Analyses of Longitudinal and Time-to-Event Data in a Randomized Clinical Trial in the Presence of a Lag Time in the Stabilization of Treatment

By

Eugenia Yu Hoi Yin

B. Sc. University of British Columbia 2000

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

DEPARTMENT OF STATISTICS

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

February 2003

© Eugenia Yu Hoi Yin, 2003

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Statistics

The University of British Columbia

6356 Agricultural Road

Vancouver, Canada

V6T 1Z2

Date:

April 17, 2003

Abstract

Randomized controlled clinical trials (RCTs) are generally considered to be the best experimental setting for assessing new medical therapies. In medical research, the evaluation of RCTs is often based on two approaches: the commonly recommended intent-to-treat (ITT) analysis, and the more controversial per-protocol (PP) approach, which respectively attempt to assess the clinical effectiveness and the efficacy of a therapy. In the presence of a variable lag time in treatment stabilization following randomization, the two approaches may differ not only in their patient inclusion and exclusion criteria, but also in their definitions of the baseline time, from which follow-up is to be measured. In this work, ITT and PP analyses are applied to the evaluation of an eye pressure lowering therapy, in data from the Collaborative Normal Tension Glaucoma Study. In this study, the therapeutic intervention consisted of achieving a 30% reduction in intra-ocular pressure, and necessitated a lag time before the lowered pressure level became stable. This thesis includes longitudinal and survival analyses, based on measurements taken on some of the main variables in this study. In this case, the PP approach defines baseline time in the treated group as the time at which treatment stabilization has been achieved. It thus loses some of the advantages of randomization, and may suffer potential bias in parameter estimation as well as diminishing statistical power in testing the treatment effect. We investigate these potential problems through some simulation work. While the ITT and the PP approaches fail to account for the delay in treatment stabilization, we also develop a multistate model for survival analysis and a piecewise linear mixed effects (LME) model for longitudinal analysis, both of which address the lag time problem in assessing the effectiveness of the therapy. Finally, we consider a baseline-adjustment approach to match the control group to the delayed treatment group for an efficacy assessment of the therapy. These methods that account for the lag time are compared to the ITT and the PP approaches, and recommendations based on their performance in our study and their general applicability are given.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I would like to thank my supervisor Dr. Michael Schulzer for his continuous guidance and patience without which the development of this thesis would not have been possible. His expertise in both statistical and medical fields has enlightened me and motivated my interest in biostatisics. He has inspired me a lot in biostatistical research, and broadened my knowledge beyond textbooks and examples.

Also, thank you to Dr. Jim Zidek for his support and confidence in me throughout my stay in UBC as both an undergraduate and a graduate student. My decision to pursue further studies in statistics was undoubtedly due to Jim's enthusiasm and encouragement. As well, I am grateful to Dr. Nancy Heckman and Dr. Richard Mathias (from the Department of Health Care and Epidemiology) for being my second readers and for their precious advice on my thesis.

Needless to say, thanks should also go to all the graduate students and faculty of the department, and to all my friends. Vivien and Lisa: many thanks for walking with me through these years, especially for all the mental and intellectual support which bonds us closely as the millennium Head-Line Team!! Mandy and Stephanie: thanks for lending an ear to me during my hard times!! Last but not least, I am most thankful for the love and support from my parents and my aunts who are always beside me.

<div align="right">EUGENIA YU</div>

The University of British Columbia
April 2003

Chapter 1

Introduction

Randomized controlled clinical trials (RCTs) are generally recognized as the best experimental setting for assessing new medical therapies. Randomization of patients to different treatments promotes comparability of the treatment groups and minimizes potential selection biases with respect to unmeasured characteristics of the patients. Differences in the response between the control and the treatment groups may then be attributed to the treatment itself rather than to some confounding factors. When there is a lag time in the stabilization of treatment following randomization, the definition of baseline from which patients are to be followed is particularly crucial in the clinical comparisons between treatment groups. Randomization alone may not be sufficient to validate the results of treatment comparisons if the baseline is defined inappropriately. In general, when a method of comparison is inappropriate, or when the assumptions underlying a correct approach are not satisfied, the results will be erroneous and may lead one to type I or type II errors.

Finding an appropriate method of comparison in the presence of a lag time in the stabilization of treatment following randomization is often difficult. The possibility of a lag time in full treatment effect was first noted by Halperin et al. [1]. Such a lag time arises, for example, when the treatment under study does not take immediate effect upon its administration to treated patients at the time of randomization. There is a delay before the intended treatment effect is achieved, and this makes comparisons between the control and the treatment groups difficult. Although randomization of patients guards

against bias in the treatment assignment and in subsequent data analyses, this lag time is likely to affect the results of comparisons but often cannot be specified precisely before the trials are conducted.

The effect of such a lag time on statistical comparison procedures has been receiving great attention by the medical and statistical communities in the past few decades. The presence of a lag time in the treatment is mostly seen in long-term treatments. Well-known examples include the Lipid Research Clinics Coronary Primary Prevention Trial (CPPT) [2], the Women's Health Trial [3] and the Physicians' Health Study [4]. In the CPPT, the treatment was a cholesterol lowering therapy. The therapy was expected to gradually reduce the amount of plaque in blood vessels and hence the risk of coronary disease. In the Women's Health Trial, a randomized controlled trial was initiated to determine if a low fat diet effectively reduced the incidence of breast cancer among high-risk group of women. The cholesterol lowering therapy and the diet intervention introduced some sort of linear lag phase where the effect of the treatment gradually increased with time. On the other hand, a different model for the lag time was used in the Physicians' Health Study where the effect of beta-carotene on cancer incidence was investigated. The experimenters believed that the drug did not affect pre-existing tumors and time was needed for new tumors to develop and become detectable. Hence a threshold lag time was assumed: the effect of the drug was not associated with tumors detected within the first two years since administration at the time of patient randomization. In all the above cases, the treatment effect was not immediate and thus introduced a lag time from the time of randomization before the treatment reached its full effect. Several authors proposed new statistical procedures to take into account the lag time in analyzing survival data. Zucker and Lakatos [5] considered a linear and a threshold lag model, and presented two weighted log rank type statistics for comparing survival curves in a non-parametric setting. Luo [6] extended their ideas to the Cox proportional hazards regression model to include lagged effects of some of the covariates. Nevertheless, little discussion in the literature has been devoted to the consequences of applying ordinary

comparison procedures without a careful adjustment for the delay. The lag time present in non-survival type data has not been widely addressed either. It is important to make clinical practitioners and designers of clinical studies aware of the problem and this motivated this study of the statistical issues related to treatment lag times.

For this work, I have focused on investigating and discussing the effect posed by a delay in treatment stabilization on the results of treatment comparisons. The comparison procedures are based on the intent-to-treat and the per-protocol principles which are commonly adopted in the evaluation of clinical trials. Longitudinal and survival data collected from the Collaborative Normal Tension Glaucoma Study [7] - [10] are used for analyses throughout this work. Approaches that take into account the lag time present in our data have been considered, in addition to some classical methods of comparison for longitudinal and survival analyses. Furthermore, a demonstration of the potential problems including bias and diminishing statistical power in applying the per-protocol approach will be given through some simulation work.

The remainder of this thesis is organized as follows. The details of the Collaborative Normal Tension Glaucoma Study and an introduction to normal tension glaucoma are first provided in Chapter 2. Chapter 3 gives a full description of the intent-to-treat and per-protocol principles and an application to the evaluation of the glaucoma study. Methodologies of modelling the longitudinal and survival data from the glaucoma study are detailed in Chapters 4 and 5, and the results of analyses are presented in Chapter 6. Simulation of longitudinal data for demonstrating the performance of the intent-to-treat and per-protocol approaches follows in Chapter 7. A general discussion and conclusions are given in Chapters 8 and 9. The thesis ends with recommendations and suggestions for future work in Chapter 10.

Chapter 2

The Collaborative Normal Tension Glaucoma Study

The Collaborative Normal Tension Glaucoma Study (CNTGS) [7]-[10] is a prospective multi-center study for investigating the effects of intra-ocular pressure (IOP) reduction on disease progression in normal tension glaucoma (NTG). Before giving the details of the design of the study in Section 2.2 and a description of the data in Section 2.3, we familiarize the readers with some basic information on the nature of the disease, its diagnosis and management to enhance their understanding of the purpose of the trial and subsequently, the methodologies and analyses which are presented in this thesis.

2.1 Normal Tension Glaucoma

2.1.1 Introduction

Glaucoma has been one of the leading causes of blindness among adults and the elderly in particular. Its definition varies across the ophthalmic community, but the disease can be referred to as a chronic ophthalmic condition characterized by optic nerve head damage, a characteristic loss of visual field and an elevated IOP. In 1857, Von Graefe [11] described a group of patients having cupping of the optic nerve head and visual field defects but with IOP levels that remained within the statistically normal range. The term "normal tension glaucoma" was then coined to describe this particular group of glaucomatous conditions.

4

2.1.2 Epidemiology

The prevalence of normal tension glaucoma was estimated in a number of studies. The estimates range from 0.3 to 4% among patients in their mid-60s [12]. The wide range is the result of the many different definitions of NTG being employed. Similar to Open-Angle Glaucoma, which comprises the largest group of patients suffering from glaucoma with elevated IOP, normal tension glaucoma is mostly asymptomatic at the early stage. There is no associated visual field loss and therefore most patients are unaware of the disease. When untreated, NTG patients will gradually lose their peripheral vision and eventually may suffer total blindness. The Glaucoma Research Foundation [13] reported a rate of blindness from glaucoma between 93 and 126 per 100,000 people over the age of 40. In particular, Open-Angle Glaucoma accounts for 19% of all blindness among African-Americans compared to 6% in Caucasians.

As has been discussed by Sassani [14], results from previous research on the risk factors for normal tension glaucoma showed that age, gender, race, diseases including migraine and diabetes, and genetic factors are associated with the development of the disease. More specifically, the prevalence increases with age; females, Asians and African Americans, and people with migraine, diabetes or family history of glaucoma are more susceptible to developing normal tension glaucoma. The most recent study of the natural history of the disease, as conducted by the Collaborative Normal Tension Glaucoma Study Group, investigated the risk factors for the progression of visual field abnormalities in NTG [10]. It was found that the female gender, the presence of migraine and disk hemorrhage contribute separately to a higher risk of progression. Asian patients have a slower rate of progression despite a high prevalence of NTG within the race, while black patients show a faster rate of progression. Moreover, age, the untreated level of IOP and self-declared family history of glaucoma were found to have no effect on the progression rate in this study.

2.1.3 Diagnosis and Management

The diagnosis of normal tension glaucoma is often made by a diagnosis of exclusion. The determination of the nature of the disease is based upon the elimination of other diseases that share similar symptoms and characteristics. All other causes leading to damage of the optic nerve and visual field loss, for example, cardiovascular abnormalities, must be eliminated, and the IOP level has to be shown repeatedly not to exceed the normal statistical upper bound (21 mm Hg) before normal tension glaucoma can be diagnosed [12].

To measure intra-ocular pressure, tonometry is used, but it has rather poor sensitivity and specificity in detecting glaucoma if used alone [15]. Therefore in practice, it is used in combination with ophthalmoscopy, which examines the appearance of the optic nerve, for early detection of the disease. Moreover, gonioscopy helps to examine the structure of the anterior chamber angle for determining whether a patient suffers from Open-Angle or Angle-Closure Glaucoma. Normal tension glaucoma shares many clinical features with Open-Angle Glaucoma, but we do not plan to discuss their similarities here. Readers can refer to the literature on glaucoma for more details. Sassani [14] provides a comprehensive reference on the subject.

Despite the fact that an IOP outside the normal range has not been documented, patients with normal tension glaucoma tend to have a wider diurnal IOP fluctuation, which might account for the glaucomatous features in the absence of a consistent elevated IOP level [12]. Furthermore, studies have shown that asymmetric normal tension glaucoma is associated with an asymmetric IOP [16]-[18], so IOP is believed to play a role in the underlying mechanism causing the disease. And with this belief, treatments for normal tension glaucoma aim at reducing IOP levels. Patients diagnosed to have an early stage of the disease are usually treated with medications. When patients show progression on medication or experience considerable visual field loss, they are treated

with laser surgery. Filtering or incisional surgery is applied upon failure of the previously mentioned treatments or upon persistence of the progression. The medical and surgical treatments all attempt to lower IOP in hope of preventing further progressive damage.

2.1.4 Visual Field Measurements

To monitor disease progression, both the optic nerve head and the visual field need to be assessed regularly. Nowadays, automated static perimetry is used to quantify visual field loss, based on the linear relationship between visual perception and the change in stimulus intensity which is measured in the logarithm scale of decibels (dB) [15]. The Humphrey Field Analyzer (HFA), which was used in the CNTGS, is one of the most commonly used automatic perimeters. In essence, perimetry based on the HFA entails estimating threshold values at each test location in the central 30 degrees of the visual field, where a threshold can be described as the minimum brightness of a stimulus a patient perceives at a particular test location. To estimate the threshold value at each location tested, stimuli are presented at that location and the intensities are decreased in $4dB$ steps until reversal, i.e., from perceived to not perceived or vice versa. The test process then reverses and the intensities increases in $2dB$ steps until the second reversal occurs, at which time the threshold determination is stopped. The last seen stimulus intensity will be used as the threshold estimate. Detailed description of the HFA and the threshold estimation procedure can be found in [19].

From these threshold values at different locations summarize the four global indices that quantify visual field loss: the mean deviation or mean defect (MD), the pattern standard deviation (PSD), the short-term fluctuation (SF) and the corrected pattern standard deviation (CPSD). The mean defect, which is an important outcome variable in our data, is a variance weighted average departure of each test location from the age-corrected value, where a threshold of stimulus intensity is observed at every test location in the visual field:

$$MD = \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{y_i - r_i}{s_i^2} \right\} \Big/ \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{1}{s_i^2} \right\}$$

where

y_i is the observed threshold

r_i is the normal age-corrected reference threshold

s_i^2 is the between-patient variance of normal field measurements

at the ith of the n test locations.

The MD measures the overall sensitivity of the retina to light. A large negative MD is suggestive of a serious overall abnormality of the visual field. On the other hand, PSD and CPSD are more effective indices for quantifying localized visual field defects. In the presence of a cataract, MD tends to have reduced specificity because cataracts are characterized by a generalized depression of thresholds over the entire field, thus leading to a decreased MD level.

Monitoring the rate of decay of MD is useful for assessing the rate of progression of normal tension glaucoma, as MD is reflective of the overall degree of visual field defects.

2.2 Motivation and Design of the Study

For all forms of glaucoma that are associated with an elevated intra-ocular pressure (IOP), treatments always involve the lowering of the IOP, and a reduced IOP has known beneficial effect on the natural history of the disease. However, for NTG patients whose IOP stays inside the statistically normal range, the usefulness of having an IOP reduction is unknown. Clinical findings suggest asymmetric NTG is often associated with asymmetric IOP levels [16]-[18]. One of the main objectives of the CNTGS was to ascertain

the role IOP reduction plays in normal tension glaucoma. The CNTGS Group compared the time-to-progression experience of an untreated group of NTG patients to a treated group in which patients received medical, laser and/or surgical treatment(s) to achieve a 30% reduction from the mean of the last three prerandomization pressure readings. The effectiveness and the efficacy of the IOP lowering strategy were assessed by using an intent-to-treat approach and a per-protocol approach, respectively. The principles underlying the two approaches are discussed in Chapter 3.

Two hundred and thirty patients from twenty-four centers were enrolled in the study. To be eligible for the study, the patients all had unilateral or bilateral normal tension glaucoma and other ophthalmic characteristics which met the criteria as described in [7]. Upon entry into the study, patients remained unrandomized until a fixation threat or progression of the study eye(s) occurred. A fixation threat can be described as having visual field defects at the point of fixation, which is the area of maximum visual acuity in human visual field. The eligible eye of each patient was then randomized to either the control group, in which the eye remained untreated, or the treatment group, in which a 30% reduction in IOP was achieved by means of medical, laser and/or surgical interventions. Most treated patients were first placed on topical medication or laser treatment. When either or both failed to reduce the IOP to the desired level, patients underwent filtering surgery. There were also cases where treated patients were given the surgical treatment immediately after randomization.

Once stabilization of the treatment effect was achieved, the patients were followed regularly until their study eyes reached the progression end point (which is defined in Section 2.3.1) or until their lifetime in the study was censored. Meanwhile, patients had their mean defect (MD) measured repeatedly and regularly, at each of their clinical visits. The time to IOP stabilization after a 30% reduction for the treated patients, the time to the progression end point and the mean defect values comprised the three main outcomes of the study. Covariate information on demographics, medical history

and baseline ophthalmic characteristics was also collected.

2.3   Description of the Data

Due to confidentiality concerns, a subset of the data analyzed by the CNTGS Group was used for this thesis. Hereafter, I will refer to this subset as the data unless stated otherwise. The data were obtained by sampling at random 97 patients from the 145 who enrolled in the study and whose study eyes met the criteria for randomization. Among the selected 97 patients, 44 were in the treated group and 53 were in the control group. Longitudinal data of the mean defect measurements (in decibels) and survival data of the time to IOP stabilization (in days) and the time to progression (in years) were available for analyses. Besides the group variable, the effects of gender, type of therapy that treated patients received, as well as age, IOP and MD levels at baseline were studied in my thesis.

2.3.1   Definition of the Progression End Point

The Collaborative Normal Tension Glaucoma Study adopted two definitions of the progression end point: the protocol definition and a definition based on the so-called four-of-five criteria. The former ensured identification of minimal visual field alterations to minimize any risk to eyes of untreated patients in the study, and the details of this definition can be found in [7]. The latter was used for the purpose of analyzing study outcomes. A computer algorithm was developed for the identification of the progression end point. In essence, progression was considered confirmed when four of five consecutive follow-up fields showed progression in a cluster of test locations relative to baseline visual fields, with at least one non-peripheral progression point (test location) common to all four fields. A progression relative to baseline fields was defined as having two or more adjacent points (they could not all be peripheral) whose MD values decreased by at least

$10dB$, relative to the average of baseline values at each of these points taken at the time of randomization. A complete description of the four-of-five criteria can be found in [7].

Chapter 3

Evaluation of Randomized Clinical Trials

3.1 Intent-to-Treat versus Per-Protocol Principles

In conducting clinical trials, treatment assignment is ideally done through randomization because randomization tends to give an unbiased comparison of the different treatment groups. In practice, however, clinicians often encounter problems of patient drop-outs, non-compliance and missing observations. Some patients do not ultimately receive the treatments to which they are preassigned. This then leads to concern about how one should analyze clinical trials in order to have a proper comparison between the treatment groups. There have been two principles that are adopted in the evaluation of clinical trials: the intent-to-treat (ITT) principle and the per-protocol (PP) principle. The former is based on the idea that *all* patients who are randomized should be included in the final analysis of the trial irrespective of the presence of drop-outs, cross-overs and non-compliance. Patients are assumed to remain in the treatment groups to which they have been randomized even if they switch to another treatment during the period they are followed. According to Lachin [20], the intent-to-treat principle refers to a set of criteria for the evaluation of the benefits and risks of a new therapy that essentially calls for the complete inclusion of all data from all patients randomized in the final analyses.

The intent-to-treat principle is contrasted with the per-protocol principle in which the main purpose of the analysis lies in the assessment of the efficacy of a treatment. With this principle, the evaluation of a clinical trial is based only on patients who actually

adhere to the treatment preassigned. Observations on dropped-out and non-compliant patients are excluded from the analysis. Lachin [20] described the principle as a strategy to select and to examine the experience of a subset of patients that meet the desired efficacy criteria for inclusion in the analysis.

It is important to distinguish between efficacy and effectiveness because their assessment entails different strategies and clinical implications. Efficacy refers to the effects of an intervention, such as a medication under ideal conditions, while effectiveness refers to how successful an intervention is in clinical practice whose conditions often deviate from the controlled conditions in efficacy studies. As described by Indrayan and Sarmukaddam [21], efficacy is evaluated under controlled conditions, whereas effectiveness is determined not only by efficacy but also by coverage, compliance, provider performance, etc.

There has been intense controversy over which of the two principles should be employed in making treatment comparisons. Advocates for the intent-to-treat approach argue that it not only provides a means of treatment comparison in an unbiased fashion, but also realistically assesses the usefulness of a treatment in clinical practice where it is infeasible to track how patients are receiving their prescribed treatments. The other group advocating the per-protocol approach argue that an intent-to-treat analysis is less powerful in detecting the presence of a treatment effect because a treatment effect is possibly diluted by the inclusion of patients who do not adhere to their treatments. Those who follow the per-protocol principle believe that by studying the subset of patients who do receive the treatments exactly as described in the protocol, the treatment effect can be truly assessed and estimated. But they have overlooked some potential problems of this approach.

Lachin [20] described some of the statistical considerations in the intent-to-treat design and in the analyses of clinical trials. He also discussed potential bias and statistical power issues related to the per-protocol analysis which he referred to as the efficacy subset analysis. The dropped-out and non-compliant patients possibly comprise a group

with demographic characteristics and health conditions substantially different from the rest of the patients who are included in the per-protocol analysis. This subset of patients being included is not identified at the time of randomization, and hence randomization of patients does not help ensure an unbiased comparison. The validity of the results from the per-protocol analysis becomes questionable; even if the results are valid, they may not be applicable to the study population in general.

In many recent studies, clinical trials are evaluated using both approaches. Results from both are reported and compared. There seems to be a belief that when a per-protocol analysis identifies a significant treatment effect, the same result obtained in the intent-to-treat analysis further demonstrates the usefulness of the treatment. And for this reason, the intent-to-treat analysis tends to serve as a confirmation tool of a treatment assessment, rather than a means of unbiased treatment assessment. It is unclear how the results obtained from the two approaches are related and can be compared as they involve essentially two different sets of patients (one set is the subset of the other). Regardless, one should be aware of the potential problems of the per-protocol approach. Moreover, because of its difference in clinical interpretation and implication from the intent-to-treat approach, one needs to be cautious when drawing conclusions from either approach.

3.2 Intent-to-Treat and Per-Protocol Analyses of the Normal Tension Glaucoma Data

In the evaluation of the Collaborative Normal Tension Glaucoma study, both the intent-to-treat and the per-protocol approaches were used for treatment comparisons. In particular, the survival experience of the treated patients who had their IOP levels reduced by 30% from the prerandomization values was compared to that of the patients in the control group who remained untreated until they reached the progression end point or their follow-up times in the trial were censored. In the intent-to-treat analysis, all the

patients who were randomized were included for treatment comparison, regardless of the fact that there were problems such as inability to achieve the desired reduced level of IOP, progression prior to IOP stabilization, treatment complications that affected visual acuity of the treated patients, and non-compliance. In contrast, patients having any of the above problems were excluded from the per-protocol analysis.

As the two approaches differ in the inclusion of patients for analyses, a substantial difference in sample size is expected when the degree of drop-outs and non-compliance is high. However, among the 145 patients who were randomized in the Collaborative Normal Tension Glaucoma Study, only five treated patients withdrew from the study before the stabilization of IOPs took place and did not meet the efficacy requirement to be included in the per-protocol analysis.

In defining the intent-to-treat and the per-protocol approaches that were adopted by the CNTGS group, not only the original criteria of inclusion of patients were used, but also two different baselines were defined. In the intent-to-treat analysis, the baseline was taken to be the time of randomization for all the patients. Even though a 30% reduction in IOP was not immediately achieved upon medical, laser or surgical intervention for the treated patients, measuring the patients starting from the time of randomization gave a reasonable assessment of an overall clinical effectiveness of the treatment as the treated patients began with the IOP-lowering therapy at randomization. On the other hand, in the per-protocol analysis, treated patients had their follow-up times measured from the baseline time at which their IOPs stabilized after a 30% reduction. Equivalently, the baseline for the control group remained at the time of randomization while that for the treatment group was shifted to the patients' individual times of IOP stabilization. This new baseline was chosen for the treatment group because having a 30% reduction in IOP was the desired treatment criterion. The treatment with the IOP reduction was supposed to have taken its full effect at the time of IOP stabilization.

Due to the small number of patients who withdrew from the study, the ITT and PP

approaches did not differ much in terms of sample size. Rather, the baseline definition was the distinguishing factor between the two approaches. Throughout this thesis, we will follow the same principles defining the intent-to-treat and the per-protocol approaches as were adopted in the evaluation of the Collaborative Normal Tension Glaucoma Study. The results of our analyses based on the two approaches will be presented in Chapter 6, and the reliability of the results from the two approaches will be discussed in Chapter 8.

Chapter 4

Linear Mixed Effects Models for the Longitudinal Mean Defect Data

To model repeated measurements collected over time, the multivariate normal distribution (MVN), generalized estimating equations (GEE) and the linear mixed effects (LME) model are amongst the popular choices. The MVN approach may not be applicable to our case because it generally works best only when the subjects have observations taken at a common set of times. Moreover, its application puts quite a strong distributional assumption on the data. In order to have more relaxed assumptions to work with, and to incorporate irregular follow-up measurements for patients enrolled in the Collaborative Normal Tension Glaucoma Study, the GEE [22] and the LME model were considered. Fitting the mean defect data using the two methods gave similar results marginally, but the LME model was chosen for further analysis of the data because it allows for random effects for covariates which vary substantially between subjects. Also, the LME model automatically takes care of problems of missing response and has the flexibility of fitting a wide variety of correlation structures for the within-patient errors and for the random effects. In particular, the data show a large between-patient variation in the baseline mean defect (MD) and in the decay pattern of MD over time. Using the LME model, we can better model the MD data by allowing different intercepts and decay rates for different patients.

4.1   The Linear Mixed Effects Model

The linear mixed effects model that was fitted to the data was described in Laird and Ware [23]. For individual i, the model has the form

$$\begin{cases} \mathbf{Y}_i = \mathbf{X}_i\beta + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i \\ \mathbf{b}_i \sim N(0, \mathbf{D}), \mathbf{e}_i \sim N(0, \mathbf{R}_i), \mathbf{b}_i \perp \mathbf{e}_i \end{cases} \tag{4.1}$$

where

$\mathbf{Y}_i$ is the vector of responses,

β is the vector of fixed effects, which are constant across subjects,

$\mathbf{b}_i$ is the vector of random effects, and is independent of $\mathbf{b}_j$ for $i \neq j$,

$\mathbf{X}_i$ and $\mathbf{Z}_i$ are the design matrices for the fixed and random effects, respectively,

$\mathbf{e}_i$ is the vector of within-subject errors, and is independent of $\mathbf{e}_j$ for $i \neq j$.

The model in Equation (4.1) has two major components: the mean structure $\mathbf{X}_i\beta$ and the covariance structure defined in terms of the distribution of the random effects and the within-subject errors. The vectors of random effects, $\mathbf{b}_i$'s, are each assumed to be normally distributed with mean $\mathbf{0}$ and covariance matrix $\mathbf{D}$, and are mutually independent of the $\mathbf{e}_i$'s. The vector of within-subject errors, $\mathbf{e}_i$, also follows the normal distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{R}_i$ of dimension $n_i \times n_i$, where n_i is the number of observations for individual i. The unknown parameters in $\mathbf{R}_i$ do not depend upon i.

