplicity is useful not only for purposes of rapid classification of new observations, but can also often yield a much simpler "model" for explaining why observations are classified or predicted in a particular manner. On the other hand, tree methods are nonparametric and nonlinear\textsuperscript{89}. In other words, no assumptions are made related to the underlying distribution of values of the dependent or independent variables. Thus, with CART we can handle numerical data that are highly skewed or multi-modal, as well as categorical predicting cofactors with either ordinal or non-ordinal structure. CART is not affected by the outliers, collinearities, heteroscedasticity or distributional error structures that affect parametric procedures. Outliers are isolated into a node and thus have no effect on splitting. Contrary to situations in parametric modeling, CART makes use of collinear variables in surrogate splits. CART has the ability to detect and reveal variable interactions in the data set and it handles effectively large data sets and the issues of higher dimensionality. Another advantage of CART
analysis is that it is a relatively automatic method. In other words, compared to the complexity of the analysis, relatively little input is required from the analyst. This is in marked contrast to other multivariable modeling methods, in which extensive input from the analyst, analysis of interim results, and subsequent modification of the method are required.\textsuperscript{89-93}

The final results of using tree methods for classification or regression can be summarized in a series of (usually few) logical if-then terminal subgroups. Thus, tree methods are particularly well suited for data mining tasks, where there is often little a priori knowledge nor any coherent set of theories or predictions regarding which variables are related and how. In those types of data analyses, tree methods can often reveal simple relationships between just a few variables that could have easily gone unnoticed using other analytic techniques. Using CART, we can identify patients who are different from the whole patient population and cannot be explained by prognostic factors; i.e. responders, whether positive or negative. Then, the responders are patients in the new treatment group, who are badly predicted by the prognostic model. For the successful identification of predictive factors, we need the assumption that all prognostic factors are already correctly accounted for in a prognostic model; this is a strong, but not unreasonable assumption.\textsuperscript{91-93} In essence, prognostic factors are predicting factors of outcome that are independent of treatment administered. Predictive factors are those factors that forecast the outcome following treatment (either in terms of tumor shrinkage or a survival benefit from treatment). Therefore, prognostic factors
are tumor and patient characteristics affecting patient outcome, whereas predictive factors define the effect of treatment on the tumor (response or resistance to particular therapeutic agent). Some factors may have both prognostic and predictive utility ⁹⁴-⁹⁷.

**Rationale**

In many clinical trials, we may want to examine the consistency of treatment difference (or resemblance) across two or more subgroups of the individuals studied. NCIC CTG MY7, which is the original clinical trial for the current study, was designed to test whether MD as induction and/or maintenance therapy improves overall survival over standard treatment of MP in patients with newly diagnosed multiple myeloma. The results reported by Shustik et al showed that dexamethasone as either induction or maintenance therapy did not significantly prolong the patients’ overall survival in comparison with MP or observation ⁶⁷. But, a preliminary subgroup analysis of MY7 data demonstrated that MD improved overall survival significantly over MP for patients with no intended use of bisphosphonate (median survival was 2.45 years for the 47 patients on the MP arm vs. 3.24 years for the 43 patients on the MD arm, HR = 0.59, 95% CI 0.37 to 0.96) while there was no general guideline indicating who should receive bisphosphonate in this trial. Assuming the subgroup effect is real, this factor may not best describe the responder subgroup of patients. This observation resulted in the proposed research project to identify patients who might benefit from MD in MY7 trial. In other
words, we would like to see if this subgroup of responders can be characterized by the biological and pathological information collected at study entry. An appropriate method of identifying such patients is based on the martingale residuals derived from fitting data on the MD arm with the estimated parameters and baseline cumulative hazard from the prognostic model built from the classical MP treatment arm Cox proportional hazard model. Using the martingale residuals, we can identify patients whose overall survival is poorly predicted by the prognostic model \(^9\), but correlated to the baseline factors of these patients. The predictive factors, which characterize the poorly fitted patients will be those identifying responders to MD in MY-7 trial. Given the issues related to data-driven cutoff points for each factor used in the analysis, it is clear that we have to use previous knowledge about the predictive factors in order to ensure a proper method of subgroup analysis. Literature review revealed only two studies using CART in multiple myeloma, which found serum beta\(_2\) microglobulin, serum albumin, platelet count, serum creatinine and age as powerful predictors of patients survival \(^7,10\). Some studies demonstrated, through other analytical methods, that patients with hypercalcemia or pancytopenia or who required simultaneous radiotherapy for a pathologic fracture benefited more from dexamethasone treatment compared to standard treatment and elderly patients benefited less \(^55,70,81\). We found no study in the literature, using CART to identify responders to MD in multiple myeloma. That's why we proposed the current study to present a classical approach for future reference in multiple myeloma clinical trials.
Chapter 2

METHODS

This thesis was a data mining and subgroup analysis of a completed clinical trial study (MY7) comparing long-acting corticosteroid (dexamethasone) with shorter-acting corticosteroid (prednisone) in combination with melphalan for multiple myeloma treatment.
Ethical Considerations

The main study from which data was taken for this thesis has already received human ethics approval and approval for this thesis project was sought and obtained from Research Ethics Board at Queen's University. Written informed consent has been obtained from all patients for participation in the original study (MY7) with one copy of the consent form retained by the subject. Participation was voluntary and patients were free to withdraw from the study at any time during its course. When biological specimens were transported, only identification numbers were used. All identifying information was held in strict confidence in locked filing cabinets, accessible only to study personnel. Computerized data files contained no identifying information; data was not shared with any employer and number codes were used in the analysis. Only grouped data are published.

Descriptive Statistics

Descriptive statistics were used to describe the basic features of the data in the study. The distribution of each variable and the frequency distribution of each value were evaluated. Measures of central tendency (mean, median and mode) and the spread of the values around the central tendency (the range and the standard deviation) were determined. Factors included in the analysis were as follows: age, sex, albumin concentration, beta_2 microglobulin level, Durie-Salmon stage, hemoglobin concentration, white blood cells, platelet counts, serum
calcium level, serum creatinine level, blood glucose level, bone marrow plasma cells, ECOG (Eastern Cooperative Oncology Group) performance status (see appendix, table 7) and radiologic skeletal assessment

**Univariate Analysis**

The associations of various patient and disease characteristics with survival were evaluated through univariate analysis using Cox regression model. Covariates measured on a continuous scale, were considered as continuous variables. The appropriate functional form of each covariate was visually investigated using lowess-smoothed plots of the martingale residuals from the Cox model. Natural log-transformation was carried out for some covariates to better fit the model assumption.

**Prognostic Model**

A prognostic model of the primary endpoint of overall survival was first built based on patients in standard treatment MP arm. Cox proportional hazards model was used to study the associations of various important baseline patients' demographics and disease characteristics with their survival and identify subsets of variables upon which the hazard function depends. All 14 covariates were used as the independent variables. Continuous covariates were used as continuous variables in Cox model except beta_2 microglobulin and albumin, which had high frequency of missing values. In order to include all patients in analysis, missing values were considered as a separate category, and those with data
were categorized according to well-known criteria from literature review. Overall, 6 covariates were considered as categorical variables (table 3). Stepwise procedure based on minimum Akaike Information Criterion (AIC) was used to determine the final prognostic model, which identified factors that the hazard function depended on. Based on this model, the baseline cumulative hazard function and coefficients for those prognostic factors were obtained. These estimates were used to derive martingale residuals for patients on MD arm (figure 3). Subjects with negative residuals were those who survived longer than expected (potential positive responders), while those with positive residuals were those who survived shorter than expected (candidates for negative responders) after being treated with MD. The next step was the development of the predictive model for the MD arm. The information (residual outliers) for the MD arm from fitting the MP arm’s prognostic model was used to develop the predictive model. This was accomplished through CART.
Figure 3: Schematic summary of statistical analysis of the current study.
Predictive Model

CART analysis was done to identify independent predictive factors for survival time after treatment. For clinical use, all independent factors were categorized mainly according to the literature review. Also, continuous covariates were investigated by the lowess-smoothed plot of the martingale residuals against the covariate in both arms to find the proper cutoff points. Various cutoff points, including those shown to be significant in published studies of multiple myeloma patients were individually analyzed. Then, continuous covariates were categorized based on proper cutoff points for CART analysis as follows: age (≤75 vs. >75 years), albumin concentration (≤3.5 g/dL vs. >3.5 g/dL vs. missing), hemoglobin concentration (<10 g/L vs. ≥10 g/L), white blood cells counts (<4000 cells/µL vs. ≥4000/mm³), platelet counts (<150,000 cells/µL vs. ≥150,000 cells/µL), serum calcium level (≤2.6 mmol/L vs. >2.6 mmol/L), serum creatinine level (<175 µmol/L vs. ≥175 µmol/L), blood glucose level (≤100 mg/dl vs. >100 mg/dl) and bone marrow plasma cells (≤40% vs. >40%). Beta₂ microglobulin level was not included in CART analysis because of excessive missing values. Also, Four categorical variables including sex (male vs. female), Durie-Salmon stage (stage 1 vs. stage 2 vs. stage 3), ECOG performance status (<2 vs. ≥2), radiologic skeletal assessment (<4 vs. ≥4 lytic bone lesions), were considered in tree construction.

In brief, for the CART analysis, martingale residuals were used as response variable and all 13 covariates in categorical shapes were used as
independent variables to develop the predictive model. The mean of the martingale residuals in a subgroup is the summary statistic in this method. Starting with all patients in the MD arm, the CART analysis first determined the proper possible predictor by which the population may well be split into two subgroups most different in the mean of martingale residuals. All 13 variables were examined at each new separation, which was binary; i.e., whether subjects are in one category versus all other categories. As mentioned before, the cutoff point of each predictor was determined based on previous studies or lowess-smoothed plots of the martingale residuals in both arms. The splitting was based on maximum reduction of within subgroup variances (homogeneity) of the martingale residuals. The splitting process was repeated on resulting subgroups until no further partitioning was reasonable, either because a subgroup was homogeneous for the martingale residuals or because the subgroup was too small to be divided further. This led to splitting up the MD arm into terminal subgroups (end nodes) with potential positive or negative response to MD. Subgroups of responders in the new treatment group were those with relative large positive or negative residuals. In other words, patients who were not well predicted (outliers in the martingale residuals) were candidates for responders; the negative outliers for positive responders and the positive outliers for negative responders. Then, patients in the MP arm were divided into groups corresponding to the same subgroups as those determined by factors obtained from the MD arm predictive model.
Stratified log rank test

The survival curves of each MP-MD pair of subgroups or combinations of subgroups were compared through log rank test stratified by subgroups. First, the subgroups in MD arm were considered one by one and sorted by the size of their mean martingale residuals. The largest negative subgroup in absolute value of the mean of martingale residuals was identified. Overall survival of the patients in that subgroup was compared with the overall survival in the paired MP arm subgroup and the $p$ value of the log rank statistic was calculated. Then, the patients identified in the next largest negative subgroup in the MD arm were added to the previously considered subgroup. Again, the overall survivals of this combination of subgroups in both arms were compared and the $p$ value of the stratified log rank test was determined. This process was repeated until there was no more negative subgroup. The point at which the $p$-value stopped decreasing $^{90}$, identified the set of positive responders and the factors involved in defining the combination of subgroups at this point were considered predictive factors for positive response. In order to identify negative responders, the above-mentioned process was repeated for all subgroups with positive mean martingale residuals. The most significant combination of positive terminal subgroups defined negative responders and the factors involved in their definition were considered predictive factors for negative response. In summary, assessment of the subgroup combinations by their size, mean martingale residuals, overall survival and the $p$ values of the stratified log-rank statistics, led to choose the
appropriate subgroups as the positive or negative responders. \( P \) values less than 0.05 were considered significant.

