Statistical Analysis of Treatment Compliance for Clinical Trials using Electronic Compliance Monitoring

Jean-Karl Sirois

Thesis submitted to the Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements for the degree of Master of Science in Biostatistics [1]

Department of Mathematics and Statistics
Faculty of Science
University of Ottawa

[1] The M.Sc. program is a joint program with Carleton University, administered by the Ottawa-Carleton Institute of Mathematics and Statistics

Abstract

Compliance, the extent to which patients follow a medication regimen, has been recognized as one of the most serious problems facing medical practice today. Recent developments in assessing compliance include electronic compliance monitors (ECM), devices that record the date and time of the release of medication from its original container. This allows utilizing ECM compliance data in statistical analyses related to clinical trials.

This thesis proposes ways of dealing with the time-varying nature of compliance. We examine the compliance behaviour from real ECM data through statistical analysis of compliance rate, followed by a time-to-event analysis with respect to first noncompliance event. Then, using discrete event simulation and proportional hazards models we compare analyses using a fixed treatment covariate and time-varying compliance covariate based on pharmacokinetic principles in estimating treatment effect. We observe a reduction of up to 40% in EMSE in favour of the latter model for treatment effect estimation.

Résumé

L'observance, la mesure selon laquelle les patients suivent un régime de traitement, est reconnue comme un sérieux problème dans la pratique médicale. La plus récente méthode pour mesurer l'observance est l'utilisation de moniteur électronique d'observance (ECM), un système qui enregistre la date et l'heure à laquelle le médicament sort de son contenant. Ceux-ci permettent d'utiliser les données ECM lors d'analyses statistiques liées aux essais cliniques.

Cette thèse propose des moyens de traiter la nature temporelle de l'observance. Nous examinerons le comportement de l'observance de vraies données ECM avec une analyse statistiques du taux d'observance, suivie d'une analyse temps-à-événement à l'égard de la première non-observance. Ensuite, en utilisant une simulation à événements discrets et des modèles de risques proportionnels nous comparerons des analyses avec une covariable de traitement fixe et une covariable d'observance temporelle basée sur des principes pharmacocinétique afin d'estimer l'effet de traitement. Nous observons une réduction d'erreur quadratique empirique (EMSE) allant jusqu'à 40% pour ce dernier lors de l'estimation de l'effet de traitement.

Acknowledgements

First and foremost, I would like to thank Dr. Allan Wilson and Information Mediary Corporation for supplying the thesis topic as well as the ECM data presented. Funding for this project was also provided in part through NSERC Discovery grants.

I would like to thank my parents for supporting me through thick and thin while completing my thesis and my brother for guiding me into the perfect field of study. I would also like to thank from the bottom of my heart my now wife Mireille for her continuous support, motivation and love.

Finally, I would like to express my sincere gratitude to my supervisors Dr. Pierre-Jérôme Bergeron and Dr. Mayer Alvo for their ongoing support throughout my Master's thesis, and whose wisdom and knowledge will forever surpass mine.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The last few decades have been filled with important phases in the development of approaches aimed at ensuring that patients who are afflicted with treatable chronic conditions continue their therapy for long periods of time. The World Health Organization [74] defines chronic conditions as follows:

"*Diseases which have one or more of the following characteristics: they are permanent, leave residual disability, are caused by nonreversible pathological alteration, require special training of the patient for rehabilitation, or may be expected to require a long period of supervision, observation or care.*"

The problem of compliance (or rather, noncompliance) extends to a variety of chronic conditions, from hypertension to cardiovascular disease. Initially, the patient was thought to hold the sole responsibility of the source of noncompliance. However, this idea of one-sided blame that assumes that patients should obey doctors' orders was refuted, explaining that from the patient's perspective, compliance is a medical construct with little meaning and that they should be encouraged to participate more actively in decision making regarding their health care [47]. This train of thought is

motivated by the initial study of noncompliance by Sackett and Haynes, stating that rather than seeing the healthcare provider's role as trying to get noncompliant patients to comply, we should emphasize the importance of negotiation within the provider-patient relationship [42]. The idea of compliance is thus associated closely with blame, be it of providers or patients. This leads to rising of the concept of *adherence* which, according to the World Health Organization, is *"a better way of capturing the dynamic and complex changes required of many players over long periods to maintain optimal health in people with chronic diseases"* [74]. Even though adherence is growing in popularity, *compliance* will be the term of choice for the purpose of this thesis.

Compliance research has focused on two different approaches: treating compliance