It can be shown that marginally, $\mathbf{Y}_i$ is independently normally distributed with mean $\mathbf{X}_i\beta$ and covariance matrix $V(\mathbf{Y}_i) = \mathbf{R}_i + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T$. The random effects introduce an extra component of variation $\mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T$ to the response variable. Furthermore, in the simplest case where $\mathbf{R}_i = \sigma^2\mathbf{I}$, i.e., the within-subject errors are mutually independent, the

individual response components within the same subject, $(Y_{i1}, Y_{i2}, ..., Y_{in_i})$ are correlated in the presence of random effects.

Estimation of the fixed and random effects, and parameters of the covariance structures can be based on least squares and maximum likelihood methods, or an empirical Bayes methodology. Details on the estimation procedure were discussed in Laird and Ware [23]. The maximum likelihood (ML) and the restricted maximum likelihood (REML) methods are by far the most popular choices of estimation procedures. REML estimates are often more efficient as REML estimation adjusts for the degrees of freedom used in the estimation of the parameters. However, when the total number of observations $\sum_{i=1}^{m} n_i$ (m is the number of subjects) is much larger than the number of unknown parameters which are to be estimated, the ML and the REML methods give very close parameter estimates.

4.2 Application to the Mean Defect Data

One of the major study questions which will be addressed in this thesis is whether a 30% reduction of IOP from the baseline value successfully slows down the rate of generalized visual field loss as measured by the rate of decay of the mean defect. Moreover, the mean defect level of the control and the treated groups would also be our study interest because a more negative level is indicative of a higher severity of normal tension glaucoma.

Preliminary plots of the individual patients' mean defect trajectories for the control and treated groups separately (see Figures 4.1 and 4.2) suggest a general trend of depression of the MD over time. Although the MD level within each patient tends to show moderate fluctuation, it does not seem too unreasonable to assume a linear model for the MD data of each of the individuals. Furthermore, we observed a large between-patient variation in the baseline MD readings at randomization and in the rate of change of MD over time. To take into account the highly variant information across patients, we fitted

Figure 4.1: The observed MD trajectories over time from randomization (in days) for the 53 control patients in the data set.
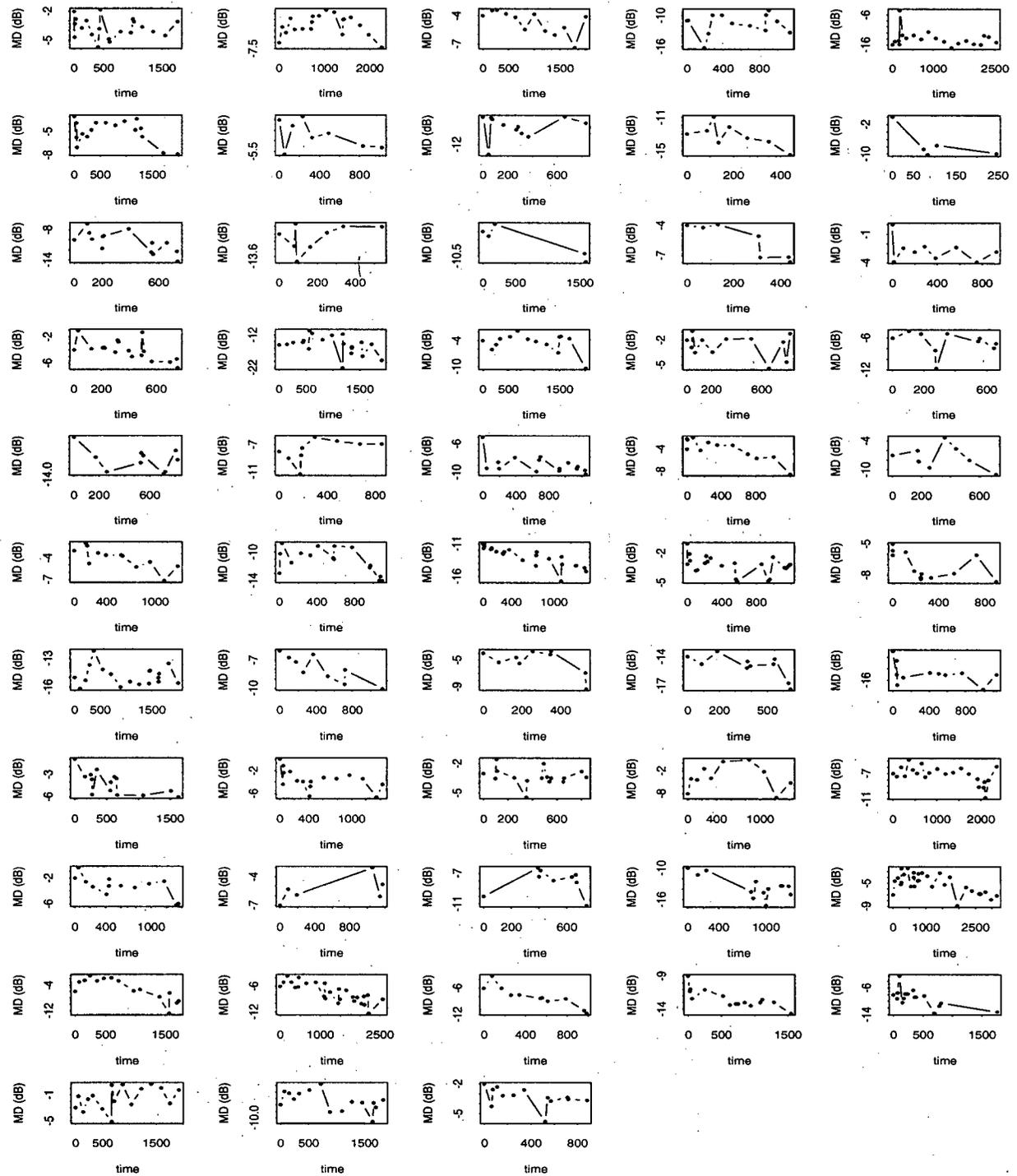
Figure 4.2: The observed MD trajectories over time from randomization (in days) for the 44 treated patients in the data set.
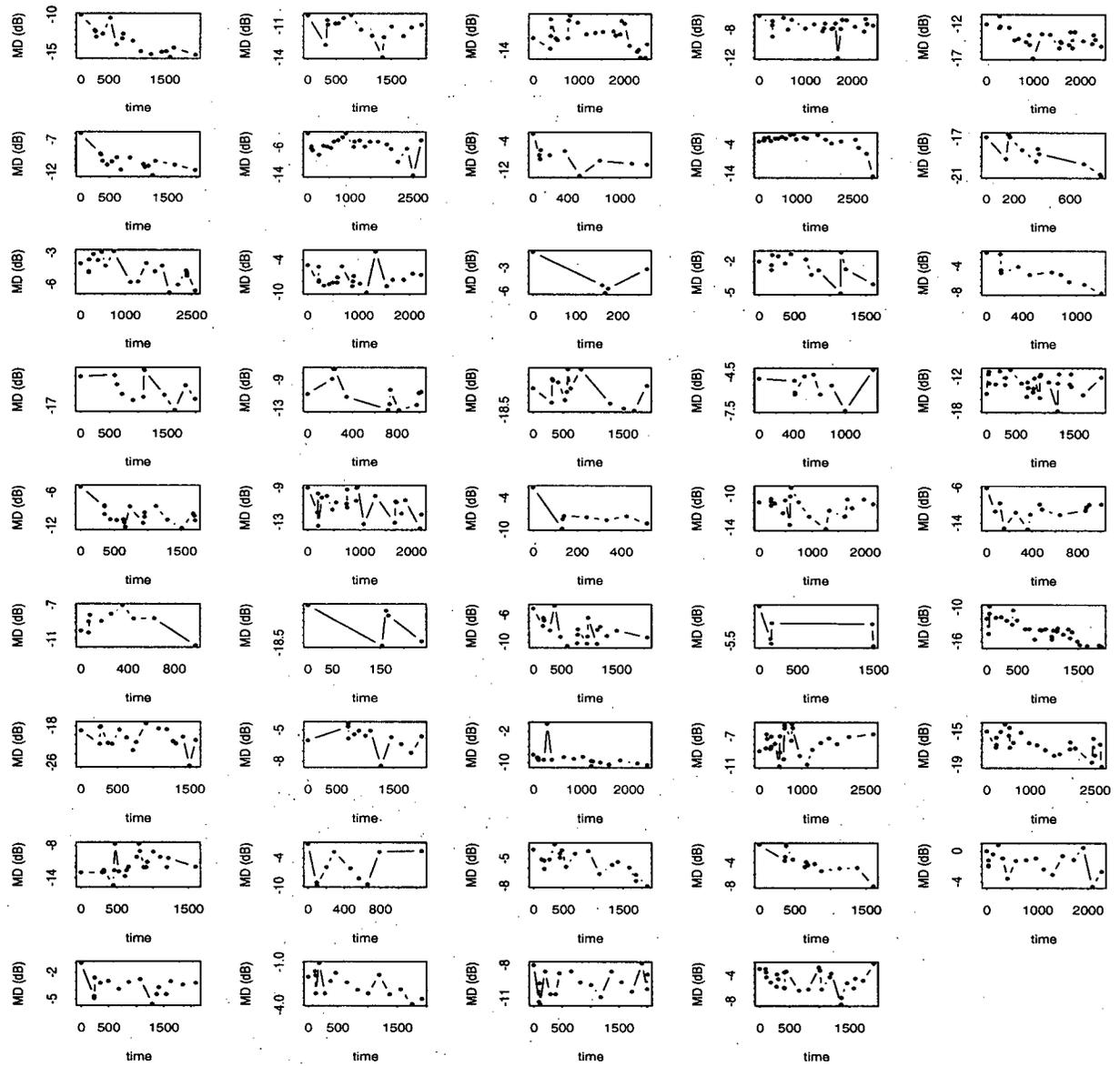
the Laird and Ware linear mixed effects (LME) model given in Equation (4.1) to the mean defect data and included as random effects the intercept and the time-slope covariate. The response variable was the MD measured over time. Covariates that comprised the fixed effects were

- the times of repeated MD measurements

- gender

- group membership

- age at baseline

- baseline MD

- baseline IOP

In both the intent-to-treat and the per-protocol analyses, the interaction between time and group was also included in order to assess any difference in the decay rate of the MD between the two treatment groups. A significant time×group interaction effect would imply a significantly different rate of change in MD and hence a different rate of disease progression between the control and the treated groups.

Various covariance structures including the independence model, the first-order continuous autoregressive model (CAR1, which is equivalent to the exponential model) and the Gaussian model were fitted to the within-patient errors in the preliminary analysis of the data. The first-order continuous autoregressive model was found to give the best fit. For the random effects, an unstructured covariance matrix was assumed.

Moreover, there is a lower level below which the MD seldom attains in normal tension glaucoma. Patients whose MD levels at baseline are close to the lower limit might show a different trend of depression and in particular a slower decay rate than those who have

less negative MD at baseline, because the closer the MD is to the lower limit, the less room is left for depression of the generalized visual field. This phenomenon is referred to as the "floor effect". In the case where the treated patients and the control patients in our study began with rather different initial MD levels, a difference between the mean MD decay rates of the two groups of patients might be the result of a potential floor effect rather than of a significant treatment effect. We checked for the presence of such a floor effect in our data by including an additional covariate into the LME model in both the ITT and the PP analyses. The covariate was an indicator variable for whether a baseline MD level was above or below -12dB. This particular value was chosen because a patient with an MD less than -12dB is generally regarded as having an advanced stage of normal tension glaucoma. The interaction between time and this indicator variable can be tested to investigate whether the decay pattern of the MD depends on the baseline MD level. If such an interaction exists, it would imply patients who have advanced normal tension glaucoma have a different MD decay rate from patients with an early or a moderate stage of the disease, regardless of whether they are treated or not.

Results from fitting the LME model under the intent-to-treat and the per-protocol approaches are given in Chapter 6, and the appropriateness of the two approaches to analyzing the MD data are discussed in Chapter 8.

Chapter 5

Multistate Models for the Time to Event Data

Classical methods of survival analysis such as the Kaplan-Meier survivor function [24] estimation, Cox regression [25] and parametric modelling have been extensively used for analyzing time to event data. While these methods can easily be applied to our study of the time to progression, we have not taken into account the possible linkage between the two events: IOP stabilization and progression in the intent-to-treat and the per-protocol analyses of the time to progression data. If we are to focus on the treated group only, the time to IOP stabilization may provide valuable information to our understanding of the hazard of progression. By including the time to IOP stabilization as a covariate, we can assess its effect on the time to progression within the treated group. However, since the time to IOP stabilization was irrelevant and thus unobserved in patients in the control group, taking into account this time information is not directly feasible when comparing the survival experience to progression between the two treatment groups. Here we consider a multistate modelling approach. By means of a multistate model, we can model the events of IOP stabilization and progression simultaneously, and flexibly deal with the partially-observed times to IOP stabilization.

5.1   Stochastic Multistate Models for Survival Time Data

In ordinary survival analysis, time to failure is the only outcome of interest. Individuals are followed until they reach the failure end point or until their lifetime is censored.

24

Within the framework of stochastic processes, the occurrence of failure can be thought of as a transition from an "alive" state to an absorbing state of failure, and the force of transition is equivalent to the hazard function for the survival time. A multistate model is an extension of the simple two-state Markov model in ordinary survival analysis. It can flexibly accommodate intermediate events that happen before failure and whose effects on the risk of failure are of study interest. It can also be used to model multiple failure events in situations where there is more than one cause of failure and the risks of failure due to different causes are to be compared.

By definition, a multistate model is a model for a stochastic process which occupies one of a set of discrete states at any time point. A transition is defined as a change in state. The state structure specifies the states and which transitions are possible. In order to define the full statistical model, one must specify the state structure as well as the form taken by the hazard functions for each possible transition. There are many different types of multistate models and we will focus on the disability model which is described in the following section. Readers are referred to Hougaard [26] for more information on the types of multistate models and their applications.

5.1.1 The Disability Model

Before failure occurs, one might observe other events which possibly affect the risk of failure. For example, in a clinical trial studying whether a heart transplantation is beneficial to patients suffering from heart diseases, patients who are randomized to the treatment group may receive heart transplantation surgery during their follow-up periods, and some may die before a suitable heart is available for transplantation. Moreover, the time of transplantation from randomization is variable across the treated patients. In order to assess whether having a transplanted heart reduces the risk of death caused by heart attacks, it would be useful to model receiving a transplantation as a separate event. Another motivating example is the development of non-fatal complications during the

Figure 5.3: The Disability Model
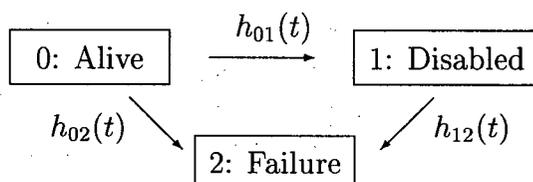
course of a disease. Patients infected with a certain disease may die with or without a complication. However, the complication which only occurs in some but not all of the patients at a variable time may be assumed to alter the risk of death. In the above two examples, the events of receiving a transplantation and the development of a complication are likely to influence the risk of failure which is to be assessed. In such cases, it is useful to model the two events: transplantation/complication and failure simultaneously. A special type of multistate model that fits in with the context similar to the examples is the disability model.

As shown in Figure 5.3, the disability model consists of three states: state 0 usually refers to the "alive" state without the "disability" which can be the heart transplantation or a disease complication; state 1 is the "disabled" state in which individuals are "disabled" but have not yet failed, and state 2 always refers to the state of failure. The hazard function for a transition from state i to state j at time t is represented by $h_{ij}(t)$. Transitions that are possible are indicated by the arrows. The disability model depicted in Figure 5.3 describes a situation where a patient can fail with and without going through the "disabled" state and all the transitions are irreversible. For instance, a patient who has made a transition from state 0 to state 1 cannot return to state 0. In other words, once a patient becomes "disabled", he/she will remain "disabled" for the rest of his/her lifetime. Also, state 2 is an absorbing state: one cannot leave state 2 once the person enters the state.

A common approach to analyzing the disability model involves separate analyses of the times to different states. The individual hazards for transitions from state i to state j, $h_{ij}(t)$ ($i, j = 0, 1, 2, i < j$), can be modelled separately by means of semi-parametric and parametric models. The modelling of an overall hazard function for the disability model has also been discussed in the literature. Andersen [27] proposed the use of transition probabilities between states to define an overall hazard function as a linear combination of the individual hazard functions. The author gave an illustration of the approach using the Steno Memorial Hospital Diabetes survival data. Hougaard [26] discussed the possibility of applying a single Cox proportional hazards model with time-dependent covariates in analyzing the disability model and exemplified an application using the well-known Stanford heart transplant data [28].

5.1.2 Application to the Normal Tension Glaucoma Data

In the Collaborative Normal Tension Glaucoma Study, patients who were randomized were followed until they reached the progression end point or their lifetimes in the trial were censored. The treatment of lowering the IOP by 30% of the pre-randomization levels for the treated group did not take immediate effect. In particular, a variable time elapsed after randomization before the desired reduction of the IOP was achieved. Upon a successful reduction, we can regard the treated patients as making a transition to the state where their IOPs remained stable but the progression end point had not yet been reached. They remained in this state until a transition to the progression state occurred. A few treated patients who were randomized in the study but not included in our data indeed failed to achieve the desired level of reduced IOP. They made transitions directly to the progression end point as in the case of the control patients. Furthermore, a 30% reduction of IOP was permanent upon IOP stabilization and the transition to the progression state was irreversible. Thus, we see that the disability model provides a reasonable description of the sequence of events observed in the study. We could define

Figure 5.4: The Disability Model for the Glaucoma Data
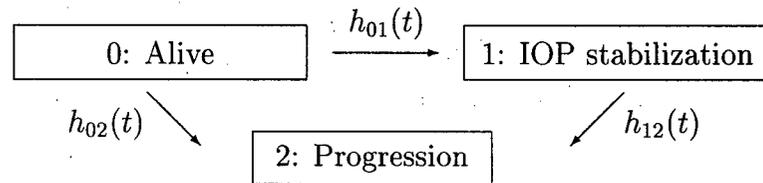
the three states in the disability model, as shown in Figure 5.4 as follows:

- State 0: Alive, in which state a patient did not have a stable 30% reduction of IOP and had not yet reached the progression end point.

- State 1: IOP stabilization, in which state a patient achieved a stable IOP after an intended reduction but had not yet reached the progression end point.

- State 2: Progression, in which state a patient reached the progression end point.

There was a need to include the state of IOP stabilization in the model because a 30% IOP reduction was fully achieved upon transition to this state and the effect of the stabilization of the treatment on the risk of progression was to be assessed.

Using the same notation in defining the individual hazards in the disability model, let $h_{01}(t)$ represent the hazard function for the time taken to achieve a stable 30% reduction in IOP as a result of medical or surgical intervention, $h_{12}(t)$ and $h_{02}(t)$ represent the hazard functions for the time to progression with and without IOP stabilization, respectively.

When we model the transition hazards separately, we can incorporate different covariates that are relevant to the different transitions. For example, when modelling $h_{01}(t)$ within the treated group, we can include the type of IOP lowering therapy as a covariate

and study if the hazard of reaching IOP stabilization differed between patients who received surgical and non-surgical treatments. When modelling $h_{12}(t)$ among the treated patients, the effect of the length of the time period awaited for IOP stabilization on the progression hazard after establishing a stable 30% IOP reduction might be of interest. In addition to the waiting time for IOP stabilization, the IOP and MD levels at stabilization can provide valuable information to be included as covariates in the model. Since by definition all patients in the control group reached the progression directly without going through the state of IOP stabilization, the modelling of the progression hazard without a stable IOP reduction, $h_{02}(t)$, involves initially all the treated and control patients. Moreover, we can add the treatment group covariate to the hazard model and assess whether the untreated and treated group had a different risk of progression without a stable IOP reduction. Treated patients who had their IOPs successfully reduced before progression will have their lifetimes censored at the time of IOP stabilization for the estimation of $h_{02}(t)$.

Within the family of Cox models [29]-[30], if the length of the waiting period to achieve IOP stabilization does not have an effect on the risk of progression with a stable IOP reduction, and if we have proportionality between the two progression hazards at all times t: $h_{12}(t) = c \times h_{02}(t)$, for some unknown c, then the three states in the disability model can be analyzed simultaneously using a single Cox model [26]. More specifically, we can fit a single Cox model $h(t)$ with a time-dependent covariate $I(t)$ as an indicator function, where $I(t) = 0$ during the period when no stabilization has been achieved, and $I(t) = 1$ during the period after the stabilization is established. We can also include other covariates such as MD and IOP at randomization, age, group membership and gender.

The Cox model with the time-dependent covariate $I(t)$ is given as follows:

$$h_i(t) = h_0(t)exp\{\alpha I_i(t) * x_{group,i} + \beta^T \mathbf{x}_i\} \tag{5.2}$$

where for the ith individual, $h_i(t)$ is the hazard function, $h_0(t)$ is the baseline hazard function and $I_i(t)$ is the time-dependent indicator variable for IOP stabilization. The

time t is measured from the time of randomization. The vector of covariates including the group covariate $x_{group,i}$ is represented by $\mathbf{x}_i$, and β is the vector of regression coefficients representing the log relative hazard ratios. Since $I_i(t)$ is relevant to the treated patients only, we make use of an interaction term between $I_i(t)$ and the group variable ($x_{group,i} = 0$ for the control group, 1 for the treated group) so that $I_i(t)$ only enters the above Cox model if patient i is treated. The significance of the coefficient for the interaction term, α, can be tested to assess the effect of IOP stabilization on the progression hazard within the treated patients. Also note that by using an indicator function for IOP stabilization, we are assuming a threshold treatment effect, i.e., at times when a treated patient had not yet shown a desired stable IOP reduction and $I_i(t) = 0$, we will assume the treatment has not taken any effect.

As mentioned earlier, some treated patients in the Collaborative Normal Tension Glaucoma Study reached the progression end point before IOP stabilization was achieved (they made direct transition from state 0 to state 2 in Figure 5.4) though they constituted only a small proportion of the treated patients. However, the data we analyzed in this thesis did not include any of these patients. Consequently, when we model the transition hazard $h_{02}(t)$, all the treated patients had censored times to progression without a stable IOP reduction. We therefore have difficulty in comparing the hazards of progression without a stable IOP reduction between the control and the treated groups. Adjusting for the group covariate in modelling the hazard $h_{02}(t)$, which could have been done if the data from treated patients who reached the progression end point prior to IOP stabilization were available, may not provide useful information to our understanding of the time-to-progression experience without IOP stabilization in our case. We might also encounter computational problems in the estimation of the group covariate coefficient. In seeing these potential problems, we will focus on the control patients only and leave out the group covariate in modelling $h_{02}(t)$ in this thesis.

Besides adopting the multistate modelling approach by means of a disability model, other non-parametric, semi-parametric and fully parametric methods will be used in each of the intent-to-treat and the per-protocol approaches to analyze the time to progression data in our study.

5.2   Baseline-Adjustment Analysis of the Time to Progression Data

While the disability model accounts for the lag time in the stabilization of the treatment by including the IOP stabilization as an intermediate event before progression occurs and by modelling the lag time as the time to IOP stabilization from randomization, this multistate modelling approach does not deliberately adjust the baseline for the control group to correspond to the delay in the treatment stabilization in the treated group. One possible method of correspondence is the use of "controls matched by covariates". In essence, the method involves matching each control with a treated patient with similar covariate such as age, gender or other baseline information. Each control will have a baseline shifted from randomization to the time of IOP stabilization of the treated patient being matched. However, in our data set, the numbers of treated and control patients are different. There are not enough treated patients to perform a one-on-one matching with the control patients and thus the method may not be applicable. Another way of adjusting the baseline can be based on the distribution which the time to IOP stabilization tends to follow. The best fitted distribution can be found in the parametric modelling of the time to IOP stabilization data. We can then shift the baseline for each control patient from the time of randomization by a time randomly generated from the best fitted parametric distribution. The lifetime of the control patients will be measured from the new baseline and that of the treated patients will be measured from their individual times of IOP stabilization. Ordinary survival analysis of the adjusted data can be performed to compare the time-to-progression experience of the two groups of patients after the stabilization of the treatment is established.

Results of analyses using the classical methods, the multistate modelling approach and the baseline adjustment approach are presented in Chapter 6. A full discussion of the performance of the different approaches can be found in Chapter 8.

Chapter 6

Results

6.1 Analysis of the Mean Defect Data

The rationale for using the linear mixed effects (LME) model to fit the mean defect data was given in Chapter 4. A full model with all the covariates listed in Section 4.2 and a time×group interaction as fixed effects was initially fitted to the MD data using both the intent-to-treat (ITT) and the per-protocol (PP) approaches. A random intercept term and a random time slope were fitted to account for the apparently different baseline MD levels and decay pattern of MD over time across patients. When we switched from the ITT to the PP analysis, the times of repeated MD measurements and the values of the baseline covariates were adjusted according to the new baseline time. The group and gender variables were parametrized in the following way:

- group: 0 for control group, 1 for treated group

- gender: 0 for female, 1 for male

Furthermore, we assumed the first-order continuous autoregressive (CAR1) model for the within-patient errors and an arbitrary covariance matrix for the random effects. The parameter estimation was based on the Restricted Maximum Likelihood (REML) method.

Table 6.1: REML coefficient estimates of the fixed effects in the LME model for the ITT and PP approaches.

	ITT		PP	
Fixed Effects	Estimate	P-value	Estimate	P-value
β_1 - time	-0.00160	<0.0001	-0.00157	<0.0001
β_2 - group	-0.993	0.005	-0.002	0.995
β_3 - baseline MD	0.834	<0.0001	0.913	<0.0001
β_4 - time×group	0.0005	0.059	0.0007	0.021

The MD data for all the 97 patients were included in the ITT analysis, while the data from one patient who dropped out of the study were excluded from the PP analysis. We tested the significance of each covariate in the full model, and both analyses showed an insignificant effect of gender, age at baseline and IOP at baseline at a 5% level. We then fitted the following reduced model:

$$Y_{ij} = \beta_{0i} + \beta_{1i}(\text{time}) + \beta_2(\text{group}) + \beta_3(\text{baseline MD}) + \beta_4(\text{time} \times \text{group}) + e_{ij}$$

$$(6.3)$$

In Equation (6.3), i is the patient index and j is the index of repeated measurements on the same individual. Y_{ij} is the jth MD measurement of the ith patient. Note that the intercept and the slope coefficients are different for different patients. We can write $\beta_{0i} = \beta_0 + b_{0i}$ and $\beta_{1i} = \beta_1 + b_{1i}$, where β_k's are the fixed effects and b_{ki}'s are the random effects.

For both the ITT and the PP analyses, model diagnostics were performed and showed that the reduced model gave a reasonable fit to the data. The normality of the random effects and the within-patient errors, and the CAR1 model for the latter seemed to be a valid assumption. The ITT and the PP estimated autocorrelation coefficients were respectively 0.21 and 0.16. This implies the correlation between the errors present in two MD observations measured one day apart was 0.21 and 0.16 on average.