**Statistical Softwares**

Univariate as well as multivariate analyses were performed by SAS System for Windows, version 9.1 (SAS Inc., Cary, NC). CART analysis and log rank tests were accomplished by R Project for Statistical Computing, version 2.6.2 ([http://www.r-project.org/](http://www.r-project.org/)).
Chapter 3

RESULTS

Study Group

Baseline characteristics of the patients in each group are outlined in table 2 (for continuous covariates) and table 3 (for categorical covariates). Briefly, the final study group for induction analysis comprised 234 and 232 patients in MP and MD arms, respectively. The male/female ratio was 1.34 in MP arm and 1.64 in MD arm. About 95% of patients had stages II-III disease in both arms. Cytogenetic data were not available at the time of study.
Table 2: Baseline characteristics of the patients in both arms.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>MP arm (n = 234)</th>
<th></th>
<th></th>
<th>MD arm (n = 232)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Minimum</td>
<td>Maximum</td>
<td>Median</td>
<td>Minimum</td>
<td>Maximum</td>
<td>Median</td>
</tr>
<tr>
<td>Age</td>
<td>41</td>
<td>91</td>
<td>71</td>
<td>42</td>
<td>88</td>
<td>71</td>
</tr>
<tr>
<td>Blood glucose level (mg/dl)</td>
<td>3.6</td>
<td>17.2</td>
<td>5.7</td>
<td>3.1</td>
<td>19</td>
<td>5.7</td>
</tr>
<tr>
<td>Hemoglobin concentration (g/L)</td>
<td>65</td>
<td>153</td>
<td>101.5</td>
<td>64</td>
<td>163</td>
<td>100</td>
</tr>
<tr>
<td>WBC counts (cells/µL)</td>
<td>1.2</td>
<td>15.4</td>
<td>5.5</td>
<td>1.1</td>
<td>22</td>
<td>5.65</td>
</tr>
<tr>
<td>Platelets counts (cells/µL)</td>
<td>52,000</td>
<td>556,000</td>
<td>212</td>
<td>44,000</td>
<td>569,000</td>
<td>212</td>
</tr>
<tr>
<td>Albumin concentration (g/L)*</td>
<td>16</td>
<td>49</td>
<td>34</td>
<td>18</td>
<td>56</td>
<td>34</td>
</tr>
<tr>
<td>Serum creatinine concentration (µmol/L)</td>
<td>37</td>
<td>964</td>
<td>104</td>
<td>50</td>
<td>809</td>
<td>107</td>
</tr>
<tr>
<td>Calcium concentration (mmol/L)</td>
<td>1.6</td>
<td>3.79</td>
<td>2.33</td>
<td>1.64</td>
<td>3.79</td>
<td>2.34</td>
</tr>
<tr>
<td>Beta\textsubscript{2} microglobulin level (mg/L)*</td>
<td>1.2</td>
<td>4620</td>
<td>204</td>
<td>1.22</td>
<td>4570</td>
<td>180</td>
</tr>
<tr>
<td>Durie-Salmon stage</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Radiologic skeletal assessment (scales based on number of lytic bone lesions)</td>
<td>0</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Bone marrow pathology (percent)</td>
<td>1</td>
<td>100</td>
<td>39</td>
<td>0</td>
<td>100</td>
<td>38</td>
</tr>
<tr>
<td>ECOG performance status</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>

* In order to include all patients in Cox analysis, missing values of beta\textsubscript{2} microglobulin level and albumin concentration were considered as a separate category, and those with data were categorized according to the well-known criteria from literature review.
Table 3: Frequency distribution of categorical variables used in Cox regression analysis.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>MP arm (n=234)</th>
<th>MD arm (n=232)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
<td>Percent</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>100</td>
<td>43</td>
</tr>
<tr>
<td>Male</td>
<td>134</td>
<td>57</td>
</tr>
<tr>
<td><strong>ECOG performance status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;2</td>
<td>146</td>
<td>62.5</td>
</tr>
<tr>
<td>≥2</td>
<td>88</td>
<td>37.5</td>
</tr>
<tr>
<td><strong>Albumin concentration</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤3.5 g/L</td>
<td>133</td>
<td>57</td>
</tr>
<tr>
<td>&gt;3.5 g/L</td>
<td>96</td>
<td>41</td>
</tr>
<tr>
<td>Missing</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td><strong>B2 microglobulin concentration</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;3.5 mg/L</td>
<td>51</td>
<td>22</td>
</tr>
<tr>
<td>≥3.5 and ≤5.5 mg/L</td>
<td>44</td>
<td>19</td>
</tr>
<tr>
<td>&gt;5.5 mg/L</td>
<td>56</td>
<td>24</td>
</tr>
<tr>
<td>Missing</td>
<td>83</td>
<td>35</td>
</tr>
<tr>
<td><strong>Durie Salmon stage</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>13</td>
<td>5.5</td>
</tr>
<tr>
<td>2</td>
<td>57</td>
<td>24.5</td>
</tr>
<tr>
<td>3</td>
<td>164</td>
<td>70</td>
</tr>
<tr>
<td><strong>Radiologic skeletal assessment</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;4 lytic bone lesions</td>
<td>122</td>
<td>52</td>
</tr>
<tr>
<td>≥4 lytic bone lesions</td>
<td>112</td>
<td>48</td>
</tr>
</tbody>
</table>
Univariate Analysis

Univariate analysis was performed to identify potential prognostic variables of survival in MP arm (table 4). Functional forms for continuous variables were determined by inspection of lowess-smoothed plots. Figures 4-10 depict the functional relationship between some of the continuous covariates and survival in MP arm using a smoothing spline. A smooth fit of the martingale residuals versus continuous covariates confirmed this relationship. Also, the presentation of both the smooth fit and the individual residuals provided insight into the influence of particular individuals on the estimate of the functional form. Moreover, the functional form for each covariate between the two arms was compared. An example is demonstrated in figure 11. They revealed similar patterns for most covariates.

Prognostic Model

All factors were considered for the multivariable analysis of Cox proportional hazards regression model (table 4). Stepwise selection process led to identification of the following variables as the prognostic factors in MP arm: sex, Durie-Salmon stage, beta_{2} microglobulin level, serum creatinine concentration and blood glucose levels. The resulting fit indicated that age, serum albumin concentration, ECOG performance status, radiologic skeletal assessment (skeletal survey), bone marrow pathology, serum calcium level, hemoglobin level, platelets and WBC counts do not have any prognostic significance in patients with multiple myeloma in MP arm.
Figure 4: Smoothing spline of martingale residuals to illustrate the functional relationship between age and survival in MP arm.
Figure 5: Smoothing spline of martingale residuals to illustrate the functional relationship between serum calcium concentration and survival in MP arm.
Figure 6: Smoothing spline of martingale residuals to illustrate the functional relationship between albumin concentration and survival in MP arm. When the smooth fit intersects the reference line in more than one point, usually the last one will be considered as the cut off point, but in reality there may be more than one cut off point for each covariate.
Figure 7: Smoothing spline of martingale residuals to illustrate the functional relationship between serum beta$_2$ microglobulin and survival in MP arm. As is clear, the data concentration to one side makes it difficult to accurately observe the smoothness of plot and estimate the cut off point. That’s the reason for log-transforming the serum beta$_2$ microglobulin data in the next plot.
Figure 8: Smoothing spline of martingale residuals to illustrate the functional relationship between log-transformed serum beta$_2$ microglobulin and survival in MP arm.
Figure 9: Smoothing spline of martingale residuals to illustrate the functional relationship between blood glucose level and survival in MP arm.
Figure 10: Smoothing spline of martingale residuals to illustrate the functional relationship between log-transformed blood glucose level and survival in MP arm.
Figure 11: Smoothing spline of martingale residuals in MP arm (red) and MD arm (black) to illustrate the functional relationship between log-transformed serum creatinine concentration and survival. More or less, they follow a similar pattern.
Table 4: Hazard ratios (HR) and $p$ values for covariates in the Cox models for overall survival in patients in MP arm. Missing values of albumin and beta$_2$ microglobulin were considered as separate categories.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Univariate Analysis</th>
<th>Full Model</th>
<th>Multivariable Final Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Crude HR (95%CI)</td>
<td>$P$</td>
<td>Adjusted HR</td>
</tr>
<tr>
<td>Age</td>
<td>1.01 (0.99-1.03)</td>
<td>0.3</td>
<td>1.006</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>1</td>
<td>0.012</td>
<td>1.34</td>
</tr>
<tr>
<td>Male</td>
<td>1.45 (1.08-1.95)</td>
<td></td>
<td>1.34</td>
</tr>
<tr>
<td>ECOG performance status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;2</td>
<td>1</td>
<td>0.13</td>
<td>1.12</td>
</tr>
<tr>
<td>≥2</td>
<td>1.25 (0.94-1.68)</td>
<td></td>
<td>1.12</td>
</tr>
<tr>
<td>Blood glucose level</td>
<td>1.025 (0.95-1.1)</td>
<td>0.5</td>
<td>1.46</td>
</tr>
<tr>
<td>Hemoglobin concentration</td>
<td>0.99 (0.98-.997)</td>
<td>0.008</td>
<td>1.001</td>
</tr>
<tr>
<td>WBC counts</td>
<td>1.06 (0.99-1.13)</td>
<td>0.086</td>
<td>1.04</td>
</tr>
<tr>
<td>Platelets counts</td>
<td>1</td>
<td>0.37</td>
<td>0.999</td>
</tr>
<tr>
<td>Albumin concentration*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤3.5 g/L</td>
<td>1</td>
<td>0.15</td>
<td>0.86</td>
</tr>
<tr>
<td>&gt;3.5 g/L</td>
<td>0.81 (0.6-1.08)</td>
<td></td>
<td>0.86</td>
</tr>
<tr>
<td>Missing</td>
<td>4.9 (0.67-35.6)</td>
<td>0.12</td>
<td>0.99</td>
</tr>
<tr>
<td>Serum creatinine concentration</td>
<td>1.002 (1.001-1.003)</td>
<td>0.0001</td>
<td>1.51</td>
</tr>
<tr>
<td>Calcium concentration</td>
<td>1.4 (0.847-2.28)</td>
<td>0.2</td>
<td>0.92</td>
</tr>
</tbody>
</table>
Table 4, (continued).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Univariate Analysis</th>
<th>Full Model</th>
<th>Multivariable Final Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Crude HR (95%CI)</td>
<td>Adjusted HR</td>
<td>Adjusted HR</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td><strong>Beta\textsubscript{2} microglobulin level</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;3.5 mg/L</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>≥3.5 and ≤5.5 mg/L</td>
<td>1.8 (1.1-3.005)</td>
<td>1.74</td>
<td>0.04</td>
</tr>
<tr>
<td>&gt;5.5 mg/L</td>
<td>2.9 (1.85-4.54)</td>
<td>2.08</td>
<td>0.007</td>
</tr>
<tr>
<td>Missing</td>
<td>2.25 (1.48-3.4)</td>
<td>1.99</td>
<td>0.002</td>
</tr>
<tr>
<td><strong>Stage</strong>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1.8 (0.82-4.04)</td>
<td>1.58</td>
<td>0.29</td>
</tr>
<tr>
<td>3</td>
<td>2.5 (1.16-5.32)</td>
<td>2.11</td>
<td>0.11</td>
</tr>
<tr>
<td><strong>Radiologic skeletal assessment</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;4 lytic bone lesions</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>≥4 lytic bone lesions</td>
<td>1.02 (0.9-1.12)</td>
<td>0.74</td>
<td>0.92</td>
</tr>
<tr>
<td><strong>Bone marrow pathology</strong></td>
<td>1.008 (1.002-1.013)</td>
<td>0.008</td>
<td>1.003</td>
</tr>
</tbody>
</table>

*Albumin level ≤3.5 g/dL was considered as the reference category.