With the same model being fitted, the two approaches gave rather different results, which are presented in Table 6.1. After adjusting for the baseline MD value, the ITT analysis gave a slightly more negative overall rate of change of MD than the PP analysis. At a 5% level, the group coefficient was found to be significantly less than zero in the ITT analysis, indicating that the mean MD level was lower in the treatment arm than in the control arm after adjusting for the baseline levels. However, the PP analysis showed no evidence of a difference between the mean MD levels of the two groups after the baseline levels were adjusted, at the same significance level. Moreover, we obtained a positive coefficient estimate for the group and time interaction term in both the ITT and PP analyses. The positive estimate suggested a slower decay rate of MD for the treated patients on average than the patients in the control group. In particular, the PP analysis showed a significant and stronger interaction (a more positive coefficient); the ITT analysis gave a marginally significant result (p=0.059). At a 5% level, the ITT analysis did not show a significantly slower depression rate of MD for the treated group, while the PP analysis demonstrated a superior treatment effect in slowing down the deterioration of the visual field.

In order to test for the presence of a floor effect, we included in the model in Equation (6.3) an extra covariate as a fixed effect:

$$X_{floor} = I(\text{baseline MD} < -12\text{dB})$$

and an interaction term between time and X_{floor} in both the ITT and the PP analyses. The interaction between the indicator variable and time was tested and was found to be insignificant (p=0.4885 for ITT and p=0.6420 for PP). The decay pattern of the MD appeared to be independent of the baseline MD level regardless of the approach of analysis.

6.1.1 Piecewise Linear Modelling Approach

To understand what actually led to the difference in the results of the ITT and the PP analyses and consequently, the conclusions made on the usefulness of an IOP reduction strategy, we looked closely at the data in hope to reveal any hidden pattern which was not observed in the individual MD trajectories. Upon examination of the average MD level over time across patients in the control and the treatment groups separately, we suspected that the MD of the treated group had rather different decay patterns before and after IOP stabilization. In Figure 6.5 where the average MD level for the two groups were plotted, we observed the following:

- Over the period (0 days, 1750 days), the average MD level for the control group appears to decrease linearly with time in spite of its mild fluctuation.

- The treated group shows a sharp monotonic decline in the average MD over the period (0 days, 170 days), which is followed by a considerably slower decay rate until about 1750 days. Notice that the median time to IOP stabilization among treated patients is 189 days. The rapid decay of MD in the beginning might possibly be the result of the IOP lowering therapy, which could have caused further depression of the generalized visual field in addition to the natural depression as observed within the untreated patients. Nevertheless, upon the stabilization of the IOP, the treated group seems to have a slower decay rate than the control group, and the average MD level curves of the two groups merge at about 1750 days. It appears that the average MD level curve for the treated group can be well approximated by a segmented linear model with a change point close to the median time of IOP stabilization. Such an observation is not apparent at the individual level, but it does suggest a potential need for fitting a piecewise linear model to the individual treated patients and a pure linear model to the individual control patients.

- There is substantial overlapping of the two average MD level curves beyond 1750

days. Moreover, the curves have a much wider fluctuation than was the case before 1750 days. A possible explanation for the large fluctuation is the loss of patients who had reached the progression end point. The mean (median) times to progression observed in our data are 1190.4 days (1122 days) and 1714.0 days (1889.5 days) for the control and the treated groups, respectively. With the increasing number of patients who left the study either because of drop-outs or reaching the progression end point as time elapses, the average MD levels of the remaining patients are subject to a larger variation and hence, a rightward funnel shape is seen in the plot.

- The average MD level curve for the treated group does not show any consistent superiority over that of the control group over the entire follow-up period. The treated group has an average MD level below the control group most of the time, and the treatment does not seem to result in a less negative MD level. In particular, the treated patients were at a disadvantage because of a dramatic decrease in MD during the period awaited for IOP reduction and stabilization even though the mean time to progression was about 500 days longer among the treated patients.

We have indeed looked to see if the piecewise linear model was necessary for the control group. Analyses of the MD data performed separately on the two groups showed that a linear model adequately described the MD trend over time for the control group while a segmented trajectory was needed for the treated group.

As the ITT approach assumed a linear MD trajectory for each treated patient and the PP approach modelled the post-IOP-stabilization data, the two approaches did not explicitly take into account the rapid decay of the MD during the period awaited for IOP stabilization within the treated group. To account for the difference in the average decay rate of MD before and after IOP stabilization among the treated patients, here we considered a linear mixed effects model which allows for a piecewise linear trajectory for the treated group but a linear trajectory for the control group.

Figure 6.5: The average MD level observed over time from randomization for the control and the treated groups.
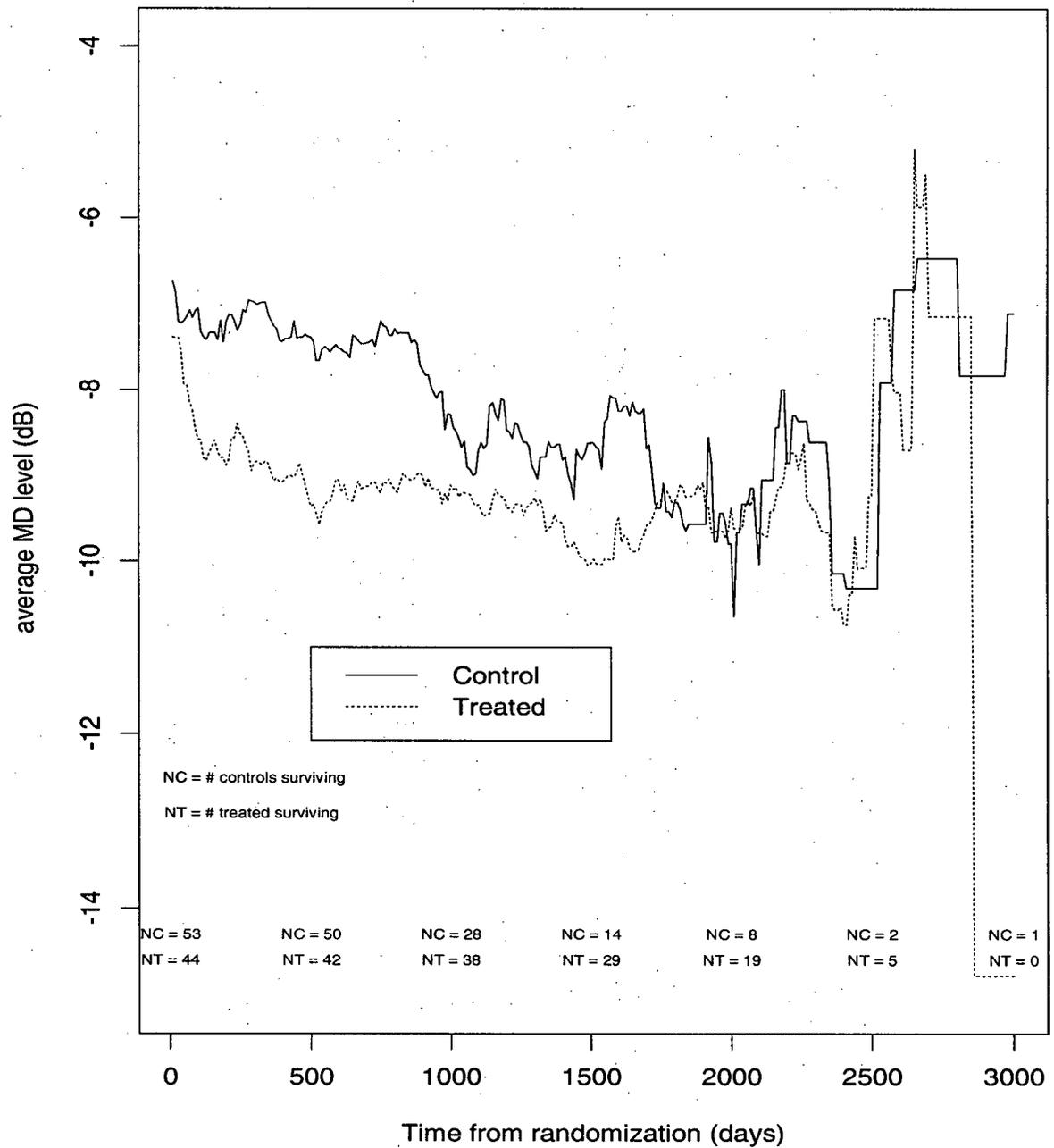
The model we fitted for this piecewise linear (PCLIN) approach was parameterized to have a mean structure

$$\beta_{0i} + \beta_1 x_{1i} + \cdots + \beta_p x_{pi} + \beta_{p+1} x_{p+1,i} \tag{6.4}$$

for the ith patient. The component $(\beta_{0i} + \beta_1 x_{1i} + \cdots + \beta_p x_{pi})$ is the same mean structure of the linear model under the ITT and the PP approaches. The extra covariate $x_{p+1,i}$ added to the structure is defined as

$$x_{p+1,i} = (\text{time} - T_{Si}) I(\text{time} > T_{Si}) \times I(\text{treatment})$$

where T_{Si} is the time taken to IOP stabilization if patient i was treated, and time is measured from the time of randomization. The covariate represented by $x_{p+1,i}$ only enters the model for the treated group and when the time of MD measurement happens after patient i had the IOP reduced and stabilized. The LME model with a mean structure given in Equation (6.4) thus fits a two-segment linear trajectory to the treated group with a change point located at the treated individuals' times of IOP stabilization. However, a linear model is fitted for the control group.

Gender, age at baseline and baseline IOP, included as fixed effects, were again found to be insignificant at a 5% level in the PCLIN modelling approach, and the final model with same covariates as in Equation (6.3) was fitted to the data:

$$Y_{ij} = \beta_{0i} + \beta_{1i} x_{1ij} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4ij} + \beta_{5i} x_{5ij} + e_{ij} \tag{6.5}$$

where for patient i,

x_{1ij} is the time of the jth MD measurement from the time of randomization,

x_{2i} is the group membership indicator (0 for control and 1 for treatment),

x_{3i} is the MD level at randomization,

$x_{4ij} = x_{1ij} \times x_{2i}$ defines the time by group interaction,

x_{5ij} is defined as $(x_{1ij} - T_{Si})I(x_{1ij} > T_{Si}) \times x_{2i}$,

e_{ij} is the random error for the jth MD measurement.

Also note that the intercept, the time of repeated measurements and the extra covariate x_{5ij} were included as both fixed and random effects. We were allowing a different post-IOP-stabilization MD decay rate for each treated patient. And for the within-patient errors and the random effects, the continuous first-order autoregressive model and an arbitrary covariance matrix were assumed, respectively. Again, the REML method was used for parameter estimation.

The results from this final model (Table 6.2) confirmed the need for two separate mean time slopes for the treated group. Model diagnostics also showed including the covariate x_{5ij} as random effect significantly improved the modelling of the individuals' MD data. The control group had a more negative estimated mean decay rate of the MD as compared to the results obtained from the ITT and the PP analyses. Before stabilization of the IOP, a treated patient was expected to have a faster decay rate than a control patient, as indicated by the negative estimate of the time by group interaction term. After stabilization, however, the decay rate slowed down considerably. The coefficient estimate for the time from IOP stabilization was not only positive, but also its magnitude was larger than the coefficient for the time by group interaction. As a result, there appeared to be a slower decline of the MD upon IOP stabilization among the treated patients, at a rate slower than the control group.

To compare statistically the rate of change of the MD level between the controls (β_C) and that of the treated patients after IOP stabilization (β_{T2}), we tested the hypotheses:

Table 6.2: REML coefficient estimates of the fixed effects in the LME model for the Three Different Approaches (ITT, PP, PCLIN).

Fixed Effects	ITT		PP		PCLIN	
	Estimate	P-value	Estimate	P-value	Estimate	P-value
β_1 - time	-0.00160	<0.0001	-0.00157	<0.0001	-0.00172	<0.0001
β_2 - group	-0.993	0.005	-0.002	0.995	-0.564	0.136
β_3 - baseline MD	0.834	<0.0001	0.913	<0.0001	0.831	<0.0001
β_4 - time×group	0.0005	0.059	0.0007	0.021	-0.0022	0.025
β_5 - post IOP stabilization	-	-	-	-	0.00307	0.002

$$H_o \quad : \quad \beta_C = \beta_{T2} \quad \text{OR} \quad \beta_4 + \beta_5 = 0 \tag{6.6}$$

$$\text{vs.} \quad H_a \quad : \quad \beta_C \neq \beta_{T2} \quad \text{OR} \quad \beta_4 + \beta_5 \neq 0$$

The corresponding test statistics is

$$T = \frac{\hat{\beta}_4 + \hat{\beta}_5}{\sqrt{Var(\hat{\beta}_4 + \hat{\beta}_5)}} \overset{\text{approx.}}{\sim} N(0,1) \text{ under } H_o$$

$Var(\hat{\beta}_4 + \hat{\beta}_5)$ can be estimated by $\hat{Var}(\hat{\beta}_4 + \hat{\beta}_5)$ which is given by

$$\hat{Var}(\hat{\beta}_4 + \hat{\beta}_5) = \hat{Var}(\hat{\beta}_4) + \hat{Var}(\hat{\beta}_5) + 2 \times \hat{Cov}(\hat{\beta}_4, \hat{\beta}_5)$$

$$= 9.85 \times 10^{-7} + 9.86 \times 10^{-7} + 2 \times (-9.4 \times 10^{-7}) = 9.1 \times 10^{-8}$$

$$\Rightarrow T = \frac{-0.0022 + 0.00307}{\sqrt{9.1 \times 10^{-8}}} = 2.83$$

$$p\text{-value} = 2 \times (1 - \Phi(|T|))$$

$$= 2 \times (1 - \Phi(2.83)) = 0.0046$$

Φ is the standard normal cumulative distribution function.

The small p-value (0.0046) showed that $\hat{\beta}_4 + \hat{\beta}_5$ was significantly different from 0. In particular, while the treated group had a significantly faster decay rate of MD during the

period awaited for IOP stabilization ($\hat{\beta}_4 < 0$, p=0.025 in Table 6.2), the deterioration slowed down to a rate significantly slower than the control rate afterwards.

The individual MD trajectories observed in the data and the mean fitted trajectory from the PCLIN approach for the control and treated groups were plotted, respectively, on the same scale in Figures 6.6 and 6.7. Each thin solid path corresponds to the observed MD level of a patient over the time period he or she was followed, and the thick solid line corresponds to the mean fitted trajectory. In Figure 6.7, the change point of the mean fitted trajectory was chosen to be the mean time to IOP stabilization. From these two plots, we saw a large between-patient variation in the MD readings at randomization as well as in the temporal trend of the MD levels. Although it is not apparent from the individual paths that the treated patients had a different MD slope before and after the IOP stabilization, the mean fitted trajectories for both groups captured the overall trend that was observed in the average MD level curves in Figure 6.5.

We also plotted in Figure 6.8 the individual fitted MD trajectories from the PCLIN approach for a random sample of 24 patients in the data. The plots with a linear trajectory correspond to control patients and those with a broken trajectory correspond to treated patients. The change point of each segmented line is the time of IOP stabilization specific to a treated patient. And the scattered points are the observed MD values. The individual fitted trajectories seem to give a reasonable fit to the data.

In order to assess the performance of the three approaches used to model the MD data, we compared the AIC, BIC indices and the log likelihood values (presented in Table 6.3) from fitting the model given in Equation (6.3) for the ITT and the PP analyses and the one given in Equation (6.5) for the PCLIN analysis.

Indeed, there is no direct method to compare the PP approach with the other two approaches based on the indices and the log likelihood because they essentially involved two different sets of patients. However, we were able to compare the ITT and the PCLIN

Figure 6.6: Individual observed MD trajectories and the mean fitted trajectory from the PCLIN approach from time of randomization for the control group.
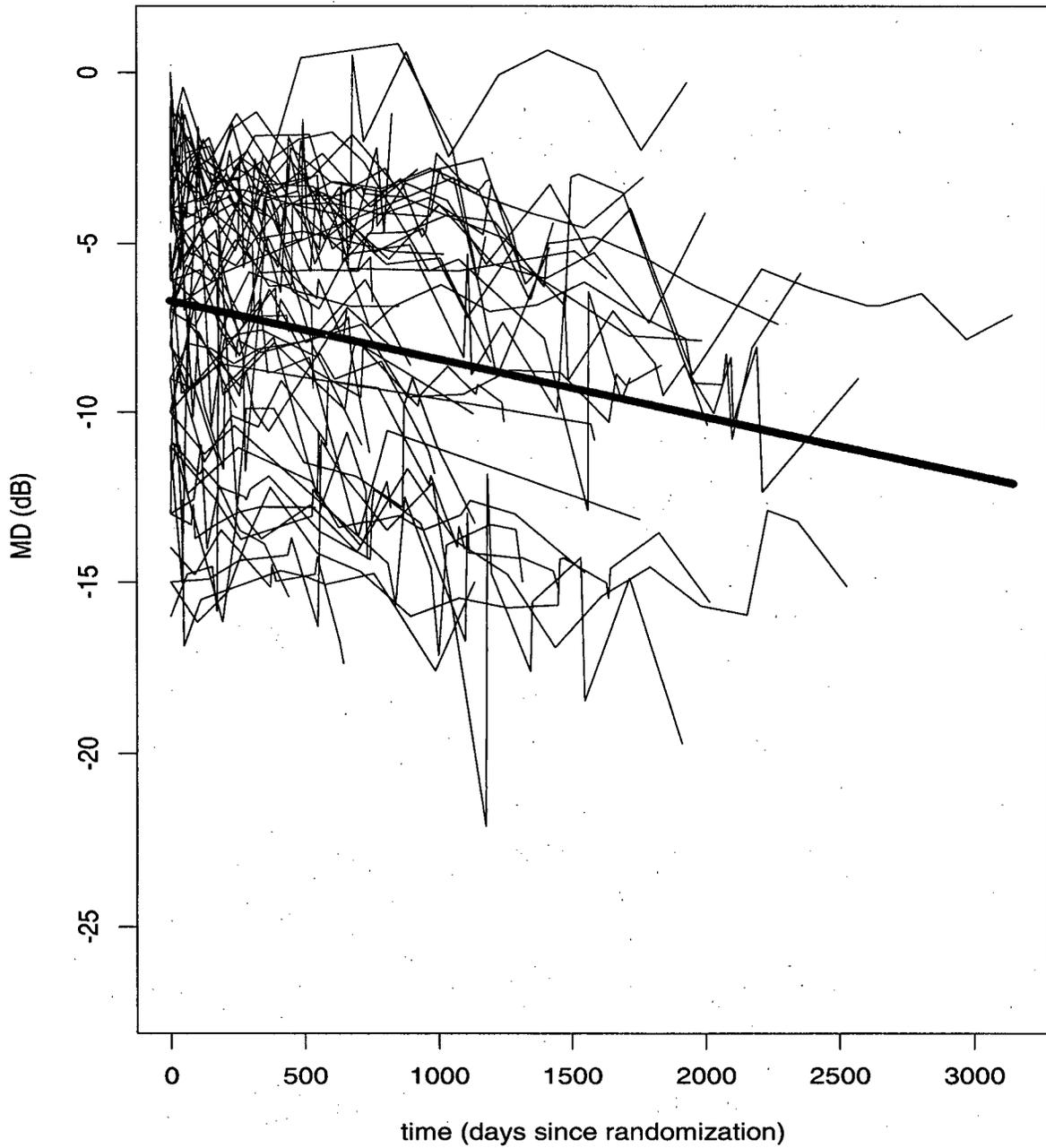
Figure 6.7: Individual observed MD trajectories and the mean fitted trajectory from the PCLIN approach from time of randomization for the treatment group.
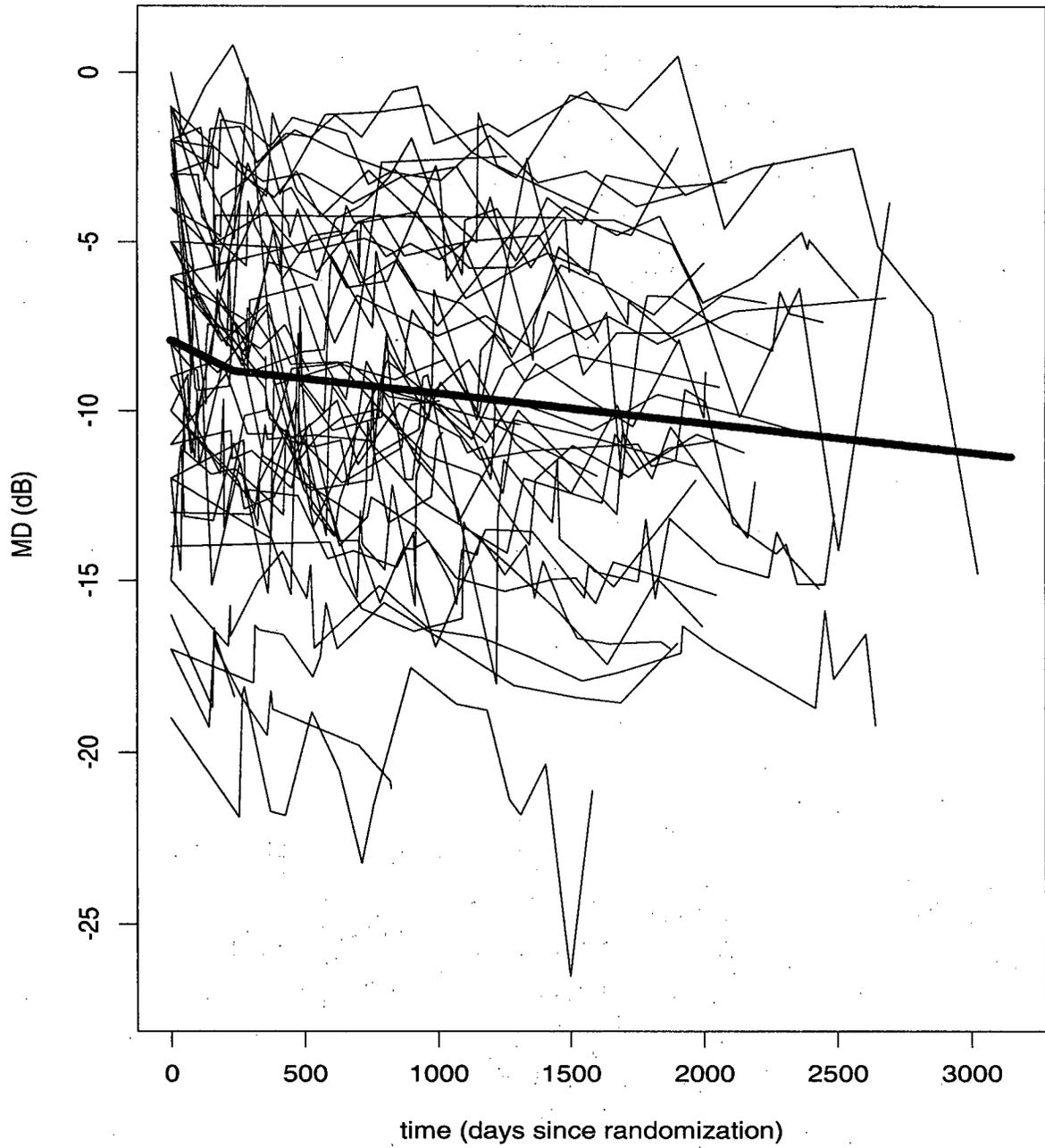
Figure 6.8: A random sample of 24 individual fitted MD trajectories from the PCLIN approach from time of randomization.
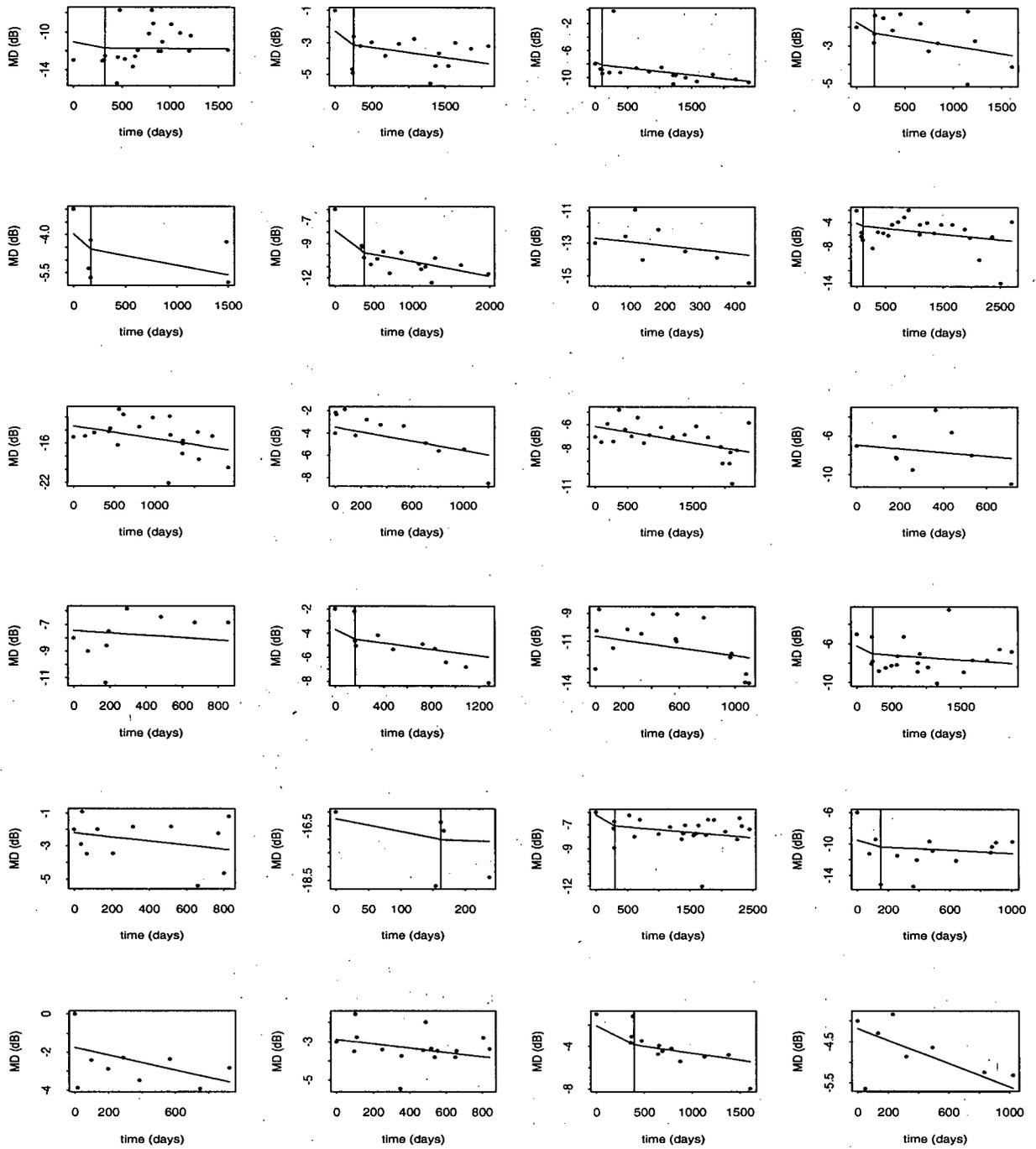
Table 6.3: Table of AICs, BICs, observed log likelihoods and within-patient mean square errors ($\hat{\sigma}^2$) from fitting the LME models for the three different approaches.

Approach	Number of Patients	Number of MD Observations	AIC	BIC	Log Likelihood	$\hat{\sigma}^2$
ITT	97	1493	5883.6	5936.7	-2931.8	2.57
PP	96	1351	5290.4	5342.1	-2635.0	2.38
PCLIN	97	1493	5876.6	5950.9	-2924.3	2.52

models because the former is nested within the latter, and both models were fitted to the same set of MD data. The log likelihood ratio test gave a χ^2 statistic of 14.99 ($= -2 \times \{-2931.8 - (-2924.3)\}$) with 4 degrees of freedom, and a corresponding p-value of 0.0047 which suggested PCLIN modelling approach outperformed the ITT approach in terms of goodness of fit. Besides, the PCLIN approach resulted in a sightly smaller mean square error than the ITT approach.

6.2   Analysis of the Time to Event Data

In Chapter 5, we have described the disability multistate model and its potential application in modelling the survival data for the two events: IOP stabilization and progression simultaneously. In this section, we will present results from the separate analyses of the time to IOP stabilization data and the time to progression data using some classical methods of survival analysis. Later in Section 6.2.3, we will show the results from modelling the individual transition hazards ($h_{12}(t)$ and $h_{02}(t)$ in Figure 5.4) and from the Cox analysis with time-dependent covariates under the multistate modelling approach. Among the 44 treated patients in our data, the time to progression information was missing for one patient and was therefore excluded from the progression survival analysis. The same patient was excluded from the analysis of the time to IOP stabilization data in order to have the same set of treated patients in both analyses.

6.2.1 Analysis of the Time to IOP Stabilization Data

In managing normal tension glaucoma, patients are placed on non-surgical therapy or surgical therapy to lower the IOP level depending on the severity of the disease. It would be useful to have some knowledge about the expected waiting times for the different therapies to effect IOP reduction. This motivated our interest to study the time to IOP stabilization among the treated patients. Note that modelling the hazard for the time to IOP stabilization is essentially the same as modelling the transition hazard $h_{01}(t)$ in the disability model (see Figure 5.4) which was discussed earlier in Chapter 5. We analyzed the time to IOP stabilization data using non-parametric, semi-parametric and parametric methods. The key covariate in the analyses was the type of treatment received. In the Collaborative Normal Tension Glaucoma Study, most patients who underwent the surgical treatment experienced failure of achieving the IOP reduction by means of an initial non-surgical treatment. Only two out of the 43 patients received only the surgical therapy. We thus categorized the treated patients into two groups according to whether they had ever undergone a filtering surgery, and we did not consider having a separate group for those two patients. Among the 43 treated patients included for analysis, 24 received non-surgical treatment only and 19 had undergone a filtering surgery. In regard to the gender composition, the female group had 27 patients and the male group had 16 patients.

Notation

The end point of interest here is the stabilization of the IOP, and for all the treated patients included in our analysis, the time to this event was observed, so there were no censored observations and the analysis was free of concerns about the censoring pattern of the data. For individual i, we defined T_{Si} to be the time taken to IOP stabilization (measured in days) and $\mathbf{X}_i$ to be the vector of covariates. We parameterized the two

categorical variables: gender and the type of treatment received variables in the following way:

- Gender (X_1): 0 for female, 1 for male

- Type of treatment received (X_2): 0 for receiving a non-surgical treatment only, 1 for ever receiving a surgical treatment

Kaplan-Meier Survival Analysis (Non-parametric)

The Kaplan-Meier survivor functions were estimated for and compared between the four different gender by treatment groups $((X_1, X_2) = (0,0),(0,1),(1,0),(1,1))$. The log rank test showed that the four groups had significantly different survival experiences (p=0.00813). However, we observed in Figure 6.9(a) substantial overlapping of the estimated survival curves for the two genders within each of the two treatment groups, which is suggestive of the absence of a gender effect. Individual analyses were carried out to compare the two treatment groups and the two gender groups separately. The results did not show a significant difference in survival experience between the opposite genders (p=0.273). The patients undergoing the filtering surgery had a significantly longer waiting period for IOP reduction and stabilization than those who just received medical or laser treatment (p=0.0054), as can be seen in Figure 6.9(b).

Cox Regression Analysis (Semi-parametric)

Despite the advantage of the Kaplan-Meier analysis which makes no assumption on the distribution of the time to IOP stabilization, the associated log rank test can only assess the effects of categorical covariates. To incorporate other continuous covariates into the modelling of the hazard of IOP stabilization, we made use of the Cox proportional hazards

Figure 6.9: (a) Estimated Kaplan-Meier survivor functions for the time to IOP stabilization of the four gender by treatment groups. (b) Estimated Kaplan-Meier survivor functions for the time to IOP stabilization of the surgical and non-surgical groups.

(a)

(b)
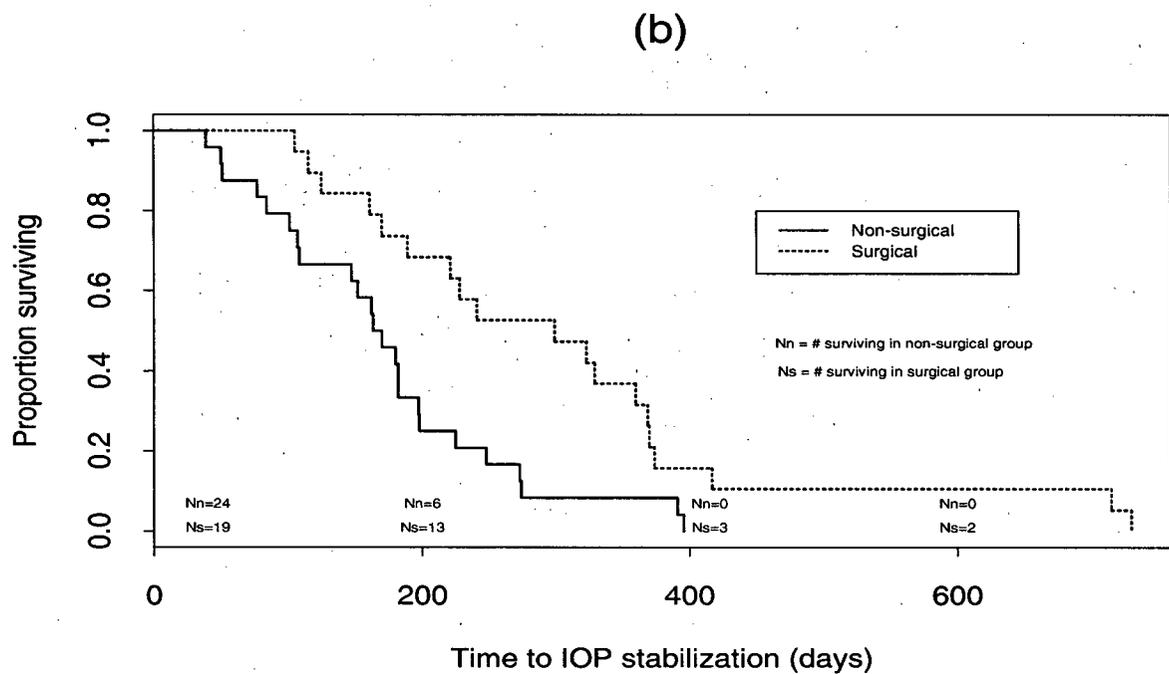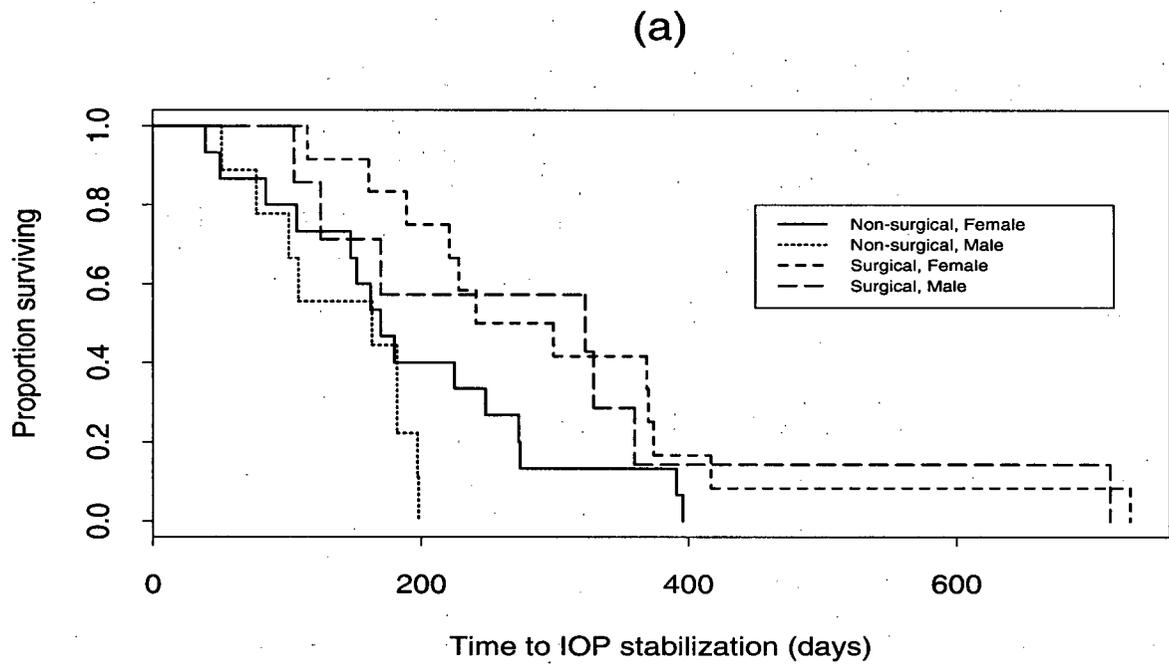
Table 6.4: Estimates of the log relative hazards ratio (β) and relative hazards ratio ($\exp(\beta)$) for the gender and treatment type covariates from the Cox regression analysis of the time to IOP stabilization data.

Covariate	$\hat{\beta}$	$\exp(\hat{\beta})$	SE($\hat{\beta}$)	P-value
Gender (X_1)	0.253	1.288	0.167	0.1300
Treatment type (X_2)	-0.984	0.374	0.343	0.0041

model:

$$h_i(t) = h_0(t)exp\{\beta^T \mathbf{x}_i\} \tag{6.7}$$

where $h_i(t)$ is the hazard function for the ith individual at time t, $\mathbf{x}_i$ is the vector of covariates, $h_0(t)$ is the baseline hazard function, and β is the vector of parameters representing the log relative hazard ratios.

The Cox regression analysis was initially performed to adjust for all the covariates including gender, type of treatment, baseline IOP, MD and age at randomization. The baseline covariates did not have a significant contribution to modelling the hazard at a 5% level. The group and gender covariates were found to be significant, but there was no suggestion of an interaction effect between the two covariates.

We then fitted a reduced model excluding the insignificant covariates, and the results are presented in Table 6.4. Interestingly, the gender effect in the reduced model was deemed insignificant as opposed to the case when fitting the full hazard model at the same significance level. Nevertheless, the positive gender coefficient indicated a higher hazard of IOP stabilization for the male patients, i.e., a female patient who belonged to the same treatment group as another male patient will be expected to wait longer for her IOP level to reduce and stabilize. And again, the surgical group was shown to have a significantly smaller hazard of IOP stabilization. According to the results based on the model with only the type of treatment covariate, the relative hazard of two patients who received and did not receive the surgical therapy was estimated to be 0.406. The

surgical therapy appeared to incur a longer waiting period before IOP stabilized after being reduced by 30%. We also investigated the presence of a possible gender-treatment interaction, which was found to be insignificant at a 5% level.

To check the validity of using the Cox model in Equation (6.7), we tested for any time dependence of the coefficient β for the covariates (individually and globally) in both the full and the reduced models, and we found no evidence of such a dependence. This implies the assumption of proportional hazards was valid for all the covariates used to model the hazard of IOP stabilization.

Parametric Modelling

As one of the classical approaches, parametric models were fitted to complete a full analysis of the time to IOP stabilization data. In seeing that only the type of treatment received and possibly gender played a significant role in determining the length of period taken to reach IOP reduction and stabilization, we included only these two covariates in the parametric models. Among the many distributions commonly used for modelling survival data, we found that the log-normal, Weibull and log-logistic distributions gave satisfactory fits (the probability plots for the three distributions resembled a straight line). Because of the simplicity of the Weibull distribution over the other two, we chose to report the results based on the Weibull hazard model, which can be parameterized as follows:

$$h(t) \quad = \quad \lambda \gamma t^{\gamma-1} \tag{6.8}$$

$$= \quad \{exp(\beta^T \mathbf{x})\} \gamma t^{\gamma-1} \tag{6.9}$$

where $h(t)$ is the hazard function at time t, $\lambda = exp(\beta^T \mathbf{x})$ depends on the vector of covariates $\mathbf{x}$ through an exponential function and linear coefficients in β. The scale parameter λ and the shape parameter γ uniquely define a Weibull distribution.

The gender effect was again shown to be insignificant (p=0.24). Refitting the same model by excluding the gender component yielded an estimated relative hazard between a surgically treated patient over a non-surgically treated patient of exp(-0.581)=0.56 (SE=0.157, p=2.2×10^{-4}).

Regardless of the method used to compare the survival experience of the surgical and the non-surgical groups, patients who ever underwent the filtering surgery to achieve a 30% reduction of IOP had a significantly longer waiting time to IOP stabilization. The presence of a gender effect is ambiguous, although all the three methods of analysis suggested the IOP of the male patients tend to reduce and stabilize faster than the female patients.

6.2.2   Analysis of the Time to Progression Data

Similar to the analysis performed in the previous section, non-parametric, semi-parametric and parametric methods were used to analyze the data for the time to progression. To investigate the effect of the lag time (brought by the delay awaited for IOP reduction and stabilization within the treated group) on the results based on the two different definitions of baseline time, the three methods were applied in both the intent-to-treat and the per-protocol analyses.

In regard to the number of patients to be included in the ITT and the PP analyses, the patient whose time to progression was missing was dropped out of the study. As a result, both analyses excluded this single patient and had a sample size of 96 patients among which 53 belonged to the control group and 43 belonged to the treated group.

Notation

Suppose we define $Y_i^{(1)}$ and $Y_i^{(2)}$ to be the actual times to progression for the ith patient according to definitions of time zero under the intent-to-treat and the per-protocol approaches, respectively. Also let $C_i^{(1)}$ and $C_i^{(2)}$ be, respectively, the censoring times to progression for the two approaches. The control group will have $Y_i^{(1)}=Y_i^{(2)}$ and $C_i^{(1)}=C_i^{(2)}$. The treated group will have $Y_i^{(2)}=Y_i^{(1)} - T_{Si}$ and $C_i^{(2)}=C_i^{(1)} - T_{Si}$, where T_{Si} is the time to IOP stabilization for the ith patient. The time to progression is either observed or censored, and we will define it by $T_i^{(j)}=\min(Y_i^{(j)}, C_i^{(j)})$ for j=1,2. Again, the following relationships hold:

- $T_i^{(1)}=T_i^{(2)}$ for the control group and

- $T_i^{(2)}=T_i^{(1)} - T_{Si}$ for the treated group.

All the time data including $Y_i^{(j)}, C_i^{(j)}$ and $C_i^{(j)}$ (i=1,2,...,96, j=1,2) were measured in years. The time to IOP stabilization T_{Si} was similarly scaled.

We also denote the vector of covariates by $\mathbf{X}_i$. In particular, we define the two categorical variables, gender and the group membership by X_1 (0 for female, 1 for male) and X_2 (0 for control, 1 for treatment), respectively. The other covariates which might have an effect on the time to progression and will be studied include the MD level, the IOP level and the age at baseline.

Kaplan-Meier Survival Analysis (Non-parametric)

For each of the ITT and PP approaches, the Kaplan-Meier survivor functions were estimated for and compared between the four groups of patients of different genders and

group membership $((X_1, X_2) = (0,0),(0,1),(1,0),(1,1))$. The log rank test found a significant difference in the survival experience to progression among the four groups (p=0.0006 for ITT and 0.005 for PP). The plots of the estimated survivor functions (Figure 6.10) showed substantial overlapping of the two gender curves within the same treatment or control group. The gender effect did not seem to explain the difference in the survival experience of the four groups. When we compared only the two treatment groups, both the ITT and the PP approaches resulted in a significant group effect. The plots of the estimated Kaplan-Meier survival curves (Figure 6.11) showed a superior survival experience of the treated group, and the two curves did not intersect over the whole course when the patients were followed. Upon comparison of the results of the ITT and the PP approaches, the log rank test demonstrated a slightly more significant difference in survival between the two treatment groups in the ITT analysis (p=0.0001 for ITT versus p=0.0009 for PP). This can also be seen from the larger separation between the two estimated Kaplan-Meier survivor functions in Figure 6.11(a) as compared to the two functions in Figure 6.11(b). In fact, the estimated Kaplan-Meier survivor function for the control group was the same under the two approaches because the baseline defined for the controls was always the time of randomization. The baseline for the treated group was shifted from the time of randomization to the individuals' times of IOP stabilization when we took the PP approach. The time to progression as defined in the PP approach was shorter than its ITT counterpart within the treated group, and hence a diminished difference between the two groups was demonstrated in the PP analysis . Although the treatment effect was found to be significant in both analyses, the PP analysis gave a slightly weaker evidence against the null hypothesis of no treatment difference. Unlike the case for the control group, the difference in the baseline definition adopted by the two approaches led to different Kaplan-Meier estimates for the treated group. The Kaplan-Meier estimate of the survivor function for the treated group did not simply undergo a rank invariant transformation when switching between the ITT and the PP approaches, because a different time awaited for IOP stabilization, T_{Si}, was subtracted from $T_i^{(1)}$ to obtain $T_i^{(2)}$. The ranks of the set of $T_i^{(1)}$'s are not preserved under the transformation

$$T_i^{(2)} = T_i^{(1)} - T_{Si}.$$

As mentioned earlier, the overlapping of the estimated survivor functions for the two gender groups within the same treatment or control group (Figure 6.10) suggested that there was no gender effect. Nevertheless, we carried out separate analyses to test for the treatment effect within each gender group and to see if they gave different results for the two genders. We compared the treated female group to the control female group $((X_1, X_2)$ = (0,0) and (0,1)), and the treated male group to the control male group $((X_1, X_2)$ = (1,0) and (1,1)), using both the ITT and the PP approaches. Interestingly, the ITT analysis revealed a difference in the treatment effect on the two genders. The log rank test demonstrated a significant difference in the time-to-progression experience between the treated female and control female patients (p=0.00064), but there was no evidence of a treatment effect within the male patients (p=0.137). On the other hand, the PP analysis showed no treatment effect in either gender group at a 5% level (p=0.471 for males and p=0.0583 for females) although the treatment effect for the female group was marginally significant. This contradicts with the results obtained from the PP analysis when we did the treatment comparison on all patients together.

Cox Regression Analysis (Semi-parametric)

Besides obtaining the non-parametric Kaplan-Meier estimates of the survivor functions, we modelled directly the hazard of progression using the Cox proportional hazards model (Equation (6.7)). Later, we repeated the Cox regression analysis within each of the two gender groups separately. We also considered the stratified Cox model with gender being the strata variable to model the progression hazard.

When we first modelled the progression hazard for all the 96 patients using the Cox proportional hazards model (Equation (6.7)), the patients' age, MD level and IOP level at baseline, gender and group membership were included as covariates. In this full model,

Figure 6.10: (a) Estimated Kaplan-Meier survivor functions for the time to progression of the four treatment-gender groups based on the intent-to-treat (ITT) approach. (b) Estimated Kaplan-Meier survivor functions for the time to progression of the four treatment-gender groups based on the per-protocol (PP) approach.
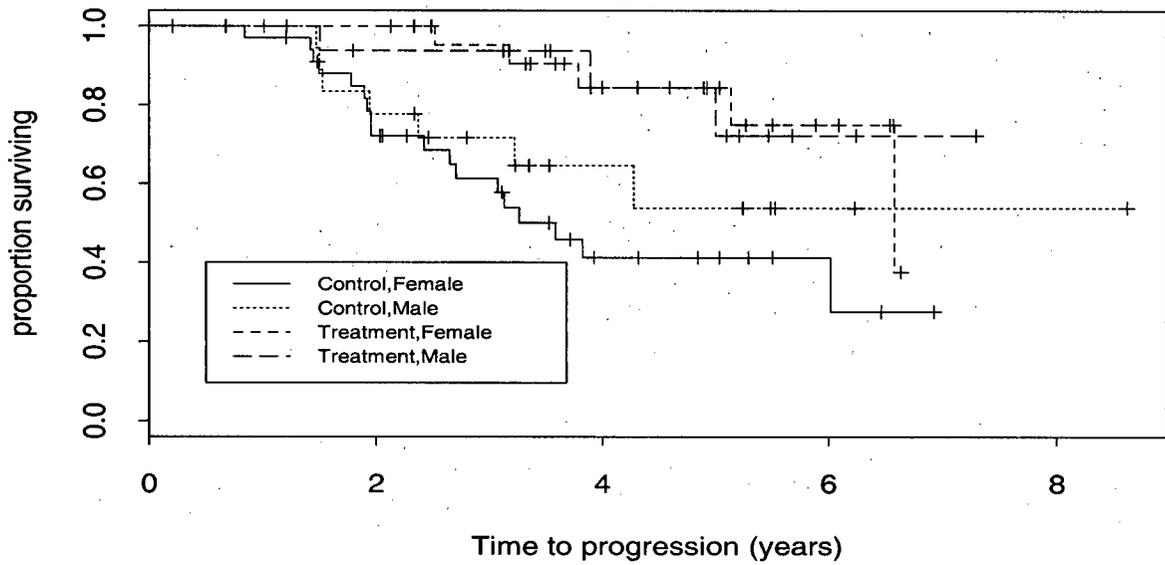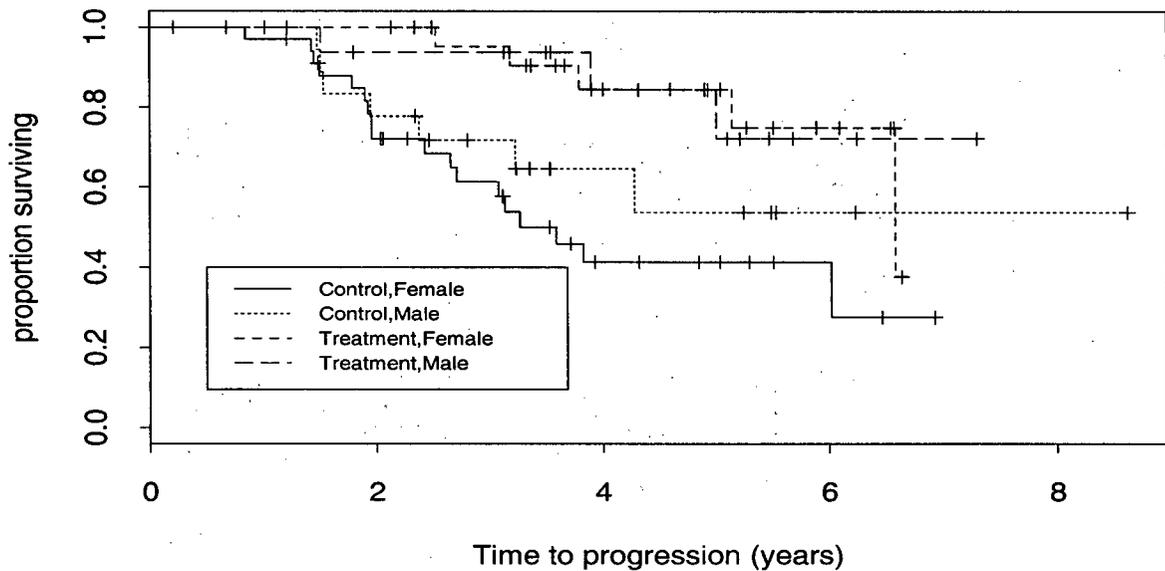
Figure 6.11: (a) Estimated Kaplan-Meier survivor functions for the time to progression of the two treatment groups based on the intent-to-treat (ITT) approach. (b) Estimated Kaplan-Meier survivor functions for the time to progression of the two treatment groups based on the per-protocol (PP) approach.
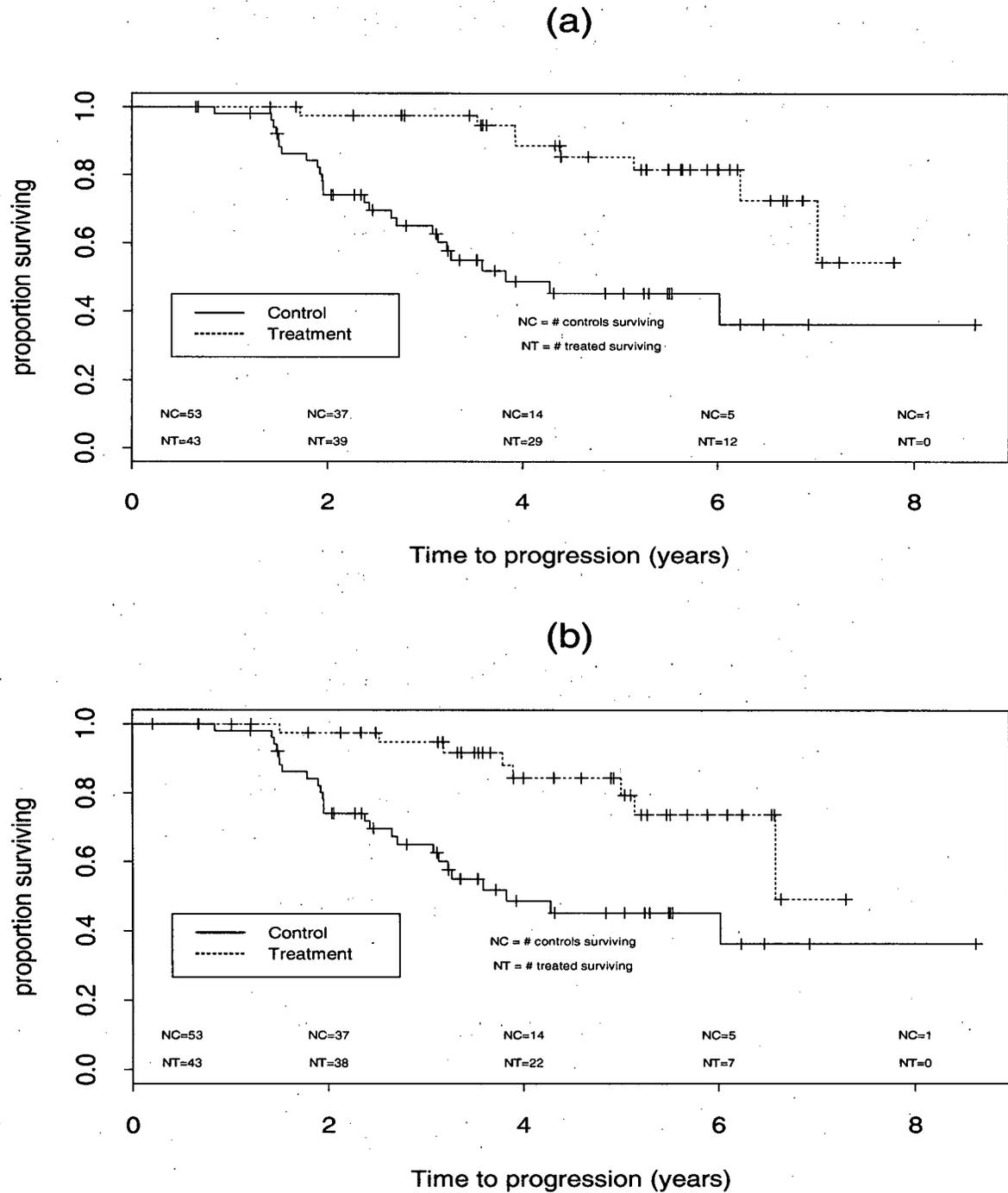
Table 6.5: Estimates of the log relative hazards ratio (β) and relative hazards ratio ($\exp(\beta)$) for the group covariate from the gender-specific Cox regression analysis of the time to progression data (M for males, F for females).