**Beta\textsubscript{2} microglobulin level ≤3.5 mg/L was considered as the reference category.

***Stage 1 was considered as the reference category.

Predictive Model

Age, bone marrow pathology, albumin and calcium concentrations, hemoglobin and blood glucose levels, platelets and WBC counts were initially considered as continuous variables for developing the prognostic model and then were categorized for the purposes of the CART analysis. Categories were determined based on inspection of martingale residual plots as well as consideration of previously reported cutoff points in multiple myeloma. For example, figure 4 demonstrates that smoothing spline intersected the reference line at age 70 to 75
years. Literature review showed different ages from 60 to 72 years as the proper
cutoff point. We developed different predictive models of CART based on cutoff
points of 70 to 75 years for age and finally, age ≤75 vs. >75 years was
considered in the final predictive model. The cutoff point(s) used for each
variable along with the frequency distribution of different categories is
demonstrated in table 5. Since more than one-third of patients had missing
values of serum beta\textsubscript{2} microglobulin, this covariate was not included in the CART
analysis. In order to carry out sensitivity analysis, different cutoff points for
continuous variables were considered in CART analysis and the results were
more or less similar. Moreover, a few variables were included at first and were
excluded later such as, past history of radiotherapy and intention to use
bisphosphonates. In all conditions, the results were not different.

The CART analysis identified 10 terminal subgroups (end nodes)
according to interaction of 8 (out of 14) predictive factors as follows (figure 12):
age (≤75 vs. >75 years), serum calcium concentration (>2.6 vs. ≤2.6 mg/dl),
Durie-Salmon stage (stage 1 vs. other stages), ECOG performance status (<2
vs. ≥2), radiologic skeletal assessment (≥4 vs. <4 lytic bone lesions), sex,
hemoglobin concentration (<100 vs. ≥100 mg/dl) and WBC count (<4,000 vs.
≥4000 cells/µL) (figure 12). The first predictive factor was age, which divided
patients into two large subgroups. Patients older than 75 years, which included
one-fifth of the patients in MD arm, had positive martingale residuals and were
candidates of negative responders. The second split of the regression tree
included calcium concentration in patients aged ≤75, which captured a total of
183 patients. Twenty-four of them were candidates of negative responders. The third most predictive factor was Durie-Salmon stage, which included 159 patients. Of these, 150 patients were categorized based on ECOG performance status. Interestingly, about 85% of those with ECOG performance status <2 (101 patients) had negative martingale residuals and were candidates of positive responders. Patients were further divided into 7 terminal subgroups based on interaction of the following predictive factors: radiologic skeletal assessment (skeletal survey) and sex on one side, hemoglobin concentration, WBC counts and radiologic skeletal assessment (skeletal survey) on the other side. In the next step, the MP arm was divided into the same ten terminal subgroups (figure 13). This was necessary to be able to compare each subgroup in MD arm to its corresponding subgroup in MP arm.
Table 5: The cutoff point used for each continuous variable along with the frequency distribution of the categories.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>MP arm n = 234 (%)</th>
<th>MD arm n = 232 (%)</th>
<th>Total n = 466 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤75 years</td>
<td>175 (75%)</td>
<td>183 (79%)</td>
<td>358 (77%)</td>
</tr>
<tr>
<td>&gt;75 years</td>
<td>59 (25%)</td>
<td>49 (21%)</td>
<td>108 (23%)</td>
</tr>
<tr>
<td><strong>Glucose concentration</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤100 mg/dl</td>
<td>97 (41.5%)</td>
<td>90 (39%)</td>
<td>187 (40%)</td>
</tr>
<tr>
<td>&gt;100 mg/dl</td>
<td>122 (52%)</td>
<td>120 (52%)</td>
<td>242 (52%)</td>
</tr>
<tr>
<td>Missing</td>
<td>15 (6.5%)</td>
<td>22 (9%)</td>
<td>37 (8%)</td>
</tr>
<tr>
<td><strong>Hemoglobin concentration</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;10 g/L</td>
<td>107 (46%)</td>
<td>112 (48%)</td>
<td>219 (47%)</td>
</tr>
<tr>
<td>≥10 g/L</td>
<td>127 (54%)</td>
<td>120 (52%)</td>
<td>247 (53%)</td>
</tr>
<tr>
<td><strong>White blood cells counts</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;4,000 cells/µL</td>
<td>43 (18.5%)</td>
<td>41 (18%)</td>
<td>84 (18%)</td>
</tr>
<tr>
<td>≥4,000 cells/µL</td>
<td>191 (81.5%)</td>
<td>191 (82%)</td>
<td>382 (82%)</td>
</tr>
<tr>
<td><strong>Platelets counts</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;150,000 cells/µL</td>
<td>44 (19%)</td>
<td>40 (17%)</td>
<td>84 (18%)</td>
</tr>
<tr>
<td>≥150,000 cells/µL</td>
<td>190 (81%)</td>
<td>192 (83%)</td>
<td>382 (82%)</td>
</tr>
<tr>
<td><strong>Serum Creatinine</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>concentration</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;175 µmol/L</td>
<td>197 (84%)</td>
<td>195 (84%)</td>
<td>392 (84%)</td>
</tr>
<tr>
<td>≥175 µmol/L</td>
<td>37 (16%)</td>
<td>37 (16%)</td>
<td>74 (16%)</td>
</tr>
<tr>
<td><strong>Calcium concentration</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤2.6 mmol/L</td>
<td>196 (84%)</td>
<td>198 (85%)</td>
<td>394 (84.5%)</td>
</tr>
<tr>
<td>&gt;2.6 mmol/L</td>
<td>38 (16%)</td>
<td>34 (15%)</td>
<td>72 (15.5%)</td>
</tr>
<tr>
<td><strong>Bone marrow pathology</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤40% plasma cells</td>
<td>131 (56%)</td>
<td>127 (55%)</td>
<td>258 (55.5%)</td>
</tr>
<tr>
<td>&gt;40% plasma cells</td>
<td>103 (44%)</td>
<td>105 (45%)</td>
<td>208 (44.5%)</td>
</tr>
</tbody>
</table>
Figure 12: Ten predictive subgroups of MD arm according to age, serum calcium concentration, Durie-Salmon stage, ECOG performance status, HGB (hemoglobin) concentration, radiologic skeletal assessment (skeletal survey) and sex formed by CART analysis. The first number in each terminal subgroup denotes mean martingale residuals (MMR) of the subgroup; the second number shows median survival (MS) in days and the third one is the number of patients in that subgroup.
Figure 13: Corresponding subgroups of MD arm in MP arm. The first number in each terminal subgroup denotes median survival (MS) in days and the second number is the number of patients in the subgroup.

Positive Responders

At the beginning, the subgroups were considered one by one and they were ordered by the size of their mean martingale residuals (numbers in red color in right lower angles of terminal subgroups in figure 12). The terminal subgroups with negative mean martingale residuals were candidates of positive responders (terminal subgroups with happy face in figure 12). First, the terminal subgroup with the largest negative mean martingale residuals in MD arm (subgroup number 1 in figure 12) was chosen and its overall survival was compared with the overall survival in the corresponding subgroup in the MP arm through stratified log rank test. The resulting \( p \) value (0.71) was plotted in figure 14. This subgroup was developed based on interaction of 7 different predictive factors, starting from age, ending by radiologic skeletal assessment (skeletal survey). It involved 23 and 20 patients in MD and MP arms, respectively. Fourteen patients were dead at the end of study in each subgroup. Median survival times of these patients in MP and MD arms were 39.5 and 42.5 months, respectively (HR = 0.86).

In the second step, the two subgroups with the largest negative mean martingale residuals in the MD arm (terminal subgroups 1 and 2 in figure 12)
were combined. This combination involved 44 patients in MD arm and 38 patients in MP arm. About 36% and 19% of the patients were alive at the end of study in each combined subgroup, respectively. Their median survival times were 41 and 36 months, respectively (HR = 0.67). The overall survival of this combination was compared with that of the corresponding combination in MP arm by stratified log rank test and the \( p \) value (0.13) was plotted in figure 14.

Then, the subgroup number 3 in figure 12 was added to the last combination in both arms. This combination included 97 patients in MD arm and 93 patients in MP arm. While 42% of these patients were alive in MD arm at the end of study, only 21% had similar fate in MP arm. Median survival times were 44.5 and 33 months, respectively. The difference was significant (HR = 0.56; \( p = 0.0014 \)) and the HR was the smallest compared to the last ones. The \( p \) value of stratified log rank statistics of two overall survivals (0.0014) was plotted in figure 14. The stratified log rank \( p \) value was the smallest compared to the ones in the first and second steps.

Finally, the last terminal subgroup with negative mean martingale residual in MD arm (subgroup number 4 in figure 12) was added to the previous combination. This covered 107 and 101 patients in MD and MP arms, respectively. About 40% and 22% in each combination in the respected arms were alive at the end of study. Median survival times were 42.5 and 33 months, respectively. The hazard ratio was 0.6. The overall survival of this combination
was compared with that of the corresponding combination in MP arm by stratified log rank statistics and the \( p \) value (0.0025) was plotted in figure 14.

The search for positive responders must be stopped at the combination of terminal subgroups, immediately before the stratified log rank \( p \) value stops decreasing \(^90\). Then, the combination of first, second and third terminal subgroups in figure 12 can be considered positive responders. The overall survival of positive responders in MD arm was plotted against the overall survival of the corresponding combination of the three subgroups in MP arm (figure 15).

In conclusion, around 42% of the patients in MD arm were identified as positive responders. All of them had 3 similar characteristics: age \( \leq 75 \) years, serum calcium concentration \( \leq 2.6 \) mmol/L and Durie-Salmon stages 2 or 3 of multiple myeloma. More than half of them had ECOG performance status \(< 2\) and HGB\(\geq 100\) mg/dl. Those with ECOG performance status \(< 2\) and HGB\(< 100\) mg/dl were among positive responders if they had less than 4 lytic bone lesions and didn’t have leukopenia. Around one-fifth were males with ECOG performance status \(\geq 2\) and \(\geq 4\) lytic bone lesions on radiologic skeletal assessment (skeletal survey).
<table>
<thead>
<tr>
<th></th>
<th>Subgroup 1</th>
<th>Subgroups 1 &amp; 2</th>
<th>Subgroups 1, 2 &amp; 3</th>
<th>Subgroups 1, 2, 3 &amp; 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Patients</td>
<td>20</td>
<td>38</td>
<td>93</td>
<td>101</td>
</tr>
<tr>
<td>Number of deaths</td>
<td>14</td>
<td>31</td>
<td>73</td>
<td>79</td>
</tr>
<tr>
<td>Median survival (months)</td>
<td>39.5</td>
<td>36</td>
<td>33</td>
<td>33</td>
</tr>
<tr>
<td>Hazard ratio (Confidence Interval)</td>
<td>0.86 (0.4-1.84)</td>
<td>0.67 (0.4-1.12)</td>
<td>0.56 (0.4-0.8)</td>
<td>0.6 (0.43-0.84)</td>
</tr>
<tr>
<td>p value of overall survival difference by stratified log rank test</td>
<td>0.71</td>
<td>0.13</td>
<td>0.0014</td>
<td>0.0025</td>
</tr>
</tbody>
</table>

![Graph showing p-values for different subgroups]
Figure 14: Determination of positive responders. Starting with subgroup 1 in the two arms, stratified log rank test compared the overall survivals. Then, the patients in subgroup 2 were added to the patients of subgroup 1 in both arms and stratified log rank test compared the overall survivals of this combination between the two arms. The process was continued to cover all the subgroups with negative martingale residuals. The $p$ value decreased every time that a subgroup was added to the previous one(s) except in the last step. Then, the combination of patients before the last step represents the positive responders.
Figure 15: Overall survival of positive responders in MD arm vs. that in MP arm. The median survival time was more than 11 months longer in positive responders in MD arm compared to the corresponding combination in MP arm [1333 days vs. 998 days, respectively; HR = 0.56 (0.4 - 0.8); and \( p = 0.0014 \)].
Negative Responders

The same method described for finding positive responders was applied to terminal subgroups with positive martingale residuals to find negative responders. Among the 6 subgroups with positive martingale residuals in MD arm, the one with the largest mean martingale residual (subgroup number 10 in figure 12) was selected and its overall survival was compared with the overall survival in the corresponding subgroup in the MP arm through stratified log rank test. The resulting $p$ value (0.0154) was plotted in figure 16. This subgroup involved 24 patients in MD arm and 27 patients in MP arm. The death rates at the end of study were 100% and 85%, respectively (HR = 2.1). Then, the two subgroups with the largest positive mean martingale residuals in the MD arm (terminal subgroups 9 and 10 in figure 12) were combined, which involved 33 patients in MD arm and 37 patients in MP arm. While only 6% of these patients in MD arm were alive at the end of study, about 27% ended alive in the combined subgroup of MP arm. The median survival time of those in MD arm was a year shorter than that in those of MP arm who had median survival of 32.6 months (HR = 1.8). The overall survival of this combination was compared with that of the corresponding combination in MP arm by stratified log rank test and the $p$ value (0.0087) was plotted in figure 16. The terminal subgroup ranked 8th in figure 12 was added to the last combination in both arms. This combination included 82 patients in MD arm and 96 patients in MP arm. While 90% of these patients were not alive in MD arm at the end of study, about 81% had similar fate in MP arm.
Median survival time in the recent combination in MD arm was 9 months shorter compared with the MP arm (HR = 1.56). The p value of stratified log rank statistics of two overall survivals was 0.00293, which was plotted in figure 16.

This method was repeated until no more subgroup with positive martingale residual was left. The search for negative responders should be stopped at the combination of subgroups, just before the stratified log rank p value stops decreasing. This occurred at the combination of 10th, 9th and 8th terminal subgroups because their stratified log rank test p value (0.00293) was the smallest among all the plotted p values in figure 16. Then, these three subgroups in MD arm were classified as negative responders, which included 82 patients with a median survival of 654 days (22 months). The median survival in the corresponding 96 patients in MP arm was 922 days (31 months). Now, it can be easily understood from the classification and regression tree analysis that, for example, multiple myeloma patients aged >75 have probably poor response to dexamethasone. Other predictive factors for negative responders were serum calcium concentration >2.6 mmol/L and Durie-Salmon stage 1. The overall survival of negative responders in MD arm was plotted against the overall survival of corresponding combination of 3 terminal subgroups in MP arm (figure 17). The hazard ratio was 1.56 (95% confidence interval 1.1-2.2; p = 0.00293).
<table>
<thead>
<tr>
<th>Subgroup</th>
<th>Subgroups</th>
<th>Subgroups</th>
<th>Subgroups</th>
<th>Subgroups</th>
<th>Subgroups</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
<td>10+9</td>
<td>10+9+8</td>
<td>10+9+8+7</td>
<td>10+9+8+7+6</td>
</tr>
<tr>
<td></td>
<td>MP MD</td>
<td>MP MD</td>
<td>MP MD</td>
<td>MP MD</td>
<td>MP MD</td>
</tr>
<tr>
<td>Number of Patients</td>
<td>27 24</td>
<td>37 33</td>
<td>96 82</td>
<td>114 101</td>
<td>121 116</td>
</tr>
<tr>
<td>Number of deaths</td>
<td>23 24</td>
<td>27 31</td>
<td>78 74</td>
<td>96 90</td>
<td>102 104</td>
</tr>
<tr>
<td>Median survival (months)</td>
<td>29.5 17</td>
<td>32.6 20.7</td>
<td>31 22</td>
<td>29 23</td>
<td>29.5 24</td>
</tr>
<tr>
<td>Hazard Ratio (Confidence Interval)</td>
<td>2.1 (1.1-3.8)</td>
<td>1.8 (1.06-3)</td>
<td>1.56 (1.1-2.2)</td>
<td>1.31 (0.98-1.75)</td>
<td>1.28 (0.97-1.68)</td>
</tr>
<tr>
<td>ρ value of overall survival difference by stratified log rank test</td>
<td>0.0154</td>
<td>0.0087</td>
<td>0.00293</td>
<td>0.0526</td>
<td>0.0513</td>
</tr>
</tbody>
</table>
Figure 16: Determination of negative responders. Starting with subgroup number 10 (SG10) in the two arms, stratified log rank test compared the overall survivals. Then, the patients in subgroup number 9 were added to the patients of subgroup number 10 in both arms and stratified log rank test compared the overall survivals of this combination between the two arms. The process was continued to cover all the subgroups with positive martingale residuals. The stratified log rank test $p$ value decreased every time that a subgroup was added to the previous combination until subgroup number 7, which increased the $p$ value. Although after adding subgroups 6 and 5, the $p$ value continuously decreased again, it never was smaller than the combination of terminal subgroups 10, 9 and 8. The latter represents the negative responders.
Figure 17: Overall survival of negative responders in MD arm vs. that in MP arm. The median survival time was 9 months shorter in negative responders in MD arm compared to the corresponding combination in MP arm [654 days vs. 922 days, respectively; HR = 1.56 (1.1 - 2.2) and $p = 0.00293$].
Non-Responders

The combination of subgroups number 4 to 7 of MD arm, which didn’t fit in positive or negative responders, composed the non-responders. This included 23% of patients in MD arm. Figure 18 illustrates their overall survival compared with that of the corresponding combination of subgroups (45 patients) in MP arm. About 11.5% and 15.5% of the patients were alive at the end of study in MD and MP arms, respectively. Their respective median survival times were 30 and 22.5 months \([\text{HR} = 0.98 (0.64 - 1.5); \text{and } p = 0.93]\). Although the difference in median survival here was about 7.5 months, the HR was almost 1.

![Figure 18: Overall survivals of non-responders (53 patients) in MD arm vs. that of corresponding subgroups in MP arm. The median survival times were 30 and 22.5 months in MD and MP arms, respectively. \([\text{HR} = 0.98 (0.64 - 1.5) \text{ and } p = 0.93]\).]
Interpretation of the Results

A multivariate analysis using a Cox regression model demonstrated a number of potential prognostic factors including Durie-Salmon stage, beta$_2$ microglobulin level, serum creatinine concentration, blood glucose levels and sex. Other factors found on univariate analysis to be potentially important,
such as hemoglobin and BM pathology, were not found to be statistically significant, likely because these factors were included in a composite risk model (e.g., hemoglobin is included in Salmon Durie staging).

Regression tree models split the input space into regions, which are described by part(s) of the predictive factors (as the input variables), and the size of the residuals (as the output variable) in each region. In the current study, martingale residuals in Cox proportional hazard model identified outliers, the tree was built on the MD arm (new treatment arm) and binary splitting was based on maximal difference of the residuals in the regions. We tried to find extreme regions of the tree model, i.e. terminal subgroups with patients who have large positive or large negative residuals. Judging the terminal subgroup combinations by their size, sign and mean of the residuals, and $p$ value of the stratified log rank statistic, we chose the subgroups of appropriate regions as the positive or negative responders. Positive responders were patients who derived benefit from the MD therapy in terms of overall survival compared to the patients with the same baseline characteristics randomized to the MP arm. Negative responders were patients under the MD treatment with survival times shorter than that of a similar group of patients (based on predictive factors) in the MP treatment arm. Non-responders were patients who revealed similar response to MP and MD; i.e., their survival time did not differ when receiving either MP or MD treatment.

The methodology of the current project aimed to find positive responders with the highest improvement in overall survival compared to the
other combinations of subgroups in the MD arm. Overall survival demonstrated an improvement of about 8% in subgroup 1 in the MD arm compared to the pair subgroup in the MP arm. When subgroup 2 was added to the subgroup 2, the improvement was around 12%. After adding subgroup 3 to the last combination, the improvement jumped to about 35%. Finally, the combination of all subgroups with negative MMR in the MD arm had about 27% improvement in OS compared to the paired combination in the MP arm. Since the combination of subgroups 1, 2 and 3 in the MD arm revealed the largest improvement in OS, its comparison with the paired combination in the MP arm showed the lowest \( p \) value. But, why the \( p \) value here was considerably smaller than the \( p \) value of the original study? The answer lies in two factor; first, the terminal subgroups in the MD arm were more homogenous than the original population and consequently they had larger difference with the corresponding subgroup(s) in the MP arm; and second, some part of the finding could be explained by chance secondary to data driven analysis (type I error).

Some of the final results of the current project are in agreement with the findings of other studies. For example, two different studies revealed that due to the higher frequency of severe pyogenic infections, hemorrhage, severe diabetes, and gastrointestinal and psychiatric complications following dexamethasone therapy in elderly patients (65 years and older), compared with MP treatment, dexamethasone-based regimens are not recommended
as first-line treatment in this subgroup of patients. On the other hand, a different study revealed that patients with hypercalcemia or pancytopenia, or those who require simultaneous radiotherapy for a pathologic fracture appeared most likely to benefit from dexamethasone therapy.

**Strengths**

In the current study, two major methods that were used to control selection bias included multivariable analysis and stratification. Fourteen covariates were used to evaluate prognosis. The relationships among these variables were complex. They could be related to one another (interaction) as well as to overall survival. The effect of one might be modified by the presence of other factors and the joint effects of two or more could be greater than the sum of their individual effects. Application of mathematical modelling techniques made it possible to consider the effects of many variables simultaneously and to adjust (control) for the effects of other variables to determine the independent effects of one. The strength of interactive tree-structured survival analysis discussed in the current project comes from the ability to extract statistically significant interactions between various covariates and find practical subgroups. This modelling made it possible to arrange variables in order of strength of their contribution and to predict the overall survival by calculating the combined effect of several variables acting together. Moreover, it has shown reliable results in the presence of moderate percentage censored data. Although Cox proportional hazard model is a very useful tool in quantifying
the effects of different covariates on survival time, and is more useful for prognostic classification, when there are many interaction terms between the covariates, the Cox proportional hazard regression model becomes more difficult to interpret for the purpose of predictive classification. Stratified analyses examined whether the effect of one variable was changed by the presence or absence of one or more other variables. Although the comparison of the crude overall survival between the two groups suggested no significant difference between the two methods of treatment, stratification made it possible to identify subgroups with significant difference in overall survival.

In the current study, cut off points derived mostly through literature review of studies that considered overall survival as the main outcome. In addition, the search for cut off points was done within the framework of a multiple regression model using martingale residuals smoothing plot to eliminate the potential influence of other predictive factors on the cutpoint. For all covariates, the cut off points were compared with those in the literature and in most cases, the established cut off points from the literature were selected to minimize data driven decisions.

**Limitations**

Currently, the standard treatment for multiple myeloma is autologous stem cell transplantation. Chemotherapeutic regimens are reserved for older patients with important co-morbidities \(^3\text{-}^6\). The presented method introduces a new
approach for evaluation of new treatment options in multiple myeloma clinical trials.

In order to prevent selection bias, beta\textsubscript{2} microglobulin, which had excessive missing information, was deleted from the multivariate CART analysis. Since beta\textsubscript{2} microglobulin is the most important prognostic factor in multiple myeloma\textsuperscript{7,99}, its exclusion was the major limitation of the current project. Missing information in beta\textsubscript{2} microglobulin leads to changes in martingale residuals and consequently causes confounding residual data because when beta\textsubscript{2} microglobulin was included in CART analysis, the missing category appeared in different splits and its interpretation was impossible.

The statistics applied in the examination of subgroups in the current thesis were secondary analyses because the objectives were not the main reason for the MY7 study. Guidelines for determining whether a finding in a subgroup analysis is real are summarized in table 6. Since a priori subgroup hypothesis was not designed in the protocol of the MY-7 study, formal statistical hypothesis testing was not possible through the current study. In addition, the original clinical trial (MY-7) was not specifically designed to have sufficient power within subgroups of interest. The posteriori-defined subgroups in the current study are suitable for formulating hypotheses. In other words, the results of the subgroup analysis study could be both informative and misleading. The decision regarding how much effort should be put into gathering more evidence depends on the potential benefits, risks and costs involved.
Table 6: Guidelines for deciding whether apparent differences in effects between subgroups are real \(^7^4\).