		$\hat{\beta}$		$exp(\hat{\beta})$		$SE(\hat{\beta})$		P-value	
Covariate		M	F	M	F	M	F	M	F
Group ITT		-1.12	-1.83	0.33	0.16	0.70	0.56	0.11	0.0011
PP		-0.98	-1.43	0.38	0.24	0.69	0.51	0.16	0.0049

the baseline values of age, MD level and IOP level were adjusted according to the ITT or the PP approach we used to analyze the data. Upon fitting the full model, we found that the group covariate satisfied the proportional hazards assumption in the PP analysis but not in the ITT analysis. It can be that the two treatment groups had proportional hazards within each gender group while the two genders had different baseline hazards. We explored this possibility by analyzing the time to progression data separately for the two genders. If indeed the assumption for the group covariate is valid within each gender group, it is also legitimate to fit a Cox model stratified by gender.

For each of the ITT and the PP approaches, we fitted the full model separately to the time to progression data specific to the two genders. In all the resulting four analyses, the proportional hazards assumption held for all of the covariates individually as well as globally. Similar to the case in the gender-specific Kaplan-Meier survival analyses presented earlier, we obtained different results for the two gender groups. Both the ITT and the PP approaches revealed a difference in the treatment effect on the two genders. More specifically, there did not seem to be a difference in the hazard of progression between the male control patients and the male treated patients. However, in the female group, the treated patients had a significantly smaller risk of progression than the control patients. The age covariate, IOP level and MD level at baseline were not associated with the hazard of progression for either gender group. We repeated the gender-specific analyses, fitted a reduced model involving only the group covariate and obtained the results in Table 6.5.

Again, the assumption of proportional hazards between the control and the treated groups was satisfied within each gender group under both the ITT and the PP approaches. Neither the ITT nor the PP analyses showed enough evidence against the null hypothesis of no treatment effect among the male patients. In contrast, the female patients who received the treatment had a significantly smaller risk of progression than those who remained untreated, as shown in both the ITT and the PP analyses. The ITT analysis gave a relative hazard of 0.16 for a treated female patient versus an untreated female patient, while the PP analysis gave a relative hazard of 0.24. In other words, within the female group, receiving an IOP reduction treatment is associated with about a 6-fold decrease in the risk of progression, as suggested by the ITT analysis, and roughly a 4-fold decrease in the PP analysis. Furthermore, the ITT approach gave a more significant result for the treatment effect on the female patients than the PP approach did. The same argument given in the Kaplan-Meier survival analysis (p.54) in this section accounted for this difference in significance. However, it is not clear how the shift in baseline within the treated group affects the assessment of a treatment effect when indeed such an effect does not exist.

As the proportional hazards assumption for the group covariate was valid within patients of the same gender, modelling the hazard of progression using the Cox model stratified by gender was justified. In essence, we were assuming two different baseline hazards for the two genders while the two genders share the same coefficient for each of the other covariates. The stratified Cox model has the form

$$h_{ij}(t) = h_{bj}(t)exp\{\beta^T \mathbf{x}_i\}, \quad j = 1, 2 \tag{6.10}$$

where $h_{ij}(t)$ is the hazard at time t for the ith individual of gender group j, $h_{bj}(t)$ is the baseline hazard of the jth gender group, $\mathbf{x}_i$ is the vector of covariates at time t not including gender, and β is the vector of parameters representing the log relative hazards ratios.

In fitting the stratified Cox model, we included the group covariate and its interaction

Table 6.6: Estimates of the log relative hazards ratio (β) and relative hazards ratio ($\exp(\beta)$) from the stratified (by gender) Cox regression analysis of the time to progression data.

Covariate		$\hat{\beta}$	$\exp(\hat{\beta})$	SE($\hat{\beta}$)	P-value
Group	ITT	-1.472	0.229	0.448	0.001
	PP	-1.203	0.300	0.43	0.005
Group×Strata(gender)	ITT	0.356	1.428	0.448	0.43
	PP	0.227	1.255	0.43	0.60

with the strata covariate, i.e., gender, and the results are presented in Table 6.6. It can be shown that by adding the group by strata interaction term to the hazard model, the result is identical to performing separate fits with only the group covariate for each stratum. For instance, the ITT log relative hazards ratio estimate ($\hat{\beta}$) for the group covariate within the male patients, which is -1.12 (Table 6.5), can be obtained by summing the ITT estimates of the log relative hazards ratios of the group covariate and the interaction term (-1.472+0.356) presented in Table 6.6. For the female patients, taking the difference of the group covariate estimate and the interaction estimate obtained by the stratified Cox regression would give the log relative hazards ratio estimate for the group covariate in the analysis focusing on only the female patients.

Despite the fact that the gender-specific ITT and PP analyses showed a significant treatment effect in the female group but not in the male group, both the ITT and the PP stratified Cox regression analyses did not suggest the presence of an interaction effect between gender and the treatment group membership (see Table 6.6). We then refitted the stratified Cox model leaving out the interaction term. The ITT approach gave a group coefficient estimate of -1.58 (p=0.0003) while the PP approach gave an estimate of -1.28 (p=0.0017). The negative coefficient implied a smaller risk of progression for the treated group and hence a favourable treatment effect. Based on the ITT analysis, the relative hazard of progression of a treated patient versus an untreated of the same gender was estimated to be exp(-1.58) = 0.206. Similarly, the PP analysis gave a relative hazard

of exp(-1.58) = 0.278.

While the stratified Cox model has the flexibility of incorporating into the hazard covariates which only satisfy the proportional hazards assumption within certain strata, the approach has certain limitations in our application. First, we cannot assess the gender effect with the stratified model because no direct estimate of the gender coefficient is available. In fact, because a different baseline hazard function is used for each gender, the relative hazard between the two genders will be time-dependent and cannot be determined unless the exact form of the baseline hazards is known. Comparison of the survival experience between two patients of opposite genders thus cannot be easily made. Second, the same coefficient for group, β_{group}, is assumed for both genders. The value of $\exp(\beta_{group})$ gives an "overall" measure of the relative risk of a treated patient versus an untreated patient of the same gender. As a result, the estimated relative risk of a treated patient versus an untreated patient within the female group is the same as that within the male group. Assuming the same relative risk for the two genders might be unrealistic in our situation where the two genders have essentially different baseline hazards.

Parametric Modelling

Various distributions common for modelling survival times were fitted to the data. The Weibull, log normal and log-logistic distributions seemed to fit the data satisfactorily and gave small and comparable deviance values. Here, we present the results from the Weibull fit because it is simpler to interpret than the other two fits. In both the ITT and the PP approaches, only the group covariate was fitted in the hazard model given by Equation (6.9), and the two corresponding group coefficients were both shown to be significantly different from zero. A coefficient estimate of -0.878 from the ITT analysis implied that the treated group and the control group had a relative hazard of exp(-0.878) = 0.416 and thus a favourable treatment effect. Similarly, the PP analysis showed a superior effect for the treated group whose hazard was estimated to be exp(-0.754) = 0.47 times of the

hazard for the control group. The results were highly significant in both the ITT and PP analyses although the former analysis gave slightly smaller p-value (p=4.96×10^{-4} for ITT and p=2.87×10^{-3} for PP).

All three methods of analysis showed the treatment significantly prolonged the time to progression, regardless of which definition of the baseline was used. Other covariates did not contribute to the modelling of the survival time or the hazard.

6.2.3 Analysis Using the Disability Model

In Section 6.2.1, we have presented the results of modelling the transition hazard $h_{01}(t)$ in the disability model (see Figure 5.4), i.e., the hazard of IOP stabilization among the treated patients. For the transition hazard $h_{12}(t)$ in the model, we could think of it as the post-IOP-stabilization hazard of progression within the treated patients. Treated patients were at risk for the transition from state 1 to state 2 starting from the time of IOP stabilization until they reached the progression end point. Thus, the time t in $h_{12}(t)$ was measured from the time of IOP stabilization. And for $h_{02}(t)$, we could interpret it as the progression hazard among patients in the control group and patients who were treated but progressed before a stable IOP reduction was achieved. Because data from the latter group were unavailable for our analysis, we focused on the control patients only in modelling $h_{02}(t)$. Control patients were at risk for the transition from state 0 to state 2 starting from the time of randomization until they reached the progression end point. The time t in $h_{02}(t)$ was measured from the time of randomization.

For both the modelling of the transition hazards $h_{12}(t)$ and $h_{02}(t)$, we have applied non-parametric, semi-parametric and fully parametric methods, and similar results were obtained from the different approaches. In the following two sections, we will focus on the results from fitting the Cox proportional hazards model given in Equation (6.7). We checked that the proportional hazards assumption was satisfied, and therefore modelling

$h_{12}(t)$ and $h_{02}(t)$ using the Cox model seemed appropriate.

Modelling of $h_{12}(t)$ within the Treated Group

An initial full Cox model with covariates including gender, age, MD and IOP levels at IOP stabilization, the type of IOP lowering therapy and the time awaited for IOP stabilization from the time of randomization was fitted to the 43 treated patients. Although the data did not show a significant effect of any of the covariates at a 5% level, we obtained a negative coefficient for both the time taken to reach IOP stabilization ($\hat{\beta}$=-0.26, p=0.85) and the type of IOP lowering therapy ($\hat{\beta}$=-1.46, p=0.24). The coefficient estimates suggested that a treated patient who took a longer time to achieve a stable IOP reduction had a smaller risk of progression, and the same applied to a patient who received the surgical treatment (type of therapy was parameterized as 0 for receiving only non-surgical treatment and 1 for ever receiving a surgical treatment). The effect of IOP at stabilization (p=0.11) may have been diluted by the other insignificant covariates included in the model, so we fitted a reduced model with only the IOP at stabilization covariate. The fit gave a coefficient estimate of 0.37 (p=0.074) and a relative hazard ratio of exp(0.37)=1.44. Though only marginally significant, the coefficient with a positive estimated value implied a higher IOP at stabilization was associated with a higher risk of progression among the treated patients. We also tested for the presence of an interaction between various pairs of covariates, but no significant results were obtained.

Modelling of $h_{02}(t)$ within the Control Group

A full Cox model with covariates including gender, age, MD and IOP levels at randomization was fitted to the 53 control patients, and no significant covariates were identified at a 5% level. Unlike in the analysis of the post-IOP-stabilization survival experience

within the treated patients where IOP at stabilization had a marginally significant effect on the post-stabilization progression hazard, IOP at randomization did not seem to affect the hazard of progression within the control group (p=0.25). There was also no evidence of a significant interaction effect between any pairs of covariates included in the full model.

Cox analysis with time-dependent IOP-stabilization indicator variable

We explored the possibility of fitting a single Cox model with time-dependent covariates for the analysis of the disability model as described in Chapter 5. Recall that if the length of the waiting period to achieve IOP stabilization does not have an effect on the risk of progression with a stable IOP reduction, and if we have proportionality between the two progression hazards $h_{12}(t)$ and $h_{02}(t)$ at all times, it would be legitimate to model an overall hazard by means of a Cox model with a time-dependent indicator variable for IOP stabilization [26]. The above analysis of the post-stabilization times to progression within the treated patients showed that the length of the time period awaited for IOP stabilization had no significant effect on the post-stabilization time to progression at a 5% level (p=0.85) and thus validated the first condition required for fitting a single Cox model. This first condition also held within each gender group (p=0.18 for male patients and p=0.69 for female patients). Since our data did not include any treated patients who reached the progression end point before IOP stabilization was established, we did not have a proper modelling of the transition hazard $h_{02}(t)$ which, in principle, should involve complete uncensored time to progression without IOP stabilization from some patients in both the untreated and treated groups. We were therefore unable to check the second condition (proportionality of the progression hazards $h_{12}(t)$ and $h_{02}(t)$). We proceeded by assuming the second condition held in our case and fitted the Cox model (in Equation (5.2)) with covariates including gender, age, MD and IOP at randomization, as well as the interaction term between the group covariate $x_{group,i}$ (0 for control group and 1 for treated

group) and the time-dependent indicator for IOP stabilization $I_i(t)$. As all the treated patients included in our data underwent IOP stabilization before reaching the progression end point, they had $x_{group,i} = 1$ by definition, and $I_i(t) = 1$ during their observed post-stabilization periods. The two variables $x_{group,i}$ and $I_i(t)$ became linearly dependent, and can lead to computational problems in the estimation of the regression coefficients for the two variables. To avoid the problem, the group main effect was excluded from the model, and consequently, we could only assess the effect of IOP stabilization on the progression hazard within the treated patients, and we were unable to assess whether the treated and untreated groups had different risks of progression.

At a 5% significance level, gender, age, MD and IOP at randomization were found to have no effect on the hazard of progression. Among the treated patients, the IOP-stabilization covariate was found to be significant ($p=7.6\times10^{-4}$), and the relative risk of having versus not having a stable 30% IOP reduction was estimated to be exp(-1.41)=0.24. This implied about a four-fold decrease in risk of progression when a stable IOP reduction was achieved. Moreover, the proportional hazards assumption held for all the time-independent covariates, so the appropriateness of using a Cox model was justified. Checking for the assumption for the IOP-stabilization indicator variable $I_i(t)$ was unnecessary because the hazard ratio involving the $I_i(t)$ term changes with time. Indeed, the property of proportional hazards is not required for time-dependent covariates for the validity of applying the Cox model in Equation (5.2).

We also fitted the same Cox model for the two genders separately. The analyses showed no evidence of a significant effect of any of the baseline covariates at a 5% level within each gender group. We then fitted a reduced model with only the IOP-stabilization covariate and the results are presented in Table 6.7. Among the male treated patients, the effect of having a stable IOP reduction on the progression hazard was insignificant although a negative estimate of β(IOP stabilization) suggested that a stabilized 30% reduction of the IOP was associated with a smaller risk of progression. On the other hand,

Table 6.7: Estimates of the log relative hazards ratio (β) and relative hazards ratio ($\exp(\beta)$) for the time-dependent IOP stabilization covariate within the treated patients from the gender-specific time-dependent Cox regression analysis (M for males, F for females).

Covariate	$\hat{\beta}$		$exp(\hat{\beta})$		$SE(\hat{\beta})$		P-value	
	M	F	M	F	M	F	M	F
IOP Stabilization	-1.06	-1.79	0.35	0.17	0.70	0.56	0.13	0.0014

a stable reduction of the IOP had a significant effect on the progression hazard among the female treated patients. Within the treated group, female patients who attained IOP stabilization had a hazard of about 6 times smaller than those who did not.

A Cox model stratified by gender was also fitted, and we again found that the baseline covariates were all insignificant at a 5% level. Refitting the stratified Cox model with only the IOP-stabilization indicator variable gave an estimate of β(IOP stabilization) of -1.54 (p=0.0004) which indicated a favourable effect of achieving a stable 30% reduction of the IOP from the prerandomization readings within the treated group. Based on this estimate, the relative risk of progression of two treated patients (of the same gender) with and without IOP stabilization was $\exp(-1.54) = 0.215$, which in turn implied an almost five-fold decrease in the risk of progression after treatment stabilization.

6.2.4 The Baseline-Adjustment Analysis

In Section 6.2.1, the Weibull distribution was found to model the time to IOP stabilization (T_{Si}) within the treated group satisfactorily. To correspond to the lag time in treatment stabilization in the treated group, an adjustment was made to the baseline for the control group. Our baseline-adjustment analysis introduced a random Weibull time shift from the time of randomization for each of the control patients in the data. Let T_{Ci} represent the random time shift for the ith patient if he or she happened to be untreated. The

variable T_{Ci} follows the Weibull distribution (see parameterization in Equation (6.9)) with estimated parameters $\hat{\lambda}=0.572$ and $\hat{\gamma}=1.68$ for the time t measured in years. The Weibull parameters were estimated from fitting the Weibull distribution to the times to IOP stabilization observed within the treated patients in our data without adjusting for any covariate information. The therapy type was the only significant covariate identified in the analysis of the time to IOP stabilization data but was not adjusted in our baseline-adjustment approach because the therapy type was irrelevant for the control patients. The Weibull distribution with no covariate adjustment was shown to also provide a reasonable fit to the time to IOP stabilization data. For each control patient, a random T_{Ci} was generated and was subtracted from the patient's original time to progression to obtain the adjusted time to progression. The values of the baseline covariates including MD, IOP and age were changed accordingly to match the shifted baseline. The time to progression for each treated patient will be the same as defined under the PP approach, i.e., the treated patients were followed from their individual time of IOP stabilization. We then carried out classical survival analysis on the new set of data with the adjustment made to the control group.

Kaplan-Meier Survival Analysis (Non-parametric)

Kaplan-Meier estimates of the survivor functions were obtained for and the time-to-progression experience was compared between the four treatment-gender groups. The log rank test showed a significant difference between the survivor functions of the four groups, but the difference did not seem to be explained by the gender effect because the estimated Kaplan-Meier survival curves for the two genders (Figure 6.12(a)) crossed within each of the two treatment groups. Especially within the treated group, the curves of the opposite genders overlapped considerably. The log rank test comparing the survivor functions for the two genders only did not find a gender effect (p=0.335). We then focused on just the two treatment groups and plotted the estimated Kaplan-Meier survivor curves

Table 6.8: Estimates of the log relative hazards ratio (β) and relative hazards ratio ($\exp(\beta)$) for the group covariate from the gender-specific Cox regression analysis under the baseline-adjustment approach (M for males, F for females).

Covariate	$\hat{\beta}$ M	$\hat{\beta}$ F	$\exp(\hat{\beta})$ M	$\exp(\hat{\beta})$ F	$SE(\hat{\beta})$ M	$SE(\hat{\beta})$ F	P-value M	P-value F
Group	-1.2	-1.81	0.3	0.16	0.70	0.56	0.087	0.0012

in Figure 6.12(b). From the plot, we observed a prolonged time to progression for the treated group, and such a favourable treatment effect was confirmed by the log rank test ($p=9.4\times10^{-5}$).

Cox Regression Analysis (Semi-parametric)

A full Cox proportional hazards model (given in Equation (6.7)) with all the covariates including gender, treatment group membership and baseline IOP, MD and age was first fitted to the adjusted data. All the covariates except the group variable were found to be insignificant at a 5% level. There was also no evidence of a group by gender interaction effect. Refitting the Cox model with only the group covariate again showed a significant treatment effect, but the assumption of the control and the treated groups having proportional hazards was invalid. We proceeded with performing separate Cox regression analyses for the two genders, and the proportional hazards assumption for every covariate that was considered in the full and the reduced models, as described below, was satisfied. A full model with all the relevant covariates including treatment group membership, baseline IOP, MD and age was fitted to the data from each gender group, and all the baseline covariates remained insignificant at a 5% level. The results from fitting the reduced model with only the group covariate for each gender are presented in Table 6.8.

Similar to the results obtained in the gender-specific analyses under the ITT and

Figure 6.12: (a) Estimated Kaplan-Meier survivor functions for the time to progression of the four gender by treatment groups based on the baseline-adjustment approach. (b) Estimated Kaplan-Meier survivor functions for the time to progression of the treated and the control groups based on the baseline-adjustment approach.

(a)

(b)
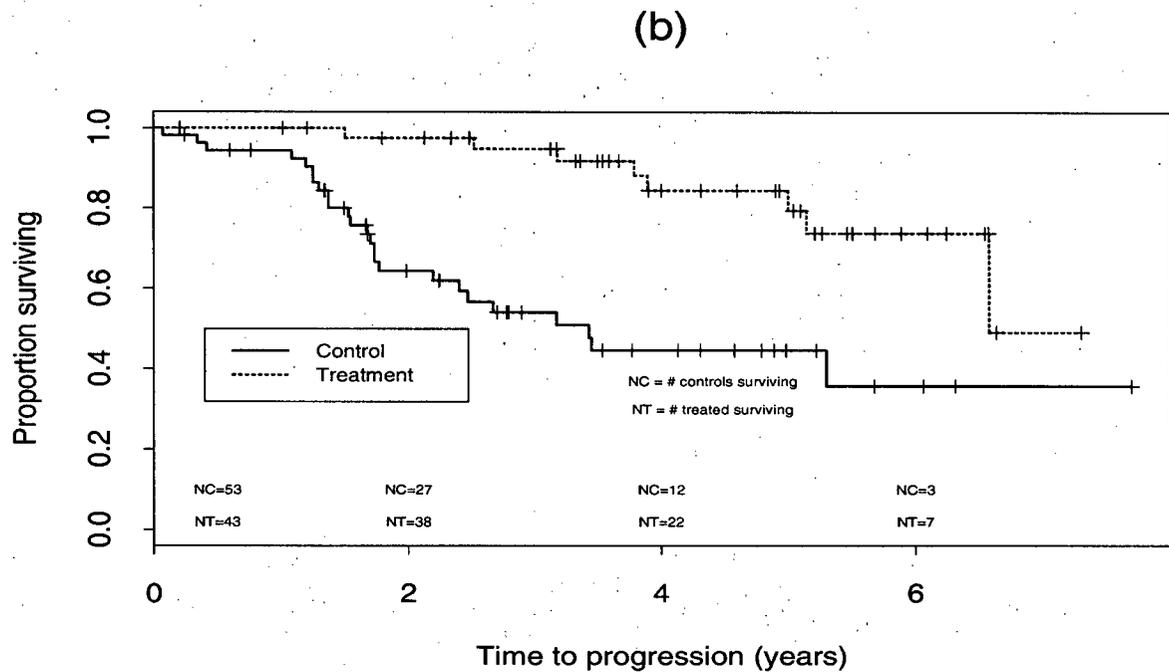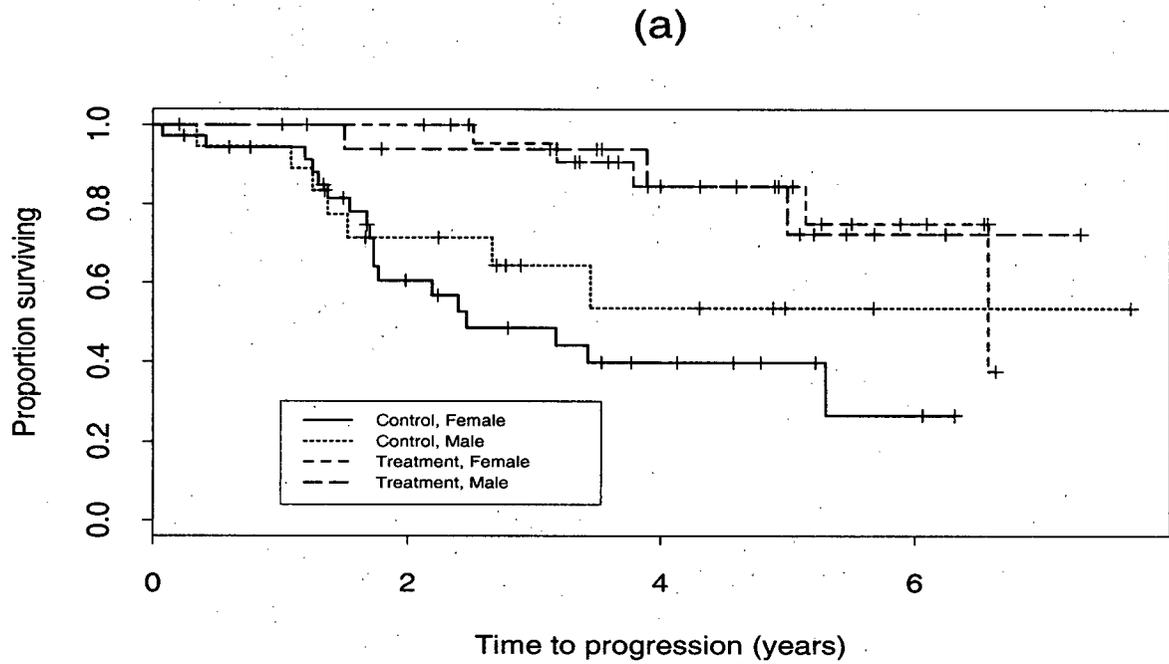
the PP approaches, the adjusted data showed a significant favourable effect of a 30% IOP reduction within the female group. The male gender seemed to benefit from the treatment, but the evidence was not strong enough to be deemed significant at a 5% level. Based on the results presented, the relative hazard of a treated female patient versus an untreated was estimated to be 0.163, which implied an approximately six-fold decrease in the risk of progression for having a 30% IOP reduction. We also carried out a stratified Cox regression analysis with gender as the strata variable. The stratified Cox model (in Equation (6.10)) did not identify a significant effect of any of the baseline covariates and the strata by group interaction. The group covariate remained significant ($p=2.2\times10^{-4}$), and the relative hazard of a treated patient versus an untreated patient was estimated to be $\exp(-1.6) = 0.202$.

Parametric Modelling

Different parametric models were fitted to the adjusted data, and the Weibull, log-normal and the log-logistic distributions provided satisfactory fits. We again chose to report the results based on the Weibull fit. After adjusting for the group membership covariate, the Weibull hazard model (given in Equation (6.9)) fitted had an estimated group coefficient of -1.13 and hence a relative hazard of $\exp(-1.13) = 0.32$ for a treated patient versus an untreated patient. The group coefficient was significant ($p=6.35\times10^{-4}$) and in particular, the risk of progression was reduced by two thirds for having a 30% IOP reduction as compared to the case when untreated.

The different methods of analysis of the mean defect data and the time to event data will be compared, and the results from each will be further discussed in Chapter 8.

Chapter 7

Simulation

7.1 The Objectives

In Chapter 6, we have seen that the intent-to-treat (ITT) and the per-protocol (PP) analyses of the mean defect (MD) data led to quite different results. Such differences can possibly be explained by the rapid decay of MD upon receiving the treatments by the treated patients, which had not been accounted for in both the ITT and the PP analyses.

An immediate concern would be how much of an impact this first phase of change within the treated group prior to IOP stabilization has on the estimates of the decay rates, the overall level of the MD and the significance of the time by group interaction effect in the ITT analysis. Consequently, the ITT and the PP analyses might lead to different results. One of the objectives of carrying out a simulation experiment here is to investigate how the results from an ITT analysis and those from a PP analysis compare in the presence of an initial phase of deterioration. Another objective is to study how the PP and the PCLIN modelling approaches perform in comparing the post-treatment-stabilization rates of progression between the treated and the untreated groups. We simulated MD data for ninety-seven patients to comprise our sample which was the same size as in the original data. The data were simulated based on the linear mixed effects (LME) model fitted under the piecewise linear (PCLIN) modelling approach, as given in Equation (6.5), using different sets of values for the decay rates of the MD for the control and the treated groups. We chose this particular LME model to be the base

model for the simulation because it gave the best fit to the original MD data among the three different approaches that were presented in Chapter 6.