**From the study itself:**

<table>
<thead>
<tr>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is the magnitude of the observed differences clinically important?</td>
</tr>
<tr>
<td>How likely is the effect to have arisen by chance, taking into account:</td>
</tr>
<tr>
<td>- The number of subgroups examined?</td>
</tr>
<tr>
<td>- The magnitude of the ( p ) value?</td>
</tr>
<tr>
<td>Was a hypothesis that the effect would be observed made before its discovery (or was justification for the effect argued for after it was found)?</td>
</tr>
<tr>
<td>Was it one of a small number of hypotheses?</td>
</tr>
</tbody>
</table>

**From other information:**

<table>
<thead>
<tr>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Was the difference suggested by comparison within rather than between studies?</td>
</tr>
<tr>
<td>Has the effect been observed in other studies?</td>
</tr>
<tr>
<td>Is there indirect evidence that supports the existence of the effect?</td>
</tr>
</tbody>
</table>
Categorization of continuous covariates may induce loss of information, may decrease the goodness of fit of the predictive model and may limit the recursive partitioning process. Although categorization was made with the assumption that such a categorization is biologically plausible, in reality, more than one cutpoint may exist. Potential confounding may arise from categorization of continuous factors and using open-ended categories. It is obvious that the obtained cut off point(s) may differ across studies depending on which data or outcome-oriented approach is used and therefore, the results may not be comparable.

In the methods discussed in the current thesis, the percentage of censoring affects the sign of the martingale residuals and consequently affects responder analysis. The higher the percentage of censoring, the more frequent the negative martingale residuals. It has been shown that the responder identification method works well on data sets with 10% to 70% censoring. As is demonstrated in figure 14, the censoring percentage in positive responders and the corresponding combination of subgroups in MP group were around 42% and 21.5%, respectively. From figure 16 it can be seen that the percentage of censoring in negative responders was about 10% while in the corresponding combination of subgroups in the MP group it was about 19%. These percentages in non-responders and in the corresponding combination of subgroups in the MP group were 11.5% and 15.5%, respectively.
Disadvantages of CART

Classification and regression trees has its own weaknesses. Insignificant modification of sample, such as eliminating several observations, may lead to radical changes in decision tree such as increase or decrease of tree complexity and changes in splitting variables and values\(^\text{100}\). Therefore, the trees would be instable, which can negatively influence the results. Furthermore, CART splits only by one variable and it doesn’t use combinations of variables at each split\(^\text{100}\). Moreover, tree is optimal at each split but it may not be globally optimal. In CART, the relative importance of variables is unknown. In addition, there was some skepticism regarding tree methodologies in general, based on unrealistic claims and poor performance of earlier techniques. Thus, some statisticians have a generalized distrust of this approach.

Bradford-Hill Criteria of Causation

Table 7 provides a summary of established nine widely used criteria to determine the strength of the association between the overall survival and MD treatment in multiple myeloma patients of the current project. Since the subgroups were homogenous according to baseline factors but the results were significantly different among subgroups receiving MD or MP, this analysis generated a hypothesis that MD treatment had differential effects on subpopulations who received this therapy. Of course the role of chance and the effects of unobserved factors shouldn’t be ignored.
Table 7: Bradford-Hill Criteria of Causality.

<table>
<thead>
<tr>
<th>Bradford-Hill Criterion</th>
<th>Predictive Factors in Multiple Myeloma (MM) Treatment by MD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strength and association</td>
<td>The strength of association between MD and MM was shown in the current project through a data mining exercise rather than from a prospectively determined hypothesis. Further studies and confirmatory analyses are needed.</td>
</tr>
<tr>
<td>Consistency</td>
<td>There are some studies showing the effects of MD on selected subgroups of MM patients (e.g., elderly patients), but testing for predictive factors in these studies has not been reported.</td>
</tr>
<tr>
<td>Specificity</td>
<td>This analysis suggests that therapy-specific predictive factors could, in theory, exist.</td>
</tr>
<tr>
<td>Temporality</td>
<td>Only baseline features were assessed for prognosis; by definition these preceded the therapy tested.</td>
</tr>
<tr>
<td>Biological gradient (dose-response)</td>
<td>While this analysis tested steroid therapy of differing potencies, a range of dose-response relations was not sufficiently tested.</td>
</tr>
<tr>
<td>Biological plausibility</td>
<td>It is plausible that the more dose intense MD regimen has advantages in patients with more severe myeloma and fewer comorbidities, while it is less likely to benefit those with less severe disease and will be poorly tolerated by those with more comorbidities.</td>
</tr>
<tr>
<td>Biological coherence</td>
<td>Unclear, needs further studies</td>
</tr>
<tr>
<td>Experimental evidence</td>
<td><em>In vivo</em> evidence supports a role for MD in the treatment of MM.</td>
</tr>
<tr>
<td>Analogy</td>
<td>The concept of dose intensity is a strong hypothesis across many cancer subtypes.</td>
</tr>
</tbody>
</table>
Conclusions

The methods of interactive tree-structured survival analysis discussed in the current thesis can be applied quickly for analysis of multiple myeloma censored survival data to confirm priori hypotheses of important differences in survival of subgroups of multiple myeloma patients to be used in clinical practice. Moreover, the application of these methods will be helpful in discovering multifaceted interactions among different covariates that are only exhibited in certain subgroups and especially for finding previously unknown interactions in subgroups. In the current work, this led to some hypotheses regarding the differences in response to dexamethasone in various subgroups of multiple myeloma patients. Validation of the final results is required through further studies before results can be implemented in clinical practice. The future studies may consist of inclusion and exclusion criteria according to the baseline significant factors found in the current project. Multiple myeloma patients could be screened and/or stratified based on these criteria and then, they will be randomized to competing therapeutic options in order to find a clinical meaningful difference in overall survival.
References


26. Greipp PR, Leong T, Bennett JM, Gaillard JP, Klein B, Stewart JA et al. **Plasmablastic morphology--an independent prognostic factor with


29. Gutierrez NC, Castellanos MV, Martin ML, Mateos MV, Hernandez JM, Fernandez M et al. Prognostic and biological implications of genetic abnormalities in multiple myeloma undergoing autologous stem cell transplantation: t(4;14) is the most relevant adverse prognostic factor, whereas RB deletion as a unique abnormality is not associated with adverse prognosis. *Leukemia* 2007; 21: 143-150.


70. Freemantle N. Interpreting the results of secondary end points and subgroup analyses in clinical trials: should we lock the crazy aunt in the attic? *BMJ* 2001; 322: 989-991.


76. Moye LA. **P-value interpretation and alpha allocation in clinical trials.** *Ann Epidemiol* 1998; 8: 351-357.


Appendices
Appendix A

Table 8: Eastern Cooperative Oncology Group (ECOG) scaling.

<table>
<thead>
<tr>
<th>Performance Status</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Asymptomatic</td>
</tr>
<tr>
<td>1</td>
<td>Symptomatic, fully ambulatory</td>
</tr>
<tr>
<td>2</td>
<td>Symptomatic, in bed&lt;50% of the day</td>
</tr>
<tr>
<td>3</td>
<td>Symptomatic, in bed&gt;50% of the day</td>
</tr>
<tr>
<td>4</td>
<td>Bedridden</td>
</tr>
<tr>
<td>5</td>
<td>Death</td>
</tr>
</tbody>
</table>
Appendix B: Summary of SAS Codes

libname Analysis "C:\Documents and Settings\keyue\Desktop\Abbas\Thesis\Analysis";

/*--------------------------------------------------------+
| Sample Code of Univariate Analysis in MP group         |
+-------------------------------------------------------*/;

proc phreg data=Analysis.my777;
where MP_VS_MD=1;
model SURVIVAL * PT_STA_ALIVEZERO (0) = age / corrb covb risklimits;
baseline out=work._surv survival=_surviv_ upper=_sdfucl_lower=_sdflcl_/ strataid=on;
run;

proc phreg data=Analysis.my777;
where MP_VS_MD=1;
model SURVIVAL * PT_STA_ALIVEZERO (0) = stage2 stage3/
corrb covb risklimits;
baseline out=work._surv survival=_surviv_ upper=_sdfucl_lower=_sdflcl_/ strataid=on;
run;

/*--------------------------------------------------------+
| Sample Code of Univariate Analysis in MD group        |
+-------------------------------------------------------*/;

proc phreg data=Analysis.my777;
where MP_VS_MD=2;
model SURVIVAL * PT_STA_ALIVEZERO (0) = age / corrb covb risklimits;
baseline out=work._surv survival=_surviv_ upper=_sdfucl_lower=_sdflcl_/ strataid=on;
run;

proc phreg data=Analysis.my777;
where MP_VS_MD=2;
model SURVIVAL * PT_STA_ALIVEZERO (0) = stage2 stage3/
corrb covb risklimits;
baseline out=work._surv survival=_surviv_ upper=_sdfucl_lower=_sdflcl_/ strataid=on;
run;
/*--------------------------------------------------------+
| Examination of functional form of covariates using        |
| Martingale Residuals in MP Group                          |
+-------------------------------------------------------*/;

proc phreg data=Analysis.my777;
  where MP_VS_MD=1;
  model SURVIVAL * PT_STA_ALIVEZERO (0) = dummy;
  output out=resid_out1 resmart=mart_res1 /order=data;
  id id;
run;

proc phreg data=Analysis.my777;
  where MP_VS_MD=2;
  model SURVIVAL * PT_STA_ALIVEZERO (0) = dummy;
  output out=resid_out2 resmart=mart_res2 /order=data;
  id id;
run;

data data_res;
merge Analysis.my777 resid_out1 resid_out2;
  By id;
run;

/*--------------------------------------------------------+
| Comparing Martingale Residuals in Each Arm or in Two arms |
| by Different Ranges of Smoothness (sm)                   |
+-------------------------------------------------------*/;

proc gplot data=data_res;
  plot mart_res1*age / vref=0 haxis=axis2 vaxis=axis1
  overlay;
  symbol i=sm60s v=dot h=1.2 w=3;
  axis1 label = (h=2 r=0 a=90 f=swiss "Martingale Residuals")
  value = (h=2.0 f=swiss);
  axis2 label = (h=2 f=swiss) value = (h=2.0 f=swiss);
  rc=gdraw('line', 2, 0, 0);
  label mart_res='Residual';
  title "Martingale Residuals Plots vs. Linear Predictor for age in MP group when i=sm60s";
run;

proc gplot data=data_res;
  plot (mart_res1 mart_res2)*age / vref=0 haxis=axis2
  vaxis=axis1
overlay;
symbol i=sm60s v=dot h=1.2 w=3;
axis1 label = (h=2 r=0 a=90 f=swiss "Martingale Residuals")
value = (h=2.0 f=swiss);
axis2 label = (h=2 f=swiss) value = (h=2.0 f=swiss);
rc=gdraw('line', 2, 0, 0);
label mart_res='Residual';
title "Martingale Residuals Plots vs. Linear Predictor for age in Both groups when i=sm60s";
run;

proc gplot data=data_res;
plot mart_res1*age / vref=0 haxis=axis2 vaxis=axis1
overlay;
symbol i=sm50s v=dot h=1.2 w=3;
axis1 label = (h=2 r=0 a=90 f=swiss "Martingale Residuals")
value = (h=2.0 f=swiss);
axis2 label = (h=2 f=swiss) value = (h=2.0 f=swiss);
rc=gdraw('line', 2, 0, 0);
label mart_res='Residual';
title "Martingale Residuals Plots vs. Linear Predictor for age in MP group when i=sm50s";
run;

proc gplot data=data_res;
plot (mart_res1 mart_res2)*age / vref=0 haxis=axis2 vaxis=axis1
overlay;
symbol i=sm50s v=dot h=1.2 w=3;
axis1 label = (h=2 r=0 a=90 f=swiss "Martingale Residuals")
value = (h=2.0 f=swiss);
axis2 label = (h=2 f=swiss) value = (h=2.0 f=swiss);
rc=gdraw('line', 2, 0, 0);
label mart_res='Residual';
title "Martingale Residuals Plots vs. Linear Predictor for age in both groups when i=sm50s";
run;
proc phreg data=Analysis.my777;
  where MP_VS_MD=1;
  model SURVIVAL * PT_STA_ALIVEZERO (0) = age;
  output out=resid_out1 xbeta=xb resdev=dev
    resmart=mart_res1;
  title "Deviance Residuals Plots vs. Linear Predictor for age in MP group";
  run;

*** Sort data by BY variables ***;
proc sort data=Analysis.my777 out=work._stsrt_;
  where MP_VS_MD=1;
run;
  title "MULTIVARIATE ANALYSIS in both groups with all covariates as the continuous variable if possible";
  footnote;