7.2 Generating Mean Defect Data

The data generation involved two major steps: the generation of the mean structure of MD specified by the fixed effects, and the generation of the random effects and the within-patient errors. The model for simulating the data has been given by Equation (6.5):

$$Y_{ij} = \beta_{0i} + \beta_{1i}x_{1ij} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4ij} + \beta_{5i}x_{5ij} + e_{ij} \qquad (7.11)$$

Here, i is the patient index and j is the index for the repeated measurements within each patient. The response Y_{ij} represents the MD measured at time x_{1ij}. Besides the time covariate, the other four covariates are the group membership (x_{2i}), the MD at randomization (x_{3i}), the time by group interaction ($x_{4ij} = x_{1ij} \times x_{2i}$) and the piecewise component (x_{5ij}) which enables the fitting of a segmented MD trajectory with a change point at the treated patients' individual times of IOP stabilization for the treated group and a linear trajectory for the control group. The piecewise component has been defined as

$$x_{5ij} = (x_{1ij} - T_{Si})I(x_{1ij} > T_{Si}) \times x_{2i}$$

where T_{Si} denotes the time taken to IOP stabilization if the ith patient belongs to the treated group; otherwise, T_{Si} is unobserved. And for $k = 0, 1, 5$, β_{ki} can be expressed as the sum of the fixed effect term β_k and the random effect term b_{ki}. For patient i (i = 1, 2,...,97), the random effects follow the normal distribution with a zero mean and a covariance matrix $\mathbf{D}$ common to all of the patients. The within-patient errors, e_{ij}'s are also normally distributed with a zero mean and a covariance structure $\mathbf{R}_i$. Hereafter, we

will refer to the fit to the original data based on the model in Equation (7.11) as "the best fit" unless stated otherwise.

7.2.1 Generation of the Mean Structure Specified by the Fixed Effects

For the simulation of the mean structure of the MD data for the ith patient, we specified the part of the model (in Equation (7.11)) involving only the fixed effects, i.e., $\mathbf{X}_i \boldsymbol{\beta}$ whose jth component is:

$$\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4ij} + \beta_5 x_{5ij}$$

The same design matrix $\mathbf{X}_i$ and the time to IOP stabilization T_{Si} as observed in the original data were used in the simulation. The times of repeated MD measurements, the group membership, the initial MD reading at randomization in the simulated data were thus exactly the same as in the original data, and the total number of MD observations from all the patients was also the same. Moreover, the estimated values of the parameters including β_0, β_2 and β_3 obtained from the best fit to the original data were used for the simulation. The set of parameters $(\beta_1, \beta_4, \beta_5)$ specifying the decay rates of the MD for the treated and control groups had values altered in a systematic way for different sets of simulations to reflect different deterioration schemes. Further discussion on the choice of parameter values is given in Section 7.2.3.

7.2.2 Generation of the Random Effects and the Within-Patient Errors

For patient i, we generated a random effects vector $\boldsymbol{b}_i = (b_{0i}, b_{1i}, b_{5i})$ from the multivariate normal distribution with mean $\mathbf{0}$ and a covariance matrix $\hat{\mathbf{D}}$. The matrix $\hat{\mathbf{D}}$ was the estimated covariance structure of the individual patients' random effects from the best fit and it was given by

$$\hat{\mathbf{D}} = \begin{bmatrix} 2.03 & 9.57 \times 10^{-4} & -9.69 \times 10^{-4} \\ 9.57 \times 10^{-4} & 1.43 \times 10^{-6} & -8.98 \times 10^{-7} \\ -9.69 \times 10^{-4} & -8.98 \times 10^{-7} & 6.8 \times 10^{-7} \end{bmatrix}$$

Because the Laird and Ware LME model in Equation (4.1) assumes no correlation between the random effects of any two different patients, and that the same covariance structure $\hat{\mathbf{D}}$ is used for all the patients, the covariance matrix for the vector of random effects for all of the patients,

$$(b_{01}, b_{11}, b_{51}, b_{02}, b_{12}, b_{52}, \cdots, b_{0,97}, b_{1,97}, b_{5,97},)$$

will be block diagonal with the matrix $\hat{\mathbf{D}}$.

For the within-patient error vector for the ith patient, e_i, we assumed the independence covariance structure, i.e., $\mathbf{R}_i = \hat{\sigma}^2 \mathbf{I}$, where $\hat{\sigma}^2$ was estimated from the best fit to the original data. Although the results from fitting the original data showed that the within-patient errors tend to follow the continuous first-order autoregressive model, the correlation between the errors present in two MD observations of the same patient measured one day apart was found to be about 0.2 on average. As the times of successive measurements as observed in the original data were at least a week apart and in the general case, several months apart, the correlation between the errors present in two consecutive MD levels would be close to 0. It was therefore reasonable to assume independence of the within-patient errors. We consequently simulated within-patient errors as independent observations from the normal distribution with mean 0 and a variance of $\hat{\sigma}^2$.

The random component in the ith patient's MD value would then be defined as $\mathbf{Z}_i b_i + e_i$, where $\mathbf{Z}_i$, the design matrix for the random effects to be used for our simulation, was the same design matrix used to fit the original data.

7.2.3 Different Combinations of the Decay Rates of the Mean Defect

In studying the effect of the first phase of rapid decline of the MD within the treated patients on the results of the ITT and the PP analyses of the MD data, one can imagine not only the absolute decay rate but also the rate relative to both the post-IOP-stabilization rate and the rate of the control group will be a determining factor that helps to explain the difference in the results of the two types of analyses. This motivated us to experiment with different combinations of the parameters $(\beta_1, \beta_4, \beta_5)$ in order to study the effect under different schemes of generalized visual field loss which was measured by the MD. These parameter combinations gave rise to the sets of $(\beta_C, \beta_{T1}, \beta_{T2})$ in Table 7.9 where

- $\beta_C = \beta_1$ represents the mean decay rate of the MD for the control group

- $\beta_{T1} = \beta_1 + \beta_4$ represents the mean decay rate of the MD for the treated group before IOP stabilization

- $\beta_{T2} = \beta_1 + \beta_4 + \beta_5$ represents the mean decay rate of MD for the treated group after IOP stabilization

The rationale behind the chosen values of the three parameters is given as follows. The natural history of normal tension glaucoma showed a gradual decay in the MD level, as was observed within the control group in our original data. We therefore assumed a negative mean time slope for the control group, i.e, $\beta_C < 0$ in our simulation. In particular, here we considered two different mean depression rates: 0.001721 and 0.003442. The former was the estimate of β_C from the best fit to the original data, and the latter was double the former. The time by group interaction coefficient, as represented by β_4 in Equation (6.5), must be non-positive so as to give a β_{T1} steeper than β_C. We saw from the original data that the treatment caused an initial decay in addition to the natural depression of the MD prior to IOP stabilization. The simulated data must also reflect such a trend, and therefore for each of the two β_C values, we tried four different β_{T1}'s:

Table 7.9: The different decay rates of the MD for the control and treated groups used in the simulation.

β_0 (Intercept) = -1.1149, β_2 (Group) = -0.5642, β_3 (MD at randomization) = 0.8309		
	β_C = -0.001721	β_C = -0.003442
$\beta_{T1} = \beta_C$	$\beta_{T2} =$ -0.001 -0.0014 -0.001721	$\beta_{T2} =$ -0.002 -0.0025 -0.003 -0.003442
$\beta_{T1} = \beta_C - 0.001256$	$\beta_{T2} =$ -0.001 -0.00125 -0.0015 -0.001721 -0.002 -0.00225	$\beta_{T2} =$ -0.0025 -0.003 -0.003442 -0.004
$\beta_{T1} = \beta_C - 2 \times 0.001256$	$\beta_{T2} =$ -0.0005 -0.001 -0.0015 -0.001721 -0.002 -0.0025	$\beta_{T2} =$ -0.002 -0.0025 -0.003 -0.003442 -0.004
$\beta_{T1} = \beta_C - 4 \times 0.001256$	$\beta_{T2} =$ -0.0007 -0.00105 -0.0014 -0.001721 -0.0021 -0.00245	$\beta_{T2} =$ -0.002 -0.0025 -0.003 -0.003442 -0.004

- $\beta_{T1} = \beta_C$

- $\beta_{T1} = \beta_C - \kappa\Delta$, for $\kappa = 1, 2, 4$

The increment Δ was chosen to be 0.001256 which was about half size of the estimated value of β_4 from the best fit to the original data. So we always had $\beta_C \geq \beta_{T1}$. The first case assumed no harm was done by the treatment during the time the treated patients were waiting for their IOPs to reduce and stabilize, and that their MD declined at the same rate as the control patients did. The other three cases showed a steeper decay in the treated group as compared to the control group before the stabilization of the treatment. Finally, for every (β_C, β_{T1}) combination, various β_{T2} values were used to simulate the MD data. The size of increments in β_{T2} varied depending on the values of β_C and β_{T1}. The β_{T2} values were chosen to reflect the following three scenarios:

1. $\beta_{T2} > \beta_C \geq \beta_{T1}$:

    the treatment results in a slower mean decay rate of the MD than the rate of the control group after the stabilization of the IOP is achieved. Situations where the mean MD trajectories of the control and the treated groups intersect after IOP stabilization will be considered. The intersection can occur within or outside the time range the patients were followed in the study.

2. $\beta_{T2} = \beta_C$:

    the treated patients share the same mean decay rate of the MD as the control patients after IOP stabilization.

3. $\beta_C > \beta_{T2} > \beta_{T1}$:

    the mean post-IOP-stabilization decay rate of the MD is slower than before IOP stabilization, but is still faster than the rate of the control group. This implies a definite unfavourable effect of the treatment.

7.3   Simulation Procedures

We followed the steps below in our simulation experiment:

1. Specify the values of $(\beta_1, \beta_4, \beta_5)$ to be used in simulation.

2. Generate the mean structure for 97 patients as described in Section 7.2.1, using the specified values of $(\beta_1, \beta_4, \beta_5)$.

3. Generate the random effects and the within-patient errors for 97 patients to obtain the random component of the MD values, as described in Section 7.2.2.

4. Sum the mean structure and the random component to obtain the MD values for the 97 patients. The vector **Y** of all the MD values from the 97 patients comprises one sample of the simulated mean defect data.

5. Repeat steps 3 and 4 to obtain a total of 500 samples. Denote the lth sample by $\mathbf{Y}^{(l)}$, l=1,2,...,500.

6. For each of the 500 samples, fit the simulated data $\mathbf{Y}^{(l)}$ using all three approaches: intent-to-treat, per-protocol and piecewise linear modelling approaches. Specifically, the first two approaches fit the model given in Equation (6.3). The ITT analysis models the data simulated for all 97 patients, while the PP analysis models the same set of data with the exclusion of one patient who did not meet the criterion for the PP analysis of the original data. The PCLIN modelling approach, on the other hand, fits the model in Equation (6.5) to the simulated data for all 97 patients. The change points of the segmented MD trajectories to be fitted for the treated patients remain the same as observed in the original data, i.e., the originally observed times of IOP stabilization.

7. For each of the 500 samples, obtain parameter estimates of the fixed effects and their standard errors from fitting the simulated data through each of the three

approaches.

7.4 Results

7.4.1 Comparison of the ITT and the PP approaches

Because the ITT and the PP approaches fit a linear trajectory to both the control and the treated groups, questions of interest might include "Is the decay rate of the MD the same for the two treatment groups?" and "How does the overall MD level compare between the two treatment groups?" To answer the first question, we can test the parallelism of the two linear trajectories corresponding to the two treatment groups by testing for the presence of a time by group interaction effect. Let $\beta_{time \times group}$ be the regression coefficient for the interaction. We would be testing the following hypotheses:

$$H_o : \beta'_C = \beta_T \quad \text{OR} \quad \beta_{time \times group} = 0$$
$$\text{vs.} \quad H_a : \beta'_C \neq \beta_T \quad \text{OR} \quad \beta_{time \times group} \neq 0$$

(7.12)

where β'_C and β_T are the time-slopes of the MD for the control and the treated groups, respectively. The null hypothesis H_o states that there is no interaction effect, and the alternative hypothesis H_a states the presence of an interaction effect.

For each combination of $(\beta_C, \beta_{T1}, \beta_{T2})$ used for simulating the MD data, we computed the p-value for the above hypothesis test for each of the 500 samples and obtained the proportion of rejecting the null hypothesis at a 5% significance level under the ITT and the PP approaches, which is presented in Tables 7.10 and 7.11. We also used the paired t-test to compare the proportions of rejecting H_o from the ITT and the PP analyses for each combination of $(\beta_C, \beta_{T1}, \beta_{T2})$. The proportion of rejection can be treated as the mean of 500 Bernoulli random variables, each of which takes on a value of 1 if H_o is rejected and a value of 0 if H_o is not rejected. By the result of the Central Limit Theorem, the normality of the proportions of rejection can be approximated as the sample size of

Table 7.10: Proportions of rejecting the null hypothesis of no time by group interaction effect in the ITT and the PP analyses, for an MD decay rate of the control group = -0.001721 and different sets of MD decay rates of the treated group.

$\beta_C = -0.001721$			
Parameter values used for simulation		Proportion of rejecting H_o	
β_{T1}	β_{T2}	ITT	PP
-0.001721	-0.001	0.706	0.732
	-0.0014	0.210	0.254
	-0.001721	0.072	0.076
-0.002977	-0.001	0.644	0.737
	-0.00125	0.302	0.426
	-0.0015	0.074	0.112
	-0.001721	0.070	0.064
	-0.002	0.236	0.138
	-0.00225	0.600	0.426
-0.004232	-0.0005	0.974	0.990
	-0.001	0.588	0.768
	-0.0015	0.104	0.162
	-0.001721	0.108	0.078
	-0.002	0.296	0.134
	-0.0025	0.874	0.728
-0.006743	-0.0007	0.756	0.950
	-0.00105	0.314	0.688
	-0.0014	0.064	0.246
	-0.001721	0.154	0.050
	-0.0021	0.640	0.279
	-0.00245	0.914	0.690

500 is reasonably large. The use of the paired t-test is thus validated. The values of the time-slope parameter for the control group estimated in either the ITT or the PP analysis, $\hat{\beta}'_C$, were very close to β_C used for simulating the data. This suggested $E(\hat{\beta}'_C)$ = β_C, where $E(.)$ represents the expected value, and that $\hat{\beta}'_C$ is an unbiased estimator or β_C.

Table 7.11: Proportions of rejecting the null hypothesis of no time by group interaction effect in the ITT and the PP analyses, for an MD decay rate of the control group = -0.003442 and different sets of MD decay rates of the treated group.

$\beta_C = -0.003442$			
Parameter values used for simulation		Proportion of rejecting H_o	
β_{T1}	β_{T2}	ITT	PP
-0.003442	-0.002	0.998	0.998
	-0.0025	0.918	0.924
	-0.003	0.336	0.356
	-0.003442	0.056	0.050
-0.004698	-0.0025	0.862	0.918
	-0.003	0.274	0.360
	-0.003442	0.080	0.058
	-0.004	0.598	0.496
-0.005953	-0.002	1.000	1.000
	-0.0025	0.818	0.914
	-0.003	0.218	0.376
	-0.003442	0.100	0.078
	-0.004	0.622	0.466
-0.008464	-0.002	0.986	0.998
	-0.0025	0.700	0.936
	-0.003	0.110	0.360
	-0.003442	0.152	0.048
	-0.004	0.818	0.490

Certain trends were noted in the rejection rates of the null hypothesis:

- For each fixed pair of (β_C, β_{T1}) used for the simulation, the paired t-test comparing the proportions of rejection showed that the ITT analysis always gave a significantly larger proportion of rejecting H_o than the PP analysis for cases where $\beta_{T2} < \beta_C$ (i.e., β_{T2} being more negative than β_C). Whereas for $\beta_{T2} > \beta_C$ (i.e., β_{T2} being less steep than β_C), the opposite trend occurred, and for most cases, the difference in the rejection proportions was significant.

  When $\beta_{T2} > \beta_C$, the slope of the treated group prior to IOP stabilization, β_{T1}, which was always more negative than β_{T2}, averaged out with β_{T2} to give a β_T that satisfied $\beta_{T1} < \beta_T < \beta_{T2}$ under the ITT approach, and the two slopes β_C' and β_T became quite comparable. However, the PP approach compared the post-IOP-stabilization slope of the treated group with the slope of the control group. The difference between β_C' and β_T under the PP approach will be more distinct and easier to detect, thus resulting in a larger proportion of rejecting the null hypothesis.

  On the other hand, when $\beta_{T2} < \beta_C$, the β_{T1} which was more negative than β_{T2} gave a slope β_T that differed more from β_C' under the ITT approach. In fact, one had $\beta_T < \beta_{T2} < \beta_C$. The PP approach, however, ignored the first portion of data for the treated group before IOP stabilization, so the slope β_T under this approach was comparable to β_{T2} used for simulating the data. Consequently, a more distinct difference between β_T and β_C' was observed in the ITT analysis than was the case in the PP analysis, and a higher proportion of rejecting the null hypothesis was obtained.

- For each fixed β_C and all cases where $\beta_{T2} = \beta_C$, the data simulated can be treated as if they were generated under the null hypothesis of no time by group interaction under the PP approach because this approach compared the slopes of the control and the treated groups after IOP stabilization was achieved. Indeed, our assumption seemed reasonable as for a 5% test, a rejection rate close to 0.05 was obtained

regardless of the values of β_C and β_{T1}.

However, from the ITT analysis, we did not obtain a proportion close to 0.05 when $\beta_{T2} = \beta_C$ and $\beta_{T2} \neq \beta_{T1}$. For each fixed β_C and $\beta_{T2} = \beta_C$, the rejection rate increased with the magnitude of β_{T1}. This agreed with our intuition: the more negative the β_{T1}, i.e, the sharper the decline of the MD during the period awaited for IOP stabilization, the greater the extent to which β_{T1} will pull down the overall slope β_T for the treated group in the ITT analysis. The difference between β_T and β_C' increased and thus led to a stronger evidence against the null hypothesis of no time by group interaction and a greater rejection rate than in the case of the PP analysis.

- For all the special cases where $\beta_C = \beta_{T1} = \beta_{T2}$, the ITT and the PP approaches resulted in similar rejection rates of the null hypothesis. The paired t-test for each of these special cases did not show a significant difference in the two rejection rates. With $\beta_C = \beta_{T1} = \beta_{T2}$, the data simulated would be under the null hypothesis of no time by group interaction in both the ITT and the PP analysis, and this explained for having rejection rates close to 0.05 for our test at a 5% level.

- For each fixed β_C, the discrepancy in the proportions of rejecting the null hypothesis under the ITT and the PP approaches increased in size as β_{T1} increased.

Summary

The simulation experiment helped to explain why different results were obtained from analyzing the mean defect data using the ITT and the PP approaches. It also gave insights into the two approaches to the evaluation of general clinical trials that study treatments entailing a lag time for their stabilization and treating diseases whose progress is monitored by following the change of a certain prognostic characteristic over time. In situations where the treatment speeds up the progression rate of a disease before its

stabilization while later inducing a favourable effect such as slowing down the progression rate, the PP analysis fails to account for the initial phase of rapid progression among the treated patients. Moreover, the approach tends to exaggerate the difference between the progression rates of the treated and the control groups. On the other hand, when the treatment causes a rapid progression not only before but also after the full effect is achieved upon stabilization, the PP analysis tends to diminish the time-group interaction effect and makes it more difficult to detect a difference between the progression rates of the treated and the control groups. Nevertheless, this alone does not necessarily mean that an ITT analysis is preferable to a PP analysis for the evaluation of a clinical trial. It is important to make it clear that an ITT analysis aims to assess the effectiveness of a therapy while a PP analysis attempts to assess the efficacy of the therapy. The PP approach adopted throughout this thesis has a number of pitfalls, which will be further discussed in Chapter 8. However, from the statistical point of view, the most appropriate approach to assessing the treatment effect would be one that accounts for the different patterns of progression before and after the full treatment effect is reached.

7.4.2 Comparison of the PP and the PCLIN approaches

Suppose we are interested in knowing how the rate of progression of a disease compares between the treated group and the untreated group after the stabilization of treatment is achieved. We will be testing the following two sets of hypotheses for the PP and the PCLIN analyses, respectively:

$$H_o : \beta'_C = \beta_T \quad \text{OR} \quad \beta_{time \times group} = 0$$
$$\text{vs. } H_a : \beta'_C \neq \beta_T \quad \text{OR} \quad \beta_{time \times group} \neq 0$$

$$(7.13)$$

where β_C' and β_T are the time-slopes of the MD for the control and for the treated groups, respectively under the PP approach.

$$H_o : \beta_C = \beta_{T2} \quad \text{OR} \quad \beta_4 + \beta_5 = 0$$
$$\text{vs.} \quad H_a : \beta_C \neq \beta_{T2} \quad \text{OR} \quad \beta_4 + \beta_5 \neq 0$$

(7.14)

where β_C and β_{T2} are the time-slopes of the MD for the control group and for the treated group after IOP stabilization, respectively under the PCLIN approach. The parameters $\beta_{time \times group}$ represents the time by group interaction under the PP approach, and $\beta_4 + \beta_5$ represents the true difference between the MD decay rates of the control and the treated groups under the PCLIN approach based on which the data were simulated. The test statistics for the above hypothesis tests will be

$$\text{PP analysis} \quad : \quad T_{PP} = \frac{\hat{\beta}_{time \times group}}{SE(\hat{\beta}_{time \times group})}$$

(7.15)

$$\text{PCLIN analysis} \quad : \quad T_{PCLIN} = \frac{\hat{\beta}_4 + \hat{\beta}_5}{SE(\hat{\beta}_4 + \hat{\beta}_5)}$$

(7.16)

The p-value for testing the hypotheses under the PCLIN approach can be obtained in the same way as presented in Section 6.1 (p.41). Again, for every set of $(\beta_C, \beta_{T1}, \beta_{T2})$ used for the simulation, the proportions of rejecting the null hypotheses in Equations (7.13) and (7.14) at a 5% significance level were computed and are summarized in Tables 7.12 and 7.13. We also used the paired t-test to compare the proportions of rejecting H_o from the PP and the PCLIN analyses for each combination of $(\beta_C, \beta_{T1}, \beta_{T2})$. The proportion of rejection can be treated as the mean of 500 Bernoulli random variables, each of which takes on a value of 1 if H_o is rejected and a value of 0 if H_o is not rejected.

Observe that for all cases where $\beta_{T2} = \beta_C$, the PP and the PCLIN approaches gave very close rejection rates in the respective tests. Especially, the PP and the PCLIN rates were close to 0.05 for a significance level of 5%. Also, the paired t-test for all these cases did not show a significant difference in the two rejection rates. It is therefore reasonable to assume the two hypothesis tests were at the same significance level, and it is legitimate

Table 7.12: Proportions of rejecting the null hypothesis of no difference in the MD decay rates between the two treatment groups after IOP stabilization in the PP and the PCLIN analyses, for an MD decay rate of the control group = -0.001721 and different sets of MD decay rates of the treated group.

$\beta_C = -0.001721$			
Parameter values used for simulation		Proportion of rejecting H_o	
β_{T1}	β_{T2}	PP	PCLIN
-0.001721	-0.001	0.732	0.732
	-0.0014	0.254	0.226
	-0.001721	0.076	0.078
-0.002977	-0.001	0.737	0.734
	-0.00125	0.426	0.418
	-0.0015	0.112	0.110
	-0.001721	0.064	0.070
	-0.002	0.138	0.204
	-0.00225	0.426	0.538
-0.004232	-0.0005	0.990	0.998
	-0.001	0.768	0.766
	-0.0015	0.162	0.180
	-0.001721	0.078	0.072
	-0.002	0.134	0.194
	-0.0025	0.728	0.790
-0.006743	-0.0007	0.950	0.956
	-0.00105	0.688	0.714
	-0.0014	0.246	0.248
	-0.001721	0.050	0.058
	-0.0021	0.279	0.328
	-0.00245	0.690	0.744

Table 7.13: Proportions of rejecting the null hypothesis of no difference in the MD decay rates between the two treatment groups after IOP stabilization in the PP and the PCLIN analyses, for an MD decay rate of the control group = -0.003442 and different sets of MD decay rates of the treated group.

$\beta_C = -0.003442$			
Parameter values used for simulation		Proportion of rejecting H_o	
β_{T1}	β_{T2}	PP	PCLIN
-0.003442	-0.002	0.998	1.000
	-0.0025	0.924	0.922
	-0.003	0.356	0.358
	-0.003442	0.050	0.052
-0.004698	-0.0025	0.918	0.928
	-0.003	0.360	0.364
	-0.003442	0.058	0.064
	-0.004	0.496	0.564
-0.005953	-0.002	1.000	1.000
	-0.0025	0.914	0.928
	-0.003	0.376	0.392
	-0.003442	0.078	0.082
	-0.004	0.466	0.502
-0.008464	-0.002	0.998	0.998
	-0.0025	0.936	0.948
	-0.003	0.360	0.380
	-0.003442	0.048	0.042
	-0.004	0.490	0.538

to compare their statistical power. Intuitively, the comparison of the slopes β'_C and β_T under the PP approach and the comparison of the slopes β_C and β_{T2} under the PCLIN approach should yield very similar results as both are comparing the slopes of the controls and of the treated patients after treatment stabilization is established, but the results from the simulation experiment tell a slightly different story. From the rejection rates presented in Tables 7.12 and 7.13, we noted the following:

- For the difference sets of slopes $(\beta_C, \beta_{T1}, \beta_{T2})$ used for simulating the data, when $\beta_{T2} \geq \beta_C$, the PP and the PCLIN approaches rejected the null hypothesis in their respective tests at about the same rate, regardless of the values of β_C and β_{T1}. In all but two cases where $\beta_{T2} \geq \beta_C$, the paired t-test did not show a significant difference in the two rejection rates. This implies that the two hypothesis tests are relatively equally powerful in detecting a difference between the slopes describing the post-IOP-stabilization decay of the MD for the treated and the control groups at a 5% level.

- For all cases where $\beta_{T2} < \beta_C$, the PCLIN approach gave a higher rejection rate than the PP approach. The paired t-test also showed a significant difference in the two rejection rates for all such cases. The hypothesis test in the PCLIN analysis appears to be more powerful in detecting a difference between the decay rates of the MD after IOP stabilization for the treated and the control groups at a 5% level.

In order to explain the above trends, we examined the parameter estimates obtained from the two approaches. The estimated values of the time-slope coefficient, or equivalently, the slope coefficient for the control group (β'_C from the PP analysis and β_C from the PCLIN analysis), were very close and they will not be presented here. There was also no systematic discrepancy between the coefficient estimates $\hat{\beta}'_C$ and $\hat{\beta}_C$ across the different sets of $(\beta_C, \beta_{T1}, \beta_{T2})$; there did not seem to be a bias in $\hat{\beta}'_C$ as an estimator of β_C. ($\hat{\beta}_C$ is an unbiased estimator of β_C.) We also looked at the following two parameters:

Table 7.14: Sample mean of the 500 estimated regression coefficients and standard errors in the PP and the PCLIN analyses, for an MD decay rate of the control group = -0.001721 and different sets of MD decay rates of the treated group.

$\beta_C = -0.001721$					
		PP		PCLIN	
Parameter values used for simulation		$\hat{\beta}_{time \times group}$ (SE) (values in 10^{-4})		$\hat{\beta}_4 + \hat{\beta}_5$ (SE) (values in 10^{-4})	
β_{T1}	β_{T2}				
-0.001721	-0.001	7.71	2.93	7.38	2.78
	-0.0014	3.56	2.84	3.21	2.73
	-0.001721	0.45	2.82	0.07	2.70
-0.002977	-0.001	7.42	2.83	7.12	2.70
	-0.00125	4.96	2.86	4.68	2.73
	-0.0015	2.39	2.87	2.15	2.72
	-0.001721	0.22	2.77	-0.08	2.64
	-0.002	-2.41	2.77	-2.70	2.64
	-0.00225	-4.99	2.75	-5.31	2.63
-0.004232	-0.0005	12.5	2.90	12.3	2.76
	-0.001	7.50	2.79	7.21	2.67
	-0.0015	2.53	2.88	2.30	2.75
	-0.001721	0.23	2.79	0.001	2.65
	-0.002	-2.61	2.88	-2.83	2.73
	-0.0025	-7.49	2.84	-7.71	2.70
-0.006743	-0.0007	10.3	2.87	10.2	2.72
	-0.00105	6.88	2.85	6.67	2.71
	-0.0014	3.56	2.82	3.21	2.73
	-0.001721	0.13	2.79	-0.02	2.65
	-0.0021	-3.72	2.77	-3.80	2.62
	-0.00245	-7.09	2.85	-7.24	2.70

- $\beta_{time \times group}$ from the PP analysis

- $\beta_4 + \beta_5$ from the PCLIN analysis

For each set of (β_C,β_{T1},β_{T2}) used for simulation, the sample mean of the estimates of the above two parameters and their standard errors across the 500 samples were computed and are tabulated in Tables 7.14 and 7.15.

The histograms of $\hat{\beta}_{time \times group}$ and $\hat{\beta}_4 + \hat{\beta}_5$ for four randomly selected combinations

Table 7.15: Sample mean of the 500 estimated regression coefficients and standard errors in the PP and the PCLIN analyses, for an MD decay rate of the control group = -0.003442 and different sets of MD decay rates of the treated group.

$\beta_C = -0.003442$		PP		PCLIN	
Parameter values used for simulation		$\hat{\beta}_{time \times group}$ (SE) (values in 10^{-4})		$\hat{\beta}_4 + \hat{\beta}_5$ (SE) (values in 10^{-4})	
β_{T1}	β_{T2}				
-0.003442	-0.002	14.7	2.76	14.5	2.64
	-0.0025	9.80	2.89	9.56	2.75
	-0.003	4.54	2.83	4.30	2.70
	-0.003442	0.19	2.78	-0.06	2.66
-0.004698	-0.0025	9.56	2.81	9.37	2.68
	-0.003	4.67	2.76	4.43	2.61
	-0.003442	0.30	2.81	0.06	2.67
	-0.004	-5.43	2.80	-5.69	2.67
-0.005953	-0.002	14.5	2.75	14.3	2.61
	-0.0025	9.53	2.90	9.36	2.74
	-0.003	4.45	2.72	4.30	2.58
	-0.003442	0.25	2.78	0.05	2.66
	-0.004	-5.26	2.83	-5.44	2.70
-0.008464	-0.002	14.5	2.80	14.4	2.65
	-0.0025	9.57	2.78	9.48	2.63
	-0.003	4.48	2.79	4.37	2.64
	-0.003442	0.05	2.85	-0.10	2.71
	-0.004	-5.47	2.87	-5.54	2.72

of $(\beta_C, \beta_{T1}, \beta_{T2})$ were also plotted in Figure 7.13, where in each histogram plotted the 500 parameter estimates obtained from the 500 samples. From the plots, we saw that the histograms of $\hat{\beta}_{time \times group}$ had slightly heavier tails (greater dispersion) than the histograms of $\hat{\beta}_4 + \hat{\beta}_5$. This agreed with what we observed from the standard errors of the estimates $\hat{\beta}_{time \times group}$ and $\hat{\beta}_4 + \hat{\beta}_5$: $SE(\hat{\beta}_{time \times group}) > SE(\hat{\beta}_4 + \hat{\beta}_5)$ for all cases. The mean $SE(\hat{\beta}_{time \times group})$ and $SE(\hat{\beta}_4 + \hat{\beta}_5)$ were compared using the t-test, and the difference was shown to be significant for every combination of $(\beta_C, \beta_{T1}, \beta_{T2})$ used in simulation.

From Tables 7.14 and 7.15, we also noticed that the $\hat{\beta}_{time \times group}$'s always had a larger sample mean than the $\hat{\beta}_4 + \hat{\beta}_5$'s regardless of the values of β_C, β_{T1} and β_{T2}, although such a trend was not apparent in the histograms. Moreover, all the histograms resembled the normal distribution, indicating that the normality assumption of $\hat{\beta}_{time \times group}$ and $\hat{\beta}_4 + \hat{\beta}_5$ used in constructing the test-statistics T_{PP} and T_{PCLIN} for the hypothesis tests was reasonably approximated.

Since $\hat{\beta}_4 + \hat{\beta}_5$ is an unbiased estimator of $\beta_4 + \beta_5$, i.e., $E(\hat{\beta}_4 + \hat{\beta}_5) = \beta_4 + \beta_5$, a property of the LME model which gives unbiased estimates of the fixed effects, the systematically larger sample means of the $\hat{\beta}_{time \times group}$'s than those of the $\hat{\beta}_4 + \hat{\beta}_5$'s suggested that $E(\hat{\beta}_{time \times group}) > E(\hat{\beta}_4 + \hat{\beta}_5) = \beta_4 + \beta_5$. To ascertain such a difference in the expected values, we used the paired t-test to compare the mean values of $\hat{\beta}_{time \times group}$ and $\hat{\beta}_4 + \hat{\beta}_5$ for each combination of $(\beta_C, \beta_{T1}\beta_{T2})$, and a significant difference in the two mean values was demonstrated for all combinations of $(\beta_C, \beta_{T1}\beta_{T2})$. In other words, based on our assumption that the MD data followed the piecewise LME model, $\hat{\beta}_{time \times group}$ under the PP approach seemed to be a biased estimator of the true difference between the post-IOP-stabilization decay rates of the MD of the control and the treated groups. Such a bias may be explained by the difference in the baseline adopted by the PP and the PCLIN approaches. The times of repeated MD measurements that entered the LME model under the two approaches were measured from two different baseline times. However, it was not clear what actually led to the discrepancy in the mean values of the $\hat{\beta}_{time \times group}$'s

Figure 7.13: Histograms of estimated coefficients $\hat{\beta}_{time \times group}$ (from the PP analysis) and $\hat{\beta}_4 + \hat{\beta}_5$ (from the PCLIN analysis) for randomly selected sets of $(\beta_C, \beta_{T1}, \beta_{T2})$ used in the simulation. (a) $\beta_C = -0.001721$, $\beta_{T1} = -0.004232$, $\beta_{T2} = -0.0015$ (b) $\beta_C = -0.001721$, $\beta_{T1} = -0.002977$, $\beta_{T2} = -0.005953$ (c) $\beta_C = -0.003442$, $\beta_{T1} = -0.005953$, $\beta_{T2} = -0.0025$ (d) $\beta_C = -0.003442$, $\beta_{T1} = -0.004698$, $\beta_{T2} = -0.003442$
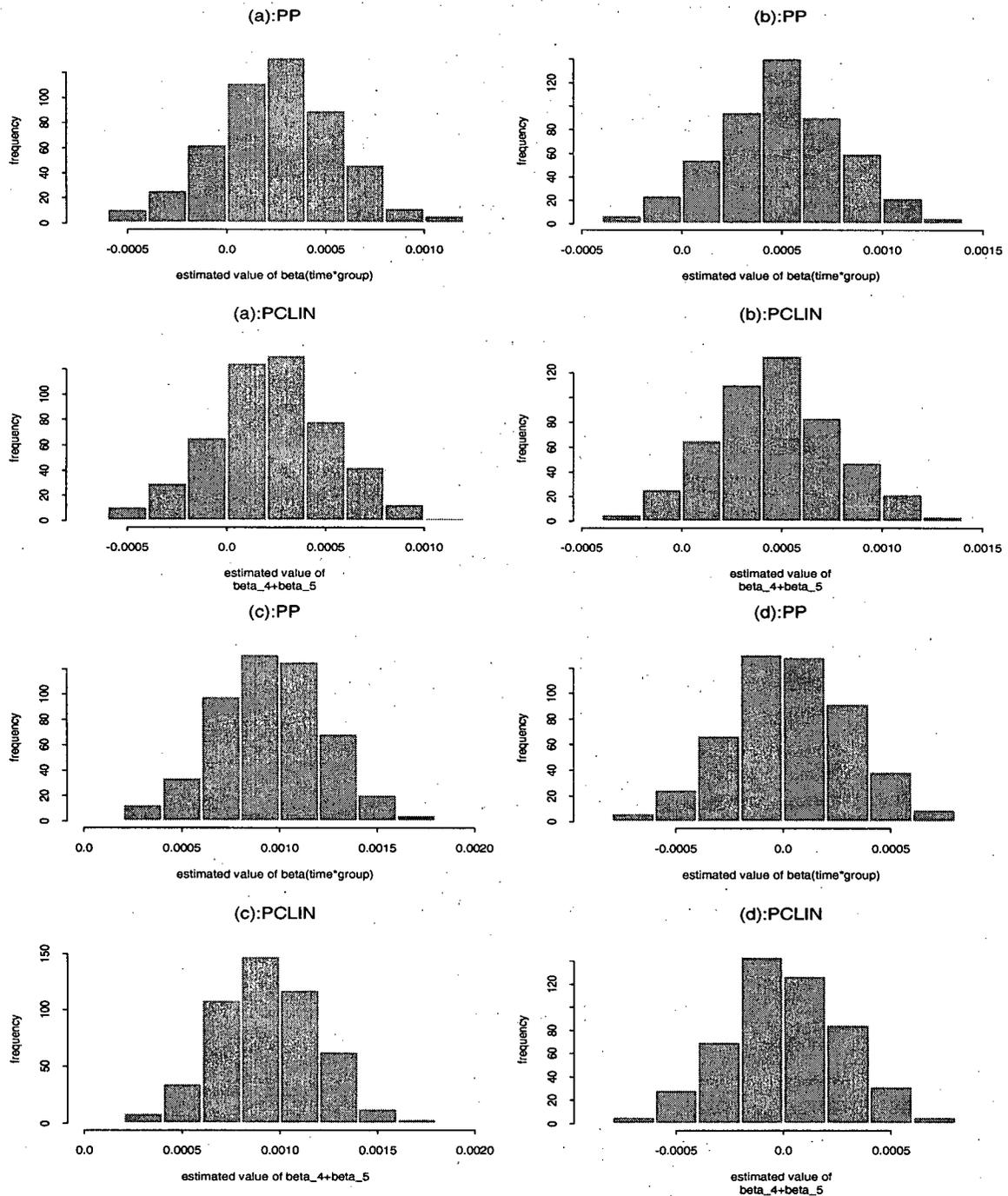
and the $\hat{\beta}_4 + \hat{\beta}_5$'s. The greater standard error of the $\hat{\beta}_{time \times group}$'s as compared to that of the $\hat{\beta}_4 + \hat{\beta}_5$'s is possibly the result of a smaller number of MD observations used for parameter estimation in the PP analysis.

We have seen that in the estimation of the true difference between the slopes of the control and the treated group after IOP stabilization, the estimator $\hat{\beta}_{time \times group}$ under the PP approach seemed to be biased and less efficient (i.e., having a larger standard error) than the estimator $\hat{\beta}_4 + \hat{\beta}_5$ under the PCLIN approach. With this knowledge, the reasons behind the trends observed in the rejection rates of the null hypotheses in Equations (7.13) and (7.14) become apparent and are given as follows. Suppose $\beta_{time \times group} = C_1$ and $\beta_4 + \beta_5 = C_2$ for some real constants C_1 and C_2. As $\hat{\beta}_{time \times group}$ and $\hat{\beta}_4 + \hat{\beta}_5$ are unbiased estimators of $\beta_{time \times group}$ and $\beta_4 + \beta_5$, we have $E(\hat{\beta}_{time \times group}) = C_1$ and $E(\hat{\beta}_4 + \hat{\beta}_5)$ $= C_2$. Also notice that the standard errors for the $\hat{\beta}_{time \times group}$'s and the $\hat{\beta}_4 + \hat{\beta}_5$'s stayed relatively constant irrespective of the estimated values of the parameters. Hence, it was reasonable to assume the standard errors were independent of the values of the parameter estimates. When $\beta_4 + \beta_5 > 0$, or $\beta_{T2} > \beta_C$, i.e., in the presence of a favourable treatment effect in slowing down the MD decay rate after achieving IOP stabilization, earlier results from comparing the mean values of $\hat{\beta}_{time \times group}$'s and $\hat{\beta}_4 + \hat{\beta}_5$'s suggested that $C_1 = E(\hat{\beta}_{time \times group}) > E(\hat{\beta}_4 + \hat{\beta}_5) = C_2$. And since $C_2 = \beta_4 + \beta_5 > 0$, we have $C_1 > C_2 > 0$. However, $SE(\hat{\beta}_{time \times group}) > SE(\hat{\beta}_4 + \hat{\beta}_5)$. Upon comparison of the relative size of the parameter estimates and their SE's, we found that

$$\frac{\hat{\beta}_{time \times group}}{\hat{\beta}_4 + \hat{\beta}_5} \simeq \frac{SE(\hat{\beta}_{time \times group})}{SE(\hat{\beta}_4 + \hat{\beta}_5)}$$

or

$$T_{PP} = \frac{\hat{\beta}_{time \times group}}{SE(\hat{\beta}_{time \times group})} \simeq \frac{\hat{\beta}_4 + \hat{\beta}_5}{SE(\hat{\beta}_4 + \hat{\beta}_5)} = T_{PCLIN}$$

(See Equations (7.15) and (7.16)). We thus expect to have

$$E(T_{PP}) = \frac{C_1}{SE(\hat{\beta}_{time \times group})} \simeq \frac{C_2}{SE(\hat{\beta}_4 + \hat{\beta}_5)} = E(T_{PCLIN})$$

It follows that on average, the p-values for the two hypothesis tests in the PP and the PCLIN analyses will be very similar and about the same rejection rates will be obtained, as observed in our simulation results.

When $\beta_4 + \beta_5 < 0$, or $\beta_{T2} < \beta_C$, i.e., in the presence of an unfavourable treatment effect (the treatment speeds up the MD decay rate after achieving IOP stabilization), we again expect to have $C_2 = E(\hat{\beta}_4 + \hat{\beta}_5) < E(\hat{\beta}_{time \times group}) = C_1$. Because $C_2 < 0$, $C_2 < C_1$ implies C_1 is closer to or above 0. And with $SE(\hat{\beta}_{time \times group}) > SE(\hat{\beta}_4 + \hat{\beta}_5)$, the following is likely to hold:

$$E(|T_{PP}|) = \left| \frac{C_1}{SE(\hat{\beta}_{time \times group})} \right| < \left| \frac{C_2}{SE(\hat{\beta}_4 + \hat{\beta}_5)} \right| = E(|T_{PCLIN}|)$$

It follows that on average, the p-value obtained from the PP analysis will be larger than that from the PCLIN analysis for the two-sided hypothesis tests. This explains the smaller rejection rates for the PP analysis.

Summary

We have seen from the simulation results that when the data were generated from the LME model under the PCLIN modelling approach, the PP analysis slightly overestimated and produced less efficient estimates of the difference between the slopes of the treated and the control groups after IOP stabilization. In particular, for cases where the post-IOP-stabilization MD within the treated patients decayed at a slower rate than the untreated patients ($\beta_{T2} > \beta_C$), although the PP and the PCLIN approach gave very similar rejection rates of the null hypotheses stating no difference between the post-IOP-stabilization slopes of the two groups, the hypothesis tests corresponding to the two

approaches seeming equally powerful was simply a coincidence resulted from the biased and the less efficient estimates of the difference. When the mean MD decay rate for the treated group after IOP stabilization was faster than that of the control group ($\beta_{T2} < \beta_C$), which was not the situation present in our original data, the simulation results suggested that the hypothesis test under the PCLIN approach was statistically more powerful than the counterpart under the PP approach. In summary, the results from the comparison between the PP and the PCLIN approaches implied that the negligence of the lag time in the PP analysis may give rise to biased parameter estimates and a decreased statistical power of hypothesis tests for treatment comparisons.

Chapter 8

Discussion

In this thesis, we investigated two different approaches to evaluate a clinical trial: the intent-to-treat (ITT) approach and the per-protocol (PP) approach. The former calls for inclusion of all the patients who are randomized in the final analysis of the trial irrespective of problems of missing data, cross-overs, drop-outs and non-compliance, while the latter analyzes a subset of patients who meet the treatment efficacy criteria. We applied the two approaches to the data from the Collaborative Normal Tension Glaucoma Study in which a delay in treatment stabilization was observed in the treated group. Besides the difference in the inclusion of patients, the two approaches we adopted differed in the definition of the baseline time. The ITT approach assumed all the patients were followed starting at the time of randomization, whereas the baseline for the treated group was shifted to the individuals' times of IOP stabilization under the PP approach. Upon IOP stabilization, the effect of having a 30% IOP reduction was regarded as fully achieved. The two baseline definitions were chosen in accordance with the principles of assessing the treatment effectiveness and the treatment efficacy in the ITT and the PP analyses, respectively.

In modelling the mean defect (MD) data, a linear mixed effects (LME) model was fitted in both the ITT and the PP analyses. In particular, the MD was assumed to decay linearly with time in each of the treated and the control groups. A random intercept and time-slope were included in the model to allow for different initial MD levels at baseline and different MD decay rates across patients. At a 5% level, gender and baseline

96

information including age and IOP did not explain the MD trend over time in either the ITT or the PP analyses. A significantly slower average MD decay rate for the treated group than the control group as a result of lowering the IOP was demonstrated in the PP analysis but not in the ITT analysis. Furthermore, the ITT analysis showed that the treated group had a more negative overall MD level than the control group, while the PP analysis did not pick out such a difference. Based on these results, quite different conclusions were reached on the usefulness of an IOP reduction strategy in treating normal tension glaucoma. The PP analysis suggested a beneficial treatment effect, but the ITT analysis concluded an ineffective or even a harmful treatment because the IOP lowering therapy not only failed to slow down the natural decay of the MD but also depressed the overall MD level.

An additional approach to analyzing the MD data, the piecewise linear (PCLIN) modelling approach, revealed that during the period awaiting the IOP to reduce and stabilize, the treated patients had on average a faster MD decay rate than both the rate after IOP stabilization and the rate of the control group. In light of this finding, the difference in the results of the ITT and the PP analyses can possibly be explained by the fact that the MD data prior to IOP stabilization of the treated patients were excluded from the PP analysis. The first phase of sharp decline of the MD pulled down the mean MD level for the treated group, which also happened to be more negative at randomization than the mean level for the control group in the ITT analysis. Besides, the slower decay rate of the MD within the treated group that followed after IOP stabilization averaged out with the initial rapid decay and resulted in an overall rate that was comparable with that of the controls in the ITT analysis. Therefore, there was weaker evidence to declare a significant time by group interaction than in the case of a PP analysis. The one patient who was included in the ITT analysis but not in the the PP analysis only had five MD observations measured from the time of randomization. The excluded patient, who belonged to the treated group, had a similar early MD trend as observed in the other

treated patients: there appeared to be a larger rate of decrease in MD before IOP stabilization than the rate afterwards. It was unlikely that the exclusion of this particular patient had led to the difference in the results from the two approaches.

Because the time to IOP stabilization was relatively short compared to the treated patients' follow-up periods, and the MD trend over time was highly variable across patients, any difference between the MD trend before and after IOP stabilization within each of the treated individuals was not easily noticeable. A linear model appeared to provide an adequate fit to the data of the treated group. However, upon fitting a segmented linear trajectory for the treated group in the PCLIN modelling approach, the modelling of the MD data was significantly improved marginally. The individual fitted trajectories of the treated patients also seemed to provide a reasonably good description of the MD observations, especially for patients whose MD patterns differed greatly before and after IOP stabilization. Although the PCLIN approach was data-driven, it was shown to outperform the ITT and the PP approaches we have used in modelling the MD data in our study.

The difference in the conclusions drawn from the ITT and the PP analyses motivated our simulation experiment which aimed to study how a different decay pattern of the MD before and after IOP stabilization within the treated group affected the results of the two approaches. We generated MD data using the best fitted LME model found from the PCLIN modelling approach. The analysis of the simulated data showed that the PP analysis tends to magnify the difference between the decay rates of the MD between the control and the treated groups in the case where the treated slope became less steep than the control slope after IOP stabilization, but the difference was diminished when the treated slope after IOP stabilization, while not as steep as it had been before the stabilization, remained steeper than the control slope. The poor performance of the PP analysis can be attributed to its exclusion of the data from the treated patients prior to IOP stabilization. And in any case where the treatment effect differed before and after

the 30% IOP reduction was attained, the PP analysis failed to account for the trend in the first phase. Besides, we simulated MD data to compare the performance of the PP and the PCLIN approaches in terms of parameter estimation and statistical power of detecting a difference between the post-IOP-stabilization decay rates of the MD of the treated and the untreated groups. Summaries of the simulation results were already given in Chapter 7.

In applying the ITT and the PP approaches to model the MD data in our study, we also looked into the floor effect issue by including in the linear mixed effects model an indicator variable classifying patients as having or not having advanced normal tension glaucoma at baseline (An MD level below $-12dB$ is generally regarded as having an advanced stage of the disease). The interaction between the indicator variable and the time-slope covariate was tested and was found to be insignificant at a 5% level in both the ITT and the PP analyses. The decay pattern of the MD appeared to be independent of the baseline MD level, as analyzed using either approach. We also compared the mean baseline MD of the treated and the control groups under the ITT and the PP approaches. When the baseline was taken to be the time of randomization for all the patients in the ITT analysis, the two-sample t-test did not suggest a significant difference between the mean MD levels at randomization of the two groups (the observed means were $-7.39dB$ for the treated group and $-6.72dB$ for the control group, p=0.49). However, in the PP analysis, the two-sample t-test showed that the two groups had significantly different baseline MD levels (the observed means for the treated and the control groups were $-8.93dB$ and $-6.72dB$, respectively, p=0.014). Recall that the PP analysis demonstrated a significantly slower linear decay rate of the MD for the treated group than the control group, and it also showed no evidence of a dependence between the MD decay rate and the initial MD level. However, we cannot rule out the possibility that the significant treatment effect was an artifact of the floor effect because as time elapsed since randomization, more and more patients reached the progression end point and the two groups of untreated and treated patients who remained in the study may have become

more and more unbalanced in terms of sample size, mean MD level and variation of the MD measurements. The floor effect may eventually be influential and result in a fallacy of a beneficial treatment effect. Moreover, the PCLIN analysis showed that the treated patients on average experienced a sharp decline in MD that was followed by a decay at an even slower rate than the control group. It remained uncertain if the slower decay rate was the result of a beneficial effect of having a 30% reduction of the IOP or, perhaps in part, of the floor effect after the large drop of the MD caused by the IOP lowering therapy. We did not analyze the floor effect in the PCLIN analysis because of the complexity of the LME model which involved many time related terms (x_{1ij}, x_{4ij} and x_{5ij} in Equation (6.5)). We could take the same approach as in the ITT and the PP analyses, i.e., we tested for an interaction effect between each of the above time related terms and the classification variable based on the initial MD level. The significance of these interaction effects may help us study whether the time slopes (β_1, $\beta_1 + \beta_4$ and $\beta_1 + \beta_4 + \beta_5$) are affected by the baseline MD level, but such an approach does not seem to provide an adequate and effective assessment of the floor effect. Besides, the interpretation of these interaction effects can be difficult because x_{4ij} and x_{5ij} are already interaction terms.

Apart from the floor effect, a statistical phenomenon called "regression to the mean", also known as the regression effect, might account for the slower MD depression rate observed among patients whose baseline MD levels were highly negative as compared to those who started with less negative MD levels even in the absence of a real treatment effect. The regression effect is usually present between any two variables which are correlated. As an example, suppose we randomly select two patients from the population of normal tension glaucoma patients and whose MD levels at baseline are both less than or greater than the baseline mean. The patient with a baseline MD level farther from the baseline mean will on average have an MD level, relative to the other patient, less far from the mean MD level at a later time, where the initial and the subsequent MD levels are likely to be correlated. This illustrates a property of the regression effect: the more extreme the starting value of the MD, the greater the regression to the mean. Patients

with more extreme negative MD levels initially will on average decrease to a lesser extent than patients with initial MD levels closer to the mean. Therefore, at any fixed time since randomization, patients with highly negative MD at baseline will be expected to show a slower MD depression than those with less negative baseline MD levels, irrespective of the existence of a treatment effect. Moreover, the degree of the regression effect depends on the strength of correlation between the two variables. A weaker correlation will lead to a stronger regression effect. When the time between two MD measurements becomes longer, the measurements will be less correlated, resulting in a greater regression to the mean. The decay rate of the MD which measures the decrease in MD relative to time will be more influenced by the regression effect as a longer time elapses from randomization. So far our discussion of the "regression to the mean" has been focused on the MD levels at baseline and at a later time. This statistical phenomenon can also be observed between the two populations of time slopes before and after IOP stabilization within the treated group. Intuitively, the two time slopes are correlated because they measure the rate of change in MD on the *same* patient over the pre-stabilization and post-stabilization periods. A regression effect between the two time slopes within the treated groups is likely to be present, and this helps to explain why a steep initial MD decay rate might have been followed by a flatter post-stabilization rate than that preceded by a less steep initial time slope.

In the analysis of the time to IOP stabilization data within the treated group, we applied classical methods including the non-parametric Kaplan-Meier method, the semi-parametric Cox regression and parametric modelling. All three methods did not show a significant effect of the age, IOP and MD level at randomization at a 5% level. There was also no strong evidence of a gender effect, although male patients appeared to take a shorter time to achieve IOP reduction and stabilization on average. Nevertheless, it is important to use caution when interpreting the results based on any of the classical methods. The three types of therapy (medical, laser and surgical) were not randomly assigned to the treated patients; the investigators decided the type of treatment the patients were

to receive. The study design was therefore observational and not randomized, and the comparison between the surgical and non-surgical groups might be subject to selection bias. Patients who had a filtering surgery likely differed from patients who received only non-surgical treatments with respect to some uncontrolled factors which affect the ease to achieve IOP reduction and stabilization. We looked to see if the MD and the IOP levels at baseline differed significantly between the two groups of treated patients as this baseline information might account for the investigators' decision, and for the failure of the non-surgical treatments in reducing the IOP to the desired level. The surgical group had an average baseline MD level of -8.84dB (SD=4.88dB) and IOP level of 17.05$mmHg$ (SD=1.55$mmHg$). The corresponding figures for the non-surgical group were -6.42dB (SD=5.49dB) and 16.83$mmHg$ (SD=2.53$mmHg$). The non-surgical group appeared to have a less negative MD and a lower IOP at the time of randomization, but the evidence was not strong enough to demonstrate a significant difference in the mean level of either variable using the two-sample t-test at a 5% level (p=0.14 for MD and p=0.74 for IOP). However, we did not have information on other demographics and ophthalmic conditions which were the basis of the choice of treatment, and thus we cannot identify any source of selection bias. Moreover, some patients who were originally assigned the medical or laser therapy switched to the surgical treatment due to failure to reduce the IOP level using the non-surgical treatment. Among the 43 treated patients, only two were initially given the surgical therapy. By counting patients who had the medical or laser treatment in the first place and switched to surgical treatment later in the surgical group, the extra waiting time before the change of treatment was included as part of the time to IOP stabilization, thus lengthening the observed time to IOP stabilization within the surgically treated group. Our results which showed a longer waiting period for IOP stabilization for the surgical group might be attributable solely to the additional time prior to the change of treatment rather than a true difference between the effects of the non-surgical (medical or laser) treatments and the filtering surgery. But because the time at which the patients switched to the surgical treatment was not available, we were unable to make appropriate adjustments to the observed time to IOP stabilization

when comparing the IOP-stabilization experiences from receiving the different types of therapy.