*** Proportional Hazards Models ***;
  options pageno=1;
proc phreg data=work._stsrt_;
  by MP_VS_MD;
  model SURVIVAL * PT_STA_ALIVEZERO (0) = SKEL_ASS2 age
    SEX_M
    STAGE2 STAGE3 PLTS logscrea PERF_STA1 B2MIC_gt3pnt5 gluc
    CA
    B2MIC_gt5pnt5 B2MIC_mis WBC h_marrow ALB_grt35
    ALB_miss/
    selection = stepwise sle=0.1 sls=0.1;
    baseline out=work._surv survival=_surviv_ upper=_sdfucl_ lower=_sdflcl_;
run; quit;
  goptions reset=all device=WIN;

** Survival plot **;
proc sort data=work._surv out=work._surv;
  by MP_VS_MD;
run;
title;
footnote;
goptions ftext=SWISS ctext=BLACK htext=1 cells;
proc gplot data=work._surv;
  by MP_VS_MD;
  label survival = 'Survival Time';
  axis2 minor=none major=(number=6)
    label=(angle=90 'Survival Distribution Function');
  symbol1 i=stepj c=BLUE l=1 width=1;
  plot _surviv_ * survival=1 / description="SDF of survival"
    frame cframe=CXF7E1C2 caxis=BLACK
    vaxis=axis2 hminor=0 name='SDF';
run;
quit;
goptions ftext= ctext= htext= reset=symbol;

*---------------------------------------------------------*
|Model Fit in MP group                                    |
*---------------------------------------------------------*

proc phreg data=work._stsrt_;
  where MP_VS_MD=1;
  model SURVIVAL * PT_STA_ALIVEZERO (0) = SKEL_ASS2 age
    SEX_M
    STAGE2 STAGE3 PLTS logscrea PERF_STA1 B2MIC_gt3pnt5
    B2MIC_gt5pnt5 B2MIC_mis WBC HGB b_marrow ALB_grt35
    ALB_miss
    gluc CA;
  baseline out=work._surv survival=_surviv_ upper=_sdfucl_
    lower=_sdflcl_
  title "Model Fit in MP group";
run;

*-----------------------------------------------------------------
|Testing Proportional Hazards Using Schoenfeld Residuals          |
*-----------------------------------------------------------------

proc phreg data=Analysis.my777;
  where MP_VS_MD=1;
  model SURVIVAL * PT_STA_ALIVEZERO (0) = SKEL_ASS2 age
    SEX_M STAGE2 STAGE3 PLTS logscrea PERF_STA1 B2MIC_gt3pnt5
    CA gluc B2MIC_gt5pnt5 B2MIC_mis WBC HGB b_marrow ALB_grt35
    ALB_miss/ ties=efron;
output out=schoenb ressch= schhgb schwbc schage schplts schca schgluc schU_MPROT schscrea schb2mic schb_marrow schalb;
run;

proc gplot data=schoenb;
plot schhgb*SURVIVAL schwbc*SURVIVAL schage*SURVIVAL schplts*SURVIVAL schca*SURVIVAL schgluc*SURVIVAL schb2mic*SURVIVAL schb_marrow*SURVIVAL schalb*SURVIVAL/vref=0 haxis=axis2 vaxis=axis1;
symbol value=dot i=sm60s h=1.2 w=3;
axis1 label = (h=2 r=0 a=90 f=swiss) value = (h=2.0 f=swiss);
axis2 label = (h=2 f=swiss) value = (h=2.0 f=swiss);
title "Schoenfeld Residuals";
run;

/*--------------------------------------------------------+
|Subgroups Definitions in the Two Groups                 |
+-------------------------------------------------------*/;

DATA Analysis.my777;
length node $10;
set Analysis.my777;

IF (MP_VS_MD = 1) and (AGE_gt76 = 2) THEN node = 'AGEMP';
IF (MP_VS_MD = 2) and (AGE_gt76 = 2) THEN node = 'AGEMD';

IF (MP_VS_MD = 1) and (AGE_gt76 = 1) and (CA_gt2pnt6 ne 1) THEN node = 'AGECAMP';
IF (MP_VS_MD = 2) and (AGE_gt76 = 1) and (CA_gt2pnt6 ne 1) THEN node = 'AGECAMD';

IF (MP_VS_MD = 1) and (AGE_gt76 = 1) and (CA_gt2pnt6 = 1) and (STAGE = 1) then node = 'CASTMP';
IF (MP_VS_MD = 2) and (AGE_gt76 = 1) and (CA_gt2pnt6 = 1) and (STAGE = 1) then node = 'CASTMD';

IF (MP_VS_MD = 1) and (AGE_gt76 = 1) and (CA_gt2pnt6 = 1) and (STAGE ne 1) and (PERF_STA1 = 0) and (HGB_gt100 = 2) then node = 'PEHGMP';
IF (MP_VS_MD = 2) and (AGE_gt76 = 1) and (CA_gt2pnt6 = 1) and (STAGE ne 1) and (PERF_STA1 = 0) and (HGB_gt100 = 2) then node = 'PEHGMD';
IF (MP_VS_MD = 1) and (AGE_gt76 = 1) and (CA_gt2pnt6 = 1) and (STAGE ne 1) and (PERF_STA1 = 0) and (HGB_gt100 ne 2) and (WBC_gt4 ne 2) then node = 'HGWBCNMP';

IF (MP_VS_MD = 2) and (AGE_gt76 = 1) and (CA_gt2pnt6 = 1) and (STAGE ne 1) and (PERF_STA1 = 0) and (HGB_gt100 ne 2) and (WBC_gt4 ne 2) then node = 'HGWBCNMD';

IF (MP_VS_MD = 1) and (AGE_gt76 = 1) and (CA_gt2pnt6 = 1) and (STAGE ne 1) and (PERF_STA1 = 0) and (HGB_gt100 ne 2) and (WBC_gt4 = 2) and (SKEL_ASS2 = 0) then node = 'HGWBCPMP';

IF (MP_VS_MD = 2) and (AGE_gt76 = 1) and (CA_gt2pnt6 = 1) and (STAGE ne 1) and (PERF_STA1 = 0) and (HGB_gt100 ne 2) and (WBC_gt4 = 2) and (SKEL_ASS2 = 0) then node = 'HGWBCPMD';

IF (MP_VS_MD = 1) and (AGE_gt76 = 1) and (CA_gt2pnt6 = 1) and (STAGE ne 1) and (PERF_STA1 = 0) and (HGB_gt100 ne 2) and (WBC_gt4 = 2) and (SKEL_ASS2 = 1) then node = 'WBCSAMP';

IF (MP_VS_MD = 2) and (AGE_gt76 = 1) and (CA_gt2pnt6 = 1) and (STAGE ne 1) and (PERF_STA1 = 0) and (HGB_gt100 ne 2) and (WBC_gt4 = 2) and (SKEL_ASS2 = 1) then node = 'WBCSAMD';

IF (MP_VS_MD = 1) and (AGE_gt76 = 1) and (CA_gt2pnt6 = 1) and (STAGE ne 1) and (PERF_STA1 ne 0) and (SKEL_ASS2 ne 1) then node = 'PESAMP';

IF (MP_VS_MD = 2) and (AGE_gt76 = 1) and (CA_gt2pnt6 = 1) and (STAGE ne 1) and (PERF_STA1 ne 0) and (SKEL_ASS2 ne 1) then node = 'PESAMD';

IF (MP_VS_MD = 1) and (AGE_gt76 = 1) and (CA_gt2pnt6 = 1) and (STAGE ne 1) and (PERF_STA1 ne 0) and (SKEL_ASS2 ne 1) and (SEX_M=1) then node = 'SASXPMP';

IF (MP_VS_MD = 2) and (AGE_gt76 = 1) and (CA_gt2pnt6 = 1) and (STAGE ne 1) and (PERF_STA1 ne 0) and (SKEL_ASS2 ne 1) and (SEX_M=1) then node = 'SASXPMMD';

IF (MP_VS_MD = 1) and (AGE_gt76 = 1) and (CA_gt2pnt6 = 1) and (STAGE ne 1) and (PERF_STA1 ne 0) and (SKEL_ASS2 ne 1) and (SEX_M ne 1) then node = 'SASXNMP';

IF (MP_VS_MD = 2) and (AGE_gt76 = 1) and (CA_gt2pnt6 = 1) and (STAGE ne 1) and (PERF_STA1 ne 0) and (SKEL_ASS2 ne 1) and (SEX_M ne 1) then node = 'SASXNMD';

run;
proc freq data=Analysis.my777; tables node; run;

/*--------------------------------------------------------+
|Survival comparison of similar subgroups in the two groups
+-------------------------------------------------------*/;

proc lifetest data=Analysis.my777(where=(node in ('HGWBCPMP','HGWBCPMD')));
time SURVIVAL * PT_STA_ALIVEZERO (0);
strata node;
run;

proc phreg data=Analysis.my777;
model SURVIVAL * PT_STA_ALIVEZERO (0) = MP_VS_MD/rl;
where node in ('HGWBCPMP','HGWBCPMD');
run;

proc lifetest data=Analysis.my777(where=(node in ('SASXPMD','SASXPMP')));
time SURVIVAL * PT_STA_ALIVEZERO (0);
strata node;
run;

proc phreg data=Analysis.my777;
model SURVIVAL * PT_STA_ALIVEZERO (0) = MP_VS_MD/rl;
where node in ('SASXPMD','SASXPMP');
run;

proc lifetest data=Analysis.my777(where=(node in ('PEHGMP','PEHGMD')));
time SURVIVAL * PT_STA_ALIVEZERO (0);
strata node;
run;

proc phreg data=Analysis.my777;
model SURVIVAL * PT_STA_ALIVEZERO (0) = MP_VS_MD/rl;
where node in ('PEHGMP','PEHGMD');
run;

proc lifetest data=Analysis.my777(where=(node in ('WBCSAMP','WBCSAMD')));
time SURVIVAL * PT_STA_ALIVEZERO (0);
strata node;
run;

proc phreg data=Analysis.my777;
model SURVIVAL * PT_STA_ALIVEZERO (0) = MP_VS_MD/rl;
where node in ('WBCSAMP', 'WBCSAMD');
run;

proc lifetest data=Analysis.my777(where=(node in ('SASXNMD', 'SASXNMP')));
time SURVIVAL * PT_STA_ALIVEZERO (0);
strata node;
run;

proc phreg data=Analysis.my777;
model SURVIVAL * PT_STA_ALIVEZERO (0) = MP_VS_MD/rl;
where node in ('SASXNMD', 'SASXNMP');
run;

proc lifetest data=Analysis.my777(where=(node in ('HGWBCNMD', 'HGWBCNMP')));
time SURVIVAL * PT_STA_ALIVEZERO (0);
strata node;
run;

proc phreg data=Analysis.my777;
model SURVIVAL * PT_STA_ALIVEZERO (0) = MP_VS_MD/rl;
where node in ('HGWBCNMD', 'HGWBCNMP');
run;

proc lifetest data=Analysis.my777(where=(node in ('PESAMP', 'PESAMD')));
time SURVIVAL * PT_STA_ALIVEZERO (0);
strata node;
run;

proc phreg data=Analysis.my777;
model SURVIVAL * PT_STA_ALIVEZERO (0) = MP_VS_MD/rl;
where node in ('PESAMP', 'PESAMD');
run;

proc lifetest data=Analysis.my777(where=(node in ('AGEMD', 'AGEMP')));
time SURVIVAL * PT_STA_ALIVEZERO (0);
strata node;
run;

proc phreg data=Analysis.my777;
model SURVIVAL * PT_STA_ALIVEZERO (0) = MP_VS_MD/rl;
where node in ('AGEMD', 'AGEMP');
run;
```sas
proc lifetest data=Analysis.my777(where=(node in ('CASTMD','CASTMP')));
time SURVIVAL * PT_STA_ALIVEZERO (0);
strata node;
run;

proc phreg data=Analysis.my777;
model SURVIVAL * PT_STA_ALIVEZERO (0) = MP_VS_MD/rl;
where node in ('CASTMD','CASTMP');
run;

proc lifetest data=Analysis.my777(where=(node in ('AGECAMP','AGECAMD')));
time SURVIVAL * PT_STA_ALIVEZERO (0);
strata node;
run;

proc phreg data=Analysis.my777;
model SURVIVAL * PT_STA_ALIVEZERO (0) = MP_VS_MD/rl;
where node in ('AGECAMP','AGECAMD');
run;

proc summary data=Analysis.my777;
  var SURVIVAL;
  class node;
  output out=WORK.SUMM(drop = _TYPE_ _FREQ_)
    MEAN =
    STD =
    MEDIAN =
    / autoname;
run;

Proc print;
run;

/*--------------------------------------------------------+
|Survival Comparison of Responder Subgroups in the Two     |
|Groups               |                                             |
+-------------------------------------------------------*/;
DATA Analysis.my777;
set Analysis.my777 ;
IF (node='PEHGMP') or (node='HGWBCPMP') or (node='SASXPMP') or (node='WBCSAMP') then nodeF='PosResMP';
IF (node='PEHGMD') or (node='HGWBCPMD') or (node='SASXPMD') or (node='WBCSAMD') then nodeF='PosResMD';
```
run;
proc freq data=Analysis.my777; tables nodeF; run;

proc lifetest data=Analysis.my777(where=(nodeF in ('PosResMP','PosResMD')));
time SURVIVAL * PT_STA_ALIVEZERO (0);
strata nodeF;
run;

proc phreg data=Analysis.my777;
model SURVIVAL * PT_STA_ALIVEZERO (0) = MP_VS_MD/rl;
where nodeF in ('PosResMP','PosResMD');
run;