In analyzing the time to progression data, both the ITT and the PP analyses were carried out. They both showed no evidence of a significant effect of the age at baseline, IOP and MD at baseline, using any of the three standard methods. When analyzing the data from all the patients using the Kaplan-Meier and the parametric modelling methods, the treated group was found to have a significantly prolonged time to progression relative to the untreated group under both the ITT and the PP approaches. In the Cox regression analysis under either the ITT or the PP approach, the treatment of having a 30% reduction in the IOP seemed to have no effect on the male patients, while female patients benefited from the treatment which significantly reduced the risk of progression. Despite the difference in the treatment effect observed between genders, the treatment by gender interaction was found to be insignificant in the Cox regression. It is worth noting that among the 96 patients included in both the ITT and the PP analyses, 34 were male and 62 were female. The number of patients who reached the progression end point was 10 and 23 within the male and the female groups, respectively. With the small number of male patients and a smaller proportion of uncensored times to progression (29.4% for males versus 37.1% for females), the test for the presence of a treatment effect within the male group will be less powerful than the same test within the female group. The treatment might indeed reduce the progression risk for the male patients, but there was not enough power to detect a treatment effect. In the case where the treatment main effect is deemed insignificant, an interaction effect between treatment group and gender will be even harder to detect.

When we applied the non-parametric method to analyze the time to IOP stabilization data and the time to progression data, the log rank test was used to compare the survival experiences of two or more populations. An alternative to the log rank test is the Wilcoxon test [31]. Readers are referred to Collett [31] for details on the statistical

properties of and the difference between the two tests. In general application, the choice of test depends on the alternative hypothesis that is to be tested. Collett discussed that when the alternative to the null hypothesis of no difference between the survival time distributions of two groups is that the hazard for an individual in one group is proportional to the hazard of a similar individual in the other group at any given time, the log rank test is more appropriate. For other types of departure from the equality of survival time distributions between two groups, the Wilcoxon test will be more suitable. Our analyses showed that the proportional hazards model fitted reasonably well to the time-to-event data. As the proportional hazards assumption held in our case, to test for a difference between the survival experiences of two different groups in our data, which can be described by an alternative hypothesis of a non-unity proportional hazards between the two groups, a log rank test seemed appropriate.

Comparing the results from the ITT and the PP analyses, we noticed that for all of the log rank test and the hypothesis tests used in the Cox regression analysis and the parametric method, the PP approach demonstrated a smaller reduced risk of progression for the treated group than the ITT approach did. The baseline set at the time of IOP stabilization for the treated group under the PP approach shortened the time to progression as compared to its ITT counterpart. Therefore, when the 30% reduction of IOP successfully slowed down the hazard to progression as suggested by the significant treatment effect, the shorter time to progression for the treated group in the PP analysis led to a smaller difference in the hazards or time to progression between the two treatment groups, thus explaining the less prominent reduction in the risk of progression for the treated group. Indeed, the group coefficient in the Cox model as estimated by the ITT and the PP approach had very close standard errors. Therefore, with a group coefficient estimate farther from the 0 value in the ITT analysis, a smaller p-value and hence a more significant result were obtained. The same was observed in the results from fitting the Weibull hazard models.

We have seen in our study that both the ITT and the PP analyses showed a beneficial treatment effect in reducing the risk of progression. However, in general applications of the two approaches to clinical trials involving a delay in treatment stabilization, two other possible scenarios may arise:

- When the treatment effect does exist but is not very strong, the ITT approach might show a significant result for the treatment effect while the PP approach does not. The time to progression within the treated group, which is measured starting from the time at which the treatment stabilization is established, is shorter than its ITT counterpart, and hence shows a weaker evidence against the null hypothesis stating no treatment effect. The argument of a PP approach being able to detect a treatment effect better than an ITT approach is not justifiable. Moreover, delaying the baseline time to the time of achieving treatment stabilization for the treated patients often results in a smaller sample size due to potential drop-outs during the pre-stabilization period. As a result, statistical tests performed in the PP analysis are prone to have a lower power.

- When the treatment effect does not exist, i.e., the 30% lowering of the eye pressure does not alter the risk of progression before or after the treatment stabilization, the ITT analysis should not reject the null hypothesis of no treatment effect, subject to a type I error of size equal to the significance level of the hypothesis test. However, the time to progression for the treated patients analyzed in the PP analysis is its ITT counterpart shortened by an amount equivalent to the waiting time for treatment stabilization. If such a waiting time is sufficiently long, the data might show a significant difference in the time-to-progression experience between the treated and the control groups, and subsequently result in a type II error. Even worse, one might conclude an unfavourable treatment effect based on the results from the PP analysis when indeed the treatment is neither beneficial nor harmful.

While the ITT and the PP approaches can lead to different results and conclusions on

the usefulness of a treatment due to the difference in their baseline definitions, analyzing data with a delay in treatment stabilization by means of a multistate model avoids the problem of defining a reasonable baseline for the control group to correspond to the lag time observed in the treatment group. A multistate model describes treatment stabilization as a separate event which occurs between the time of randomization and the time of reaching an end point of interest. For our study, we applied the disability model and performed separate analyses of the times of transition to the two states: the state of IOP stabilization and the state of disease progression. An advantage of modelling hazards for the times to different states separately is the flexibility of incorporating different covariates that are relevant to the various transitions between states. In addition to modelling the individual transition hazards, we fitted Cox and stratified Cox models with a time-dependent indicator variable for treatment stabilization. If the data from treated patients who reached the progression end point before IOP stabilization were available, fitting the Cox model (Equation (5.2)) with the group covariate and the time-dependent IOP-stabilization indicator variable would allow us to assess both the effect of a stable IOP reduction on the progression hazard within the treated patients, and the difference between the time-to-progression experiences of the treated and the control groups. However, such data were unavailable for our analysis, and to avoid computational problems (p.65), we did not include the group covariate in the Cox model and thus could not assess whether the treated and untreated groups had different risks of progression. When we analyzed the data from patients of both genders, we obtained a highly significant favourable result for the effect of a stable IOP reduction on the progression hazard within the treated group ($p=7.6\times10^{-4}$). The result showed about a four-fold decrease in risk of progression after a stable IOP reduction was achieved. The gender-specific analyses, on the other hand, gave rather different results for the opposite genders. The effect of having a stable reduced IOP was found to be insignificant among the male treated patients, while within the female treated patients, a stable reduced IOP had a favourable effect of reducing the risk of progression to about one-sixth of the same risk in the case of not having a stabilized IOP reduction.

As discussed earlier, the presence of a minimum MD level that can be attained in the course of normal tension glaucoma poses a potential floor effect which complicates the comparison of the MD decay rates between the control and the treated groups and may lead one to a false impression of the existence of a treatment effect. Such a floor effect might also affect the comparison of the time-to-progression experiences between treatment groups. As the progression end point in our study was defined as having a decrease of at least $10dB$ in MD relative to the level at the time of randomization, patients who began with an MD level close to the lower bound had limited room to achieve the required decrease and hence took a longer time to reach the progression end point. These patients tend to show a smaller risk of progression not as a result of a treatment effect but rather a floor effect. Indeed, when the treated and the control groups have significantly different mean MD levels at baseline or at a later time point during the course of the disease, the floor effect may come into play and affect treatment comparisons. In the presence of both a floor effect and a beneficial treatment effect which prolongs the survival time to progression, if the treated group happens to have a more negative mean MD level than the untreated group on average, the treatment effect will be biased in favour of the treatment (an overestimation of a beneficial treatment effect). On the contrary, if the control group had a mean MD level closer to the lower limit, the beneficial treatment effect will be masked by the floor effect. The pseudo-reduction in the risk of progression caused by the floor effect within the control group may be about the same extent as the true reduction caused by the treatment itself within the treated group. Although we did not explicitly address the floor effect issue in the analysis of our time-to-event data, it is important to be aware of this common phenomenon and its potential impact on the results of treatment group comparisons.

The PP approach adopted throughout this thesis compared a control group from the time of randomization to a treated group from the time at which the stabilization of the treatment was established. It attempted to assess the efficacy of a 30% IOP reduction in slowing down the progression of normal tension glaucoma, but the approach was

unsatisfactory because of its two major pitfalls: firstly, the baseline was defined at the time of randomization for the control group and at the time of individuals' times of IOP stabilization for the treated group. The two-sample t-test showed that the mean baseline MD level was significantly lower within the treated group (p=0.014). Indeed, during the period awaited for IOP reduction and stabilization, the MD of the treated patients dropped considerably. The paired t-test comparing the mean MD levels at randomization and at the time of IOP stabilization within the treated group gave a significant result (p=2.0×10^{-4}). By comparing the two groups from two different baselines without taking into consideration the treatment lag time, the desirable properties of randomization of treatment assignment will be lost. The difference between the mean baseline MD levels of the two groups in the PP analysis illustrates the lack of homogeneity of different treatment groups. Unlike the experimental approach of an ITT analysis, a PP analysis is essentially observational because it analyzes only a subset of patients who are originally randomized and meet the efficacy criteria. Among patients who qualify for inclusion for the PP analysis, the control group will most likely differ from the treated group with respect to other baseline and demographic characteristics besides the treatment assignment. Problems of confounding and bias in results may then arise: any difference between the two groups may then be attributed to some uncontrolled factors other than the treatment effect. An earlier discussion of the floor effect already pointed out that with different baseline mean MD levels of the two groups, the assessment of the treatment effect can be confounded by the floor effect. Here we see that randomization alone does not guard against bias in treatment group comparison if an inappropriate approach such as the PP approach is used.

Secondly, the PP approach seemed to give biased estimates of the time by group interaction effect in the analysis of the longitudinal MD data, as suggested by the simulation results. The simulation study also found that the PP approach was less powerful than the PCLIN approach in detecting a difference between the decay rates of the treated and the control groups under some circumstances. Although in our study the number of patients

who were originally randomized but were excluded from the PP analysis was small, one can imagine in cases where a large number of patients does not meet the efficacy criteria, especially when the treatment lag time is long compared to the patients' follow-up period, the sample size in the PP analysis will be reduced greatly. The number of patients or observations included in the PP analysis is usually smaller than the planned sample size at the design stage, but how much of a loss in sample size due to the exclusion of patients from analysis is often difficult to predict before a clinical trial is conducted. As a result, the desired power of statistical tests can never be achieved. Recruiting more patients to compensate for the loss in sample size is not only uneconomical but also infeasible in most of the cases.

In seeing the potential problems of bias and diminishing power of the PP approach in the analysis of the glaucoma data and in the simulation experiment, in order to have a valid efficacy assessment in the presence of a lag time in treatment stabilization, we must make appropriate adjustments to the control group to correspond to the delayed treated group. We exploited our findings about the distribution that the time to IOP stabilization data followed in our baseline-adjustment approach. The approach adjusted the baseline for the control group to correspond to the lag time in treatment stabilization in the treated group. We simply shifted the baseline for each control patient from the time of randomization by a time randomly generated from the Weibull distribution originally fitted to the times to IOP stabilization for the treated patients in our data. By adjusting the baseline in this way, we were assuming that in the absence of a treatment effect, the time taken to reach IOP stabilization followed the Weibull distribution. In comparison to the ITT and the PP approaches, our baseline-adjustment approach demonstrated a stronger treatment effect, i.e., the treated group showed a larger reduction in the risk of progression relative to the control group than in the ITT and the PP analyses. Such a trend was not readily observed in the plot of Kaplan-Meier estimates of survivor functions (Figure 6.12(b)). The gender-specific Cox regression analyses gave an estimate of the relative hazard of a treated versus an untreated patient of 0.3 for the male gender

and an estimate of 0.16 for the female gender. Correspondingly, the ITT and the PP estimates were respectively 0.33 and 0.38 for the male gender, and 0.16 and 0.24 for the female gender. In the parametric analysis, a Weibull fit adjusted for the group membership covariate gave a coefficient estimate of -1.13 whereas the ITT and the PP analyses estimated the group coefficient to be -0.878 and -0.754, respectively. A stronger treatment effect shown by the baseline-adjustment approach as compared to the PP approach was indeed expected because the seemingly shorter time to progression among the untreated patients demonstrated in the PP analysis was truncated as a result of the baseline shift, and the difference between the time-to-progression experience between the treated and the untreated groups became even more evident.

While the adjustment made to the baseline of the control group helped to reduce the imbalance of the two treatment groups at their respective baselines induced by the lag time in treatment stabilization, any approaches involving baseline matching between treatment groups do not guarantee comparable groups at the adjusted baselines. Furthermore, these approaches often fail to make use of all the available observations originally recorded. For example, repeated measurements taken at times before the adjusted baseline may have to be discarded because they do not have meaningful contribution to a longitudinal or a time-series analysis.

In summary, the piecewise linear model gave the most reasonable fit to the MD data as it accounted for the sharp MD decline before the stabilization of the IOP within the treated group, while the ITT and the PP approach fitted a linear model and the lag time was ignored. For the survival data, the multistate model provided an appropriate means of assessing the effect of IOP stabilization on the time-to-progression experience in normal tension glaucoma. It was particularly useful because it can flexibly accommodate the event of IOP stabilization whose time varied across the treated patients.

Chapter 9

Conclusions

For the analysis of clinical trials in the presence of a lag time in the stabilization of treat-ment, an efficacy-assessment approach entails comparisons between treatment groups af-ter the establishment of treatment stabilization. In evaluating the efficacy of a 30% IOP lowering therapy for normal tension glaucoma in our study, we adopted a per-protocol (PP) approach that compared a control group from the time of randomization and a treated group from the patients' individual times of treatment stabilization (IOP stabi-lization), and all the data from the treated patients collected prior to IOP stabilization were excluded from the analysis. As treated patients generally experience changes in dis-ease state during the period awaited for treatment stabilization, neglecting the lag time led to the invalidity of our PP approach for efficacy assessment. Based on the results from our simulation study of the MD data, we saw that failure to appropriately account for the lag time in the PP analysis can result in bias in parameter estimation and dimin-ishing power of statistical tests. Therefore, a plausible method for assessing treatment effects after the stabilization of treatment should take the lag time into consideration and make proper baseline time matching between the treatment groups. In any case, it is important to be aware of the fact that an adjustment made to the baseline times may produce heterogeneous treatment groups that differ with respect to certain base-line characteristics. Our proposed PCLIN approach in the longitudinal analysis and the multistate modelling approach with the use of a single Cox model with time-dependent treatement stabilization covariate in the time-to-progression survival analysis not only incorporated information on IOP stabilization into the modelling of the data, but also

followed the intent-to-treat (ITT) principle. By including all patients who were random-ized in the final analysis of the trial as in the case of an ITT analysis, both approaches tend to preserve homogeneity of the treatment groups and hence provide an unbiased means of treatment group comparisons. On the contrary, in using the PP approach, which analyzed a subset of patients/data which fulfilled the treatment efficacy criteria, randomization of the patients was lost and a reduced sample size resulted. Subsequent analyses of this subset of data was prone to selection bias and loss in statistical power.

In conclusion, when there is a delay in treatment stabilization, an ITT analysis that takes into account the lag time remains an ideal approach to the evaluation of a clinical trial. In analyzing the time-to-progression data, the multistate modelling approach is preferred to the baseline-adjustment approach because the former evaluates the overall clinical effectiveness of the IOP reduction therapy in an unbiased fashion. Although the latter seemed to provide a reasonable baseline time correspondence between the treated and the control groups, and showed a better efficacy assessment than our PP approach, the assessment was likely to be biased.

Chapter 10

Future Research

The main focus of this thesis has been on interpreting the difference in the results and comparing the performance of the intent-to-treat and the per-protocol analyses in evaluating clinical trials in the presence of a delay in treatment stabilization. The intent-to-treat (ITT) approach included all the patients who were randomized into the final analysis and patients were followed starting from the time of randomization. While the criteria according to the ITT principle as previously described compose the only definition of the ITT approach, there can be different ways of defining a PP approach because the treatment efficacy criteria which a PP approach is based on vary according to how experimenters view a therapy as efficacious. In the context of our study, a PP analysis aims to assess a treatment effect after the stabilization of treatment is established. We could have defined a different set of efficacy criteria for our PP analysis which assumes a different baseline other than the time of randomization for the control patients or includes a different subset of patients for analysis. Our baseline-adjustment approach used in analyzing the time-to-progression data adjusted the baseline times for the control patients by making use of the distribution which the time to IOP stabilization within the treated group followed. Exploring other possible methods of baseline adjustment for better time correspondence between treatment groups remains a future research topic on efficacy assessment.

In the simulation experiment, the mean defect data were generated based on the piecewise linear mixed effects model, and no missing data problem due to drop-outs was

addressed. In light of the presence of a possible floor effect imposed by the lower limit of the MD level that can be attained, all the simulated MD values that fall below the lower bound should be excluded from the analysis. We can treat patients as if they drop out of the study at the time when their MD reaches the lower limit, and consequently, in the simulation experiment, we could regard the MD data that fall outside the realistic range as non-ignorable missing data. In particular, a more negative decay rate will result in a higher proportion of missing data, and the impact the missing observations have on the results of the ITT and the PP analysis requires further research. Investigating such an impact through simulations can lead to a better understanding of the floor effect. Moreover, patients whose MD levels at baseline are close to the lower limit probably show a different trend of depression and especially a slower MD decay rate than those who have less negative MD at baseline. We might simulate different MD slopes for patients with a wide range of different initial MD levels and study the floor effect on the outcome.

To account for the delay in the treatment effect present in our data, we fitted a piecewise linear mixed effects model to the mean defect data within the treated group, and fitted the multistate disability model to the time-to-progression data by means of a time-dependent Cox model with an indicator variable for IOP stabilization. In both the longitudinal and the survival analyses, we assumed a threshold treatment effect. Different decay rates of the MD before and after IOP stabilization were assumed for the treated patients. The decay rate did not change gradually in the vicinity of the time of IOP stabilization; the treated patients followed a different decay pattern after IOP stabilization. Similarly, by using a time-dependent indicator variable for IOP stabilization in the Cox model, the treatment of having a 30% IOP reduction will be treated as not having taken any effect before the stabilization and as having reached its full effect afterwards. The abrupt change from no treatment effect to a full effect at the time of IOP stabilization might not be realistic and does not seem to describe accurately the effect of the IOP lowering treatment. We can consider a linear lag model or other non-threshold types of lag models for the treatment effect to better capture a gradual achievement of

a full effect over time. The use of lagged models was mostly discussed in the context of survival analysis. Implementing the same idea in longitudinal analysis is worthy of future work, and success in the implementation will enhance our understanding of the effect a treatment has on the course of a disease in the presence of a lag time in treatment stabilization . On the other hand, when analyzing the multistate disability model, we can apply a different approach proposed by Andersen [27], who used transition probabilities in the parameter estimation procedure. This approach involves fewer assumptions than the time-dependent Cox regression analysis that was applied in this thesis. It is also more applicable in general cases where the treatment groups do not have proportional hazards. However, the parameter estimation procedure is far more complex and computationally intensive.

Throughout this thesis, the MD data and the time-to-event data were analyzed separately. The MD level was correlated with the time to progression because the MD is a measure of the generalized visual field loss and the progression end point was determined by a persistent presence of a cluster of depression points in the visual field based on the four-of-five criteria adopted by the Collaborative Normal Tension Glaucoma Study. Being able to incorporate the information on the degree of generalized visual field loss into the analysis of the time-to-progression data will most likely improve the modelling of the hazard of progression and therefore provide insights into the prediction of the survival rates (without reaching the progression end point) for patients with normal tension glaucoma. In particular, as repeated measurements of the MD are available in our data, one can consider modelling the longitudinal data and the survival data simultaneously. Wulfsohn and Tsiatis [32] proposed a joint modelling approach which made use of a joint likelihood of the covariate repeatedly measured and the survival process for parameter estimation. They considered the linear mixed effects model for the covariate process and the Cox proportional hazards model for the survival process. The two models were used for the separate analyses of the MD data and the time to event data in our study and they were found to provide reasonably good fits to the data. Therefore, the joint

modelling approach seems to be readily applicable to our case.

All of the above recommendations for future research deal with the issue of a lag time in treatment stabilization in the evaluation of clinical trials. Some of the suggested analytical approaches such as adjusting baseline times for the control group are ad-hoc and do not have a general application in practice. If clinical researchers foresee a lag time for the achievement of a full treatment effect, it would be useful to come up with remedial actions at the design stage of a clinical trial. For example, problems of patient drop-outs before treatment stabilization and inability to achieve full treatment effect are associated with reduced sample sizes. Measures that correct for patient loss in the design of clinical trials will free researchers from future problems regarding insufficient statistical power of treatment comparison procedures. The corrective measures are especially advantageous as problems brought about by the treatment delay are often difficult to rectify after the trial is carried out. In any case, it is equally important to address the treatment lag issue at the design stage and in the analysis of a clinical trial.

Bibliography

[1] Halperin, M., Rogot, E. Gurian, J. and Ederer, F. Sample Sizes for Medical Trials with Special Reference to Long-term Therapy. *Journal of the Chronic Diseases.* 1968;21:13-24.

[2] Lipid Research Clinics Program. The Coronary Primary Prevention Trial: Design and Implementation. *Journal of the Chronic Diseases.* 1979;32:609-31.

[3] Self, S., Prentice, R., et al. Statistical design of the Women's Health Trial. *Controlled Clinical Trials.* 1988;9:119-36.

[4] Physicians' Health Study Steering Committee. *Physicians' Health Study Protocol.* Brookline Mass: Harvard Medical School, Department of Medicine;1983.

[5] Zucker, D. and Lakatos, E. Weight Log Rank Type Statistics for Comparing Survival Curves When There Is a Time Lag in the Effectiveness of Treatment. *Biometrika.* 1990;77:853-864.

[6] Luo, X., Turnbull, B. W., et al. Regression for Censored Survival Data with Lag Effects. *Communications in Statistics. Theory and Methods.* 1994;23:3417-3438.

[7] Collaborative Normal Tension Glaucoma Study Group. Comparison of Glaucomatous Progression between Untreated Patients with Normal Tension Glaucoma and Patients with Therapeutically Reduced Intraocular Pressure. *American Journal of Ophthalmology.* 1998;126:487-497.

[8] Collaborative Normal Tension Glaucoma Study Group. The Effectiveness of Intraocular Pressure Reduction in the Treatment of Normal Tension Glaucoma. *American Journal of Ophthalmology.* 1998;126:498-505.

[9] Collaborative Normal Tension Glaucoma Study Group. Natural History of Normal Tension Glaucoma. *Ophthalmology.* 2001;108:247-253.

[10] Collaborative Normal Tension Glaucoma Study Group. Risk Factors of Progression of Visual Field Abnormalities in Normal Tension Glaucoma. *American Journal of Ophthalmology.* 2001;131:699-708.

[11] Graefe, V. Uber die Iridectomie bei Glaucom und uber den glaucomatosen process. *Graefe Arch Ophthalmol.* 1857;3:456.

[12] Wilensky, J.T. and Gieser, D. K. Low-Tension Glaucoma. In: Weinstein G. W. ed. *Open-Angle Glaucoma.* New York: Churchill Livingstone;1986:58.

117

[13] The Glaucoma Research Foundation. *Gleams.* 2001;18(3):4.

[14] Sassani, J. W., ed. *Ophthalmic Fundamentals: Glaucoma.* New Jersey: Slack;1999.

[15] Andersen, K. E. and Jeppesen, O. K. *Classifying Visual Field Data.* Master's Thesis. Department of Mathematical Sciences, Aalborg University;1998:9.

[16] Cartwright, M. J. and Anderson, D. R. Correlation of Asymmetric Damage with Asymmetric Intraocular Pressure in Normal Tension Glaucoma. *Arch Ophthalmol.* 1998;106:888-890.

[17] Crichton, A., Drance, S. M., Douglas, G. R. and Schulzer, M. Unequal Intraocular Pressure and Its Relation to Asymmetric Visual Field Defects in Low-Tension Glaucoma. *Ophthalmology.* 1989;96:1312-1314.

[18] Haefliger, I. O. and Hitchings, R. A. Relationship between Asymmetry in Visual Field Defects and Intraocular Pressure Difference in an Untreated Normal Tension Glaucoma Population. *Arch Ophthalmol.* 1990;68:564-567.

[19] Heijl, A. Humphrey Field Analyzer. In: Drance S. and Anderson D. ed. *Automatic Perimetry in Glaucoma. A Practical Guide.* Orlando: Grune & Stratton;1985:132.

[20] Lachin, J. M. Statistical Considerations in the Intent-to-Treat Principle. *Controlled Clinical Trials.* 2000;21:167-189.

[21] Indrayan, A. and Sarmukaddam, S. B. *Medical Biostatistics.* New York: Marcel Dekker;2001:53.

[22] Burton, P., Gurrin, L, and Sly, P. Tutorial in Biostatistics: Extending the Simple Linear Regression Model to Account for Correlated Responses: An Introduction to Generalized Estimating Equations and Multi-level Mixed Modelling. *Statistics in Medicine.* 1998;17:1261-1291.

[23] Laird, N. M. and Ware, J. H. Random-Effects Models for Longitudinal Data. *Biometrics.* 1982;38:963-974.

[24] Kaplan, E. L. and Meier, P. Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association.* 1958;53:457-481.

[25] Cox, D. R. Regression Models and Life Tables (with Discussion). *Journal of the Royal Statistical Society. B.* 1972;34:187-202.

[26] Hougaard, P. *Analysis of Multivariate Survival Data.* New York: Springer;2000.

[27] Andersen, P. K. Multistate Models in Survival Analysis: A Study of Nephropathy and Mortality in Diabetes. *Statistics in Medicine.* 1988;7:661-670.

[28] Crowley J. and Hu, M. Covariance Analysis of Heart Transplant Survival Data. In: Hougaard, P. *Analysis of Multivariate Survival Data.* New York: Springer;2000.

[29] Cox, D. R. and Oakes, D. *Analysis of Survival Data.* New York: Chapman and Hall;1984. Chapter 8.

[30] Lawless, J. F. *Statistical Models and Methods for Lifetime Data.* New York: John Wiley & Sons;1982. 344-345,365,392-394.

[31] Collett, D. *Modelling Survival Data in Medical Research.* New York: Chapman and Hall;1994. 43-45.

[32] Wulfsohn, M. S. and Tsiatis, A. A. A Joint Model for Survival and Longitudinal Data Measured with Error. *Biometrics.* 1994;53:330-339.