DATA Analysis.my777;
set Analysis.my777;
IF (node='AGEMP') or (node='AGECAMP') or (node='CASTMP') or (node='HGWBCNMP') or (node='PESAMP') or (node= 'SASXNMP')
then nodeF='NegResMP';
IF (node='AGEMD') or (node='AGECAMD') or (node='CASTMD') or (node='HGWBCNMD') or (node='PESAMD') or (node= 'SASXNMD')
then nodeF='NegResMD';
run;
proc freq data=Analysis.my777; tables nodeF; run;

proc lifetest data=Analysis.my777(where=(nodeF in ('NegResMP','NegResMD')));
time SURVIVAL * PT_STA_ALIVEZERO (0);
strata nodeF;
run;

proc phreg data=Analysis.my777;
model SURVIVAL * PT_STA_ALIVEZERO (0) = MP_VS_MD/rl;
where nodeF in ('NegResMP','NegResMD');
run;

proc summary data=Analysis.my777;
  var SURVIVAL;
class nodeF;
  output out=WORK.SUMM(drop = _TYPE_ _FREQ_) MEAN = STD = MEDIAN = / autoname;
run;
Proc print;
run;
DATA Analysis.my777;
set Analysis.my777(DROP=node12);

IF (node='HGWBCPMP') or (node='SASXPMP') then
node12='PosResMP';
IF (node='HGWBCPMD') or (node='SASXPMD') then
node12='PosResMD';

IF (node='AGECAMP') or (node='CASTMP') then
node12='NegResMP';
IF (node='AGECAMD') or (node='CASTMD') then
node12='NegResMD';

run;

proc freq data=Analysis.my777; tables node12 ; run;
proc lifetest data=Analysis.my777(where=(node12 in ('PosResMP','PosResMD')));
time SURVIVAL * PT_STA_ALIVEZERO (0);
strata node12;
run;
proc phreg data=Analysis.my777;
model SURVIVAL * PT_STA_ALIVEZERO (0) = MP_VS_MD/rl;
where node12 in ('PosResMP','PosResMD');
run;

proc lifetest data=Analysis.my777(where=(node12 in ('NegResMP','NegResMD')));
time SURVIVAL * PT_STA_ALIVEZERO (0);
strata node12;
run;
proc phreg data=Analysis.my777;
model SURVIVAL * PT_STA_ALIVEZERO (0) = MP_VS_MD/rl;
where node12 in ('NegResMP','NegResMD');
run;

proc summary data=Analysis.my777;
var SURVIVAL;
class node12;
output out=WORK.SUMM(drop = _TYPE_ _FREQ_) MEAN = STD = MEDIAN = / autoname;
run;
Proc print;
run;
DATA Analysis.my777;
set Analysis.my777;

IF (node='PEHGMP') or (node='HGWBCPMP') or (node='SASXPPMP')
then node123='PosResMP';
IF (node='PEHGMD') or (node='HGWBCPMD') or (node='SASXPMMD')
then node123='PosResMD';

IF (node='AGECAMMP') or (node='CASTMP') or (node='AGEEMP')
then node123='NegResMP';
IF (node='AGECAMD') or (node='CASTMD') or (node='AGEEMD')
then node123='NegResMD';

run;

proc freq data=Analysis.my777; tables node123 ; run;

proc lifetest data=Analysis.my777(where=(node123 in ('PosResMP','PosResMD')));
time SURVIVAL * PT_STA_ALIVEZERO (0);
strata node123;
run;
proc phreg data=Analysis.my777;
model SURVIVAL * PT_STA_ALIVEZERO (0) = MP_VS_MD/rl;
where node123 in ('PosResMP','PosResMD');
run;

proc lifetest data=Analysis.my777(where=(node123 in ('NegResMP','NegResMD')));
time SURVIVAL * PT_STA_ALIVEZERO (0);
strata node123;
run;
proc phreg data=Analysis.my777;
model SURVIVAL * PT_STA_ALIVEZERO (0) = MP_VS_MD/rl;
where node123 in ('NegResMP','NegResMD');
run;

proc summary data=Analysis.my777;
var SURVIVAL;
class node123;
output out=WORK.SUMM(drop = _TYPE_ _FREQ_)
    MEAN =
    STD =
    MEDIAN =
    / autoname;
run;
**Proc print;**
**run;**

**DATA** Analysis.my777;
**set** Analysis.my777;

IF (node='AGECAMP') or (node='CASTMP') or (node='AGEMP') or (node='PESAMP') then node1234 = 'NegResMP';
IF (node='AGECAMD') or (node='CASTMD') or (node='AGEMD') or (node='PESAMD') then node1234 = 'NegResMD';

**run;**

**proc freq data=Analysis.my777; tables node1234 ; run;**

**proc lifetest data=Analysis.my777(where=(node1234 in ('NegResMP','NegResMD'))); time SURVIVAL * PT_STA_ALIVEZERO (0); strata node1234;**
**run;**

**proc phreg data=Analysis.my777; model SURVIVAL * PT_STA_ALIVEZERO (0) = MP_VS_MD/rl; where node1234 in ('NegResMP','NegResMD');**
**run;**

**proc summary data=Analysis.my777; var SURVIVAL; class node1234; output out=WORK.SUMM(drop = _TYPE_ _FREQ_) MEAN = STD = MEDIAN = / autoname;**
**run;**

**Proc print;**
**run;**

**DATA** Analysis.my777;
**set** Analysis.my777;

IF (node='AGECAMP') or (node='CASTMP') or (node='AGEMP') or (node='PESAMP') or (node='HGWBCNMP') then node12345 = 'NegResMP';
IF (node='AGECAMD') or (node='CASTMD') or (node='AGEMD') or (node='PESAMD') or (node='HGWBCNMD') then node12345 = 'NegResMD';

**run;**

**proc freq data=Analysis.my777; tables node12345 ; run;**

107
proc lifetest data=Analysis.my777(where=(node12345 in ('NegResMP','NegResMD')));
time SURVIVAL * PT_STA_ALIVEZERO (0);
strata node12345;
run;
proc phreg data=Analysis.my777;
model SURVIVAL * PT_STA_ALIVEZERO (0) = MP_VS_MD/rl;
where node12345 in ('NegResMP','NegResMD');
run;

proc summary data=Analysis.my777;
var SURVIVAL;
class node12345;
output out=WORK.SUMM(drop = _TYPE_ _FREQ_) MEAN = STD = MEDIAN = / autoname;
run;
Proc print;
run;

DATA Analysis.my777;
set Analysis.my777;
IF (node='HGWBCNMP') or (node='PESAMP') or (node='SASXNMP') or (node='WBCSAMP') then nodeN='NonResMP';
IF (node='HGWBCNMD') or (node='PESAMD') or (node='SASXNMD') or (node='WBCSAMD') then nodeN='NonResMD';
run;
proc freq data=Analysis.my777; tables nodeN; run;

proc lifetest data=Analysis.my777(where=(nodeN in ('NonResMP','NonResMD')));
time SURVIVAL * PT_STA_ALIVEZERO (0);
strata nodeN;
run;

proc phreg data=Analysis.my777;
model SURVIVAL * PT_STA_ALIVEZERO (0) = MP_VS_MD/rl;
where nodeN in ('NonResMP','NonResMD');
run;

proc summary data=Analysis.my777;
var SURVIVAL;
class nodeN;
output out=WORK.SUMM(drop = _TYPE_ _FREQ_)
MEAN =
STD =
MEDIAN =
/ autoname;
run;
Proc print;
run;
Appendix C: Summary of R Codes

```r
library(foreign)
library(survival)
library(MASS)
MY7.dat <- read.table("C:\Documents and Settings\keyue\Desktop\Abbas\Thesis\Analysis\R\MY77.txt", header=T, sep=" 	")
names(MY7.dat)
library(survival)
library(rpart)
library(MASS)

################################################################
# Multivariate Cox analysis
################################################################
options(contrasts=c("contr.treatment", "contr.treatment"))

MY7.cox.cont <- coxph(Surv(survival, PT_STA_ALIVEZERO) ~ factor(SEX_M) + factor(STAGE) + factor(B2MIC_4codes) + factor(SKEL_ASS2) + factor(ALB_gt35) + factor(PERF_STA1) + factor(loggluc_gt1pnt72) + factor(CA_gt2pnt6) + logscrea + AGE + PLTS + WBC + HGB + b_marrow, data = MY7.dat, subset=MP_VS_MD==1)

summary(MY7.cox.cont)

MY7.pl.cox2 <- stepAIC(MY7.cox.cont)
MY7.pl.cox2 <- stepAIC(MY7.cox.cont, direction = c("both"))
MY7.pl.cox2

MY7.trt <- MY7.dat[MY7.dat$MP_VS_MD=="2",]
MY7.cont <- MY7.dat[MY7.dat$MP_VS_MD=="1",]
dim(MY7.trt)

fit.haz <- basehaz(MY7.pl.cox2)
fit.haz1 <- approx(fit.haz$time,fit.haz$haz, xout= MY7.trt$survival, yleft=0, rule=2, method="constant",f=0)
par(mfrow=c(1,1))
plot(fit.haz1, main="Baseline Cumulative Hazard", xlab="Time in Months", ylab="Cumulative Hazard")
lines(fit.haz$time,fit.haz$haz)
lines(fit.haz$time,fit.haz$haz, type="s")
resid.fit.placebo <- residuals(MY7.pl.cox2, type="martingale")
par(mfrow=c(1,1))
plot(residuals(MY7.pl.cox2))

fit.linpred <- predict(MY7.pl.cox2, newdata=MY7.trt, type="lp")
resid.fit <- MY7.trt$PT_STA_ALIVEZERO- exp(fit.linpred)*fit.haz1$y
```

110
par(mfrow=c(1,2))
plot(resid.fit, main="MD Group Residuals Based on the MP Model")
plot(resid.fit.placebo, main="MP Group Residuals")

# Building the Tree

MY7.trt$resid <- resid.fit

tree.treat3 <- rpart(resid~ factor(SEX_M) + factor(STAGE) + factor(WBC_gt4) + factor(screa_gt175) + factor(AGE_gt76) + factor(ALB_gt35) + factor(b_marrow_gt40) + factor(HGB_gt100) + factor(PLTS_gt150) + factor(SKEL_ASS2) + factor(PERF_STA1) + factor(CA_gt2pnt6), data = MY7.trt)
printcp(tree.treat3)
par(mfrow=c(1,1))
plot(tree.treat3, uniform=T, main="Tree Predictive Model Based on Martingale Residuals")
text(tree.treat3, digits=3, pretty=3, all=TRUE, use.n=TRUE)

# Definition of Subgroups in MD arm

node.1  <-  MY7.trt$AGE_gt76==1& MY7.trt$CA_gt2pnt6==1& MY7.trt$STAGE!=1 & MY7.trt$PERF_STA1==0 & MY7.trt$HGB_gt100!=2 & MY7.trt$WBC_gt4==2& MY7.trt$ SKEL_ASS2==0
sum(node.1)

node.2  <-  MY7.trt$AGE_gt76==1& MY7.trt$CA_gt2pnt6==1& MY7.trt$STAGE!=1 & MY7.trt$PERF_STA1!=0 & MY7.trt$ SKEL_ASS2==1& MY7.trt$SEX_M==1
sum(node.2)

node.3  <-  MY7.trt$AGE_gt76==1& MY7.trt$CA_gt2pnt6==1& MY7.trt$STAGE!=1 & MY7.trt$PERF_STA1==0 & MY7.trt$HGB_gt100==2 & MY7.trt$SKEL_ASS2==0
sum(node.3)

node.4  <-  MY7.trt$AGE_gt76==1& MY7.trt$CA_gt2pnt6==1& MY7.trt$STAGE!=1 & MY7.trt$PERF_STA1==0 & MY7.trt$HGB_gt100!=2 & MY7.trt$WBC_gt4==2& MY7.trt$ SKEL_ASS2!=0
sum(node.4)
node.5 <- MY7.trt$AGE_gt76==1 & MY7.trt$CA_gt2pnt6==1 &
MY7.trt$STAGE!=1 & MY7.trt$PERF_STA1!=0 & MY7.trt$SKEL_ASS2==1 &
MY7.trt$SEX_M!=1
sum(node.5)

node.6 <- MY7.trt$AGE_gt76==1 & MY7.trt$CA_gt2pnt6==1 &
MY7.trt$STAGE!=1 & MY7.trt$PERF_STA1==0 & MY7.trt$HGB_gt100!=2 &
MY7.trt$WBC_gt4!=2
sum(node.6)

node.7 <- MY7.trt$AGE_gt76==1 & MY7.trt$CA_gt2pnt6==1 &
MY7.trt$STAGE!=1 & MY7.trt$PERF_STA1!=0 & MY7.trt$SKEL_ASS2!=1
sum(node.7)

node.8 <- MY7.trt$AGE_gt76!=1
sum(node.8)

node.9 <- MY7.trt$AGE_gt76==1 & MY7.trt$CA_gt2pnt6==1 &
MY7.trt$STAGE==1
sum(node.9)

node.10 <- MY7.trt$AGE_gt76==1 & MY7.trt$CA_gt2pnt6!=1
sum(node.10)

#########################################################################
# Applying the Subgroups Criteria Separating for the all Data Set     #
#########################################################################

node.1 <- MY7.dat$AGE_gt76==1 & MY7.dat$CA_gt2pnt6==1 &
MY7.dat$STAGE!=1 & MY7.dat$PERF_STA1==0 & MY7.dat$HGB_gt100!=2 &
MY7.dat$WBC_gt4==2 & MY7.dat$SKEL_ASS2==0
sum(node.1)

node.2 <- MY7.dat$AGE_gt76==1 & MY7.dat$CA_gt2pnt6==1 &
MY7.dat$STAGE!=1 & MY7.dat$PERF_STA1!=0 & MY7.dat$SKEL_ASS2==1 &
MY7.dat$SEX_M==1
sum(node.2)

node.3 <- MY7.dat$AGE_gt76==1 & MY7.dat$CA_gt2pnt6==1 &
MY7.dat$STAGE!=1 & MY7.dat$PERF_STA1==0 & MY7.dat$HGB_gt100==2
sum(node.3)

node.4 <- MY7.dat$AGE_gt76==1 & MY7.dat$CA_gt2pnt6==1 &
MY7.dat$STAGE!=1 & MY7.dat$PERF_STA1==0 & MY7.dat$HGB_gt100!=2 &
MY7.dat$WBC_gt4==2 & MY7.dat$SKEL_ASS2!=0
sum(node.4)
node.5 <- MY7.dat$AGE_gt76==1 & MY7.dat$CA_gt2pnt6==1 & MY7.dat$STAGE!=1 & MY7.dat$PERF_STA1!=0 & MY7.dat$SKEL_ASS2==1 & MY7.dat$SEX_M!=1
sum(node.5)

node.6 <- MY7.dat$AGE_gt76==1 & MY7.dat$CA_gt2pnt6==1 & MY7.dat$STAGE!=1 & MY7.dat$PERF_STA1==0 & MY7.dat$HGB_gt100!=2 & MY7.dat$WBC_gt4!=2
sum(node.6)

node.7 <- MY7.dat$AGE_gt76==1 & MY7.dat$CA_gt2pnt6==1 & MY7.dat$STAGE!=1 & MY7.dat$PERF_STA1!=0 & MY7.dat$SKEL_ASS2==1
sum(node.7)

node.8 <- MY7.dat$AGE_gt76!=1
sum(node.8)

node.9 <- MY7.dat$AGE_gt76==1 & MY7.dat$CA_gt2pnt6==1 & MY7.dat$STAGE==1
sum(node.9)

node.10 <- MY7.dat$AGE_gt76==1 & MY7.dat$CA_gt2pnt6!=1
sum(node.10)

# Defining the Corresponding Subgroups
##
nodes <- node.1*1 + node.2*2 + node.3*3 + node.4*4 + node.5*5 + node.6*6 + node.7*7 + node.8*8 + node.9*9
MY7.dat$nodes <- factor(nodes)
survdiff( Surv(survival, PT_STA_ALIVEZERO) ~ MP_VS_MD, data = MY7.dat[node.1,])
s1 <- 0.706
summary(coxph( Surv(survival, PT_STA_ALIVEZERO) ~ MP_VS_MD, data = MY7.dat[node.1,]) )
survdiff( Surv(survival, PT_STA_ALIVEZERO) ~ MP_VS_MD, data = MY7.dat[node.1 | node.2,])
s2 <- 0.126
summary(coxph( Surv(survival, PT_STA_ALIVEZERO) ~ MP_VS_MD+strata(nodes), data = MY7.dat[node.1|node.2, ] ))
d2 <- 0.143

survdiff( Surv(survival, PT_STA_ALIVEZERO) ~ MP_VS_MD, data = MY7.dat[node.1 | node.2 | node.3,])
s3 <- 0.00122

summary(coxph( Surv(survival, PT_STA_ALIVEZERO) ~ MP_VS_MD+strata(nodes), data = MY7.dat[node.1|node.2|node.3, ] ))
d3 <- 0.00127

survdiff( Surv(survival, PT_STA_ALIVEZERO) ~ MP_VS_MD, data = MY7.dat[node.1 | node.2 | node.3 | node.4,])
d4 <- 0.00229

# Determine the Positive Responders

plot(c(s1,d2,d3,d4), xlab="Subgroup 1                                           Subgroups 1 & 2                                        Subgroups 1, 2 & 3          Subgroups 1, 2, 3 & 4
1 & 2 Subgroups 1, 2 & 3 Subgroups 1, 2, 3 & 4
", ylab="p value", type="o", lty=1,pch=19, main=" ")

# Detection of Negative Responders #

summary(coxph( Surv(survival, PT_STA_ALIVEZERO) ~ MP_VS_MD+strata(nodes), data = MY7.dat[node.10,]))
d5 <- 0.0154

summary(coxph( Surv(survival, PT_STA_ALIVEZERO) ~ MP_VS_MD+strata(nodes), data = MY7.dat[node.10 | node.9,]))
d6 <- 0.0087

summary(coxph( Surv(survival, PT_STA_ALIVEZERO) ~ MP_VS_MD+strata(nodes), data = MY7.dat[node.10 | node.9 | node.8,]))
d7 <- 0.00293

summary(coxph( Surv(survival, PT_STA_ALIVEZERO) ~ MP_VS_MD+strata(nodes), data = MY7.dat[node.10 | node.9 | node.8 | node.7,]))
d8 <- 0.0526

summary(coxph( Surv(survival, PT_STA_ALIVEZERO) ~ MP_VS_MD+strata(nodes), data = MY7.dat[node.10 | node.9 | node.8 | node.7 | node.6,]))
d9 <- 0.0513
summary(coxph( Surv(survival, PT_STA_ALIVEZERO) ~ MP_VS_MD + strata(nodes), data = MY7.dat[node.10| node.9| node.8| node.7| node.6| node.5,])))
d10 <- 0.0383

# Determine the negative Responders
plot(c(5,6,7,8,9,10), c(d5,d6,d7,d8,d9,d10), xlab="Subgroup 10
Subgroups 10,9
Subgroups 10,9,8
Subgroups 10,9,8,7
Subgroups 10,9,8,7,6
Subgroups 10,9,8,7,6,5", ylab="p value", type="o", lty=1,pch=19, main="")

library(survival)
options(contrasts=c("contr.treatment","contr.treatment"))
a1 <- survfit( Surv(survival, PT_STA_ALIVEZERO) ~ factor(MP_VS_MD), data = MY7.dat[node.1|node.2|node.3,])
b1 <- coxph( Surv(survival, PT_STA_ALIVEZERO) ~ factor(MP_VS_MD), data = MY7.dat[node.1|node.2|node.3,])

plot( survfit( Surv(survival, PT_STA_ALIVEZERO) ~ factor(MP_VS_MD), data = MY7.dat[node.1|node.2|node.3,]),xlim=c(0,3000), xlab="Days", ylab="Probability", main="", lty=1,lwd=3, col=2:3)

legend(1,0.2, c("MP","MD"), lty=1,lwd=3, col=2:3)
legend(6,0.8, print(c("", round(summary(b1)$sctest[3],3))),bty="" )
legend(6,0.7, print(c("", round(summary(b1)$coef[2],3))),bty="" )
summary(b1)

library(survival)
options(contrasts=c("contr.treatment","contr.treatment"))
a1 <- survfit( Surv(survival, PT_STA_ALIVEZERO) ~ factor(MP_VS_MD), data = MY7.dat[node.10| node.9| node.8,])
b1 <- coxph( Surv(survival, PT_STA_ALIVEZERO) ~ factor(MP_VS_MD), data = MY7.dat[node.10| node.9| node.8,])
plot( survfit( Surv(survival, PT_STA_ALIVEZERO) ~ factor(MP_VS_MD), data = MY7.dat[node.10| node.9| node.8],xlim=c(0,3000),xlab="Days",ylab="Probability",main=" ", lty=1,lwd=3, col=2:3)
legend(1,0.2, c("MP","MD"), lty=1,lwd=3, col=2:3)
legend(6,0.8, print(c("Log-Rank, p=", round(summary(b1)$sctest[3],3))),bty="n" )
legend(6,0.7, print(c("H-R=", round(summary(b1)$coef[2],3))),bty="n" )
summary(b1)

# Survival Plot of Non-Responders

library(survival)
options(contrasts=c("contr.treatment","contr.treatment"))
a1 <- survfit( Surv(survival, PT_STA_ALIVEZERO) ~ factor(MP_VS_MD), data = MY7.dat[node.7| node.6| node.5| node.4])
b1 <- coxph( Surv(survival, PT_STA_ALIVEZERO) ~ factor(MP_VS_MD), data = MY7.dat[node.7| node.6| node.5| node.4])
plot( survfit( Surv(survival, PT_STA_ALIVEZERO) ~ factor(MP_VS_MD), data = MY7.dat[node.7| node.6| node.5| node.4]),xlim=c(0,3000),xlab="Days",ylab="Probability",main=" ", lty=1,lwd=3, col=2:3)
legend(1,0.2, c("MP","MD"), lty=1,lwd=3, col=2:3)
legend(6,0.8, print(c("Log-Rank, p=", round(summary(b1)$sctest[3],3))),bty="n" )
legend(6,0.7, print(c("H-R=", round(summary(b1)$coef[2],3))),bty="n" )
summary(b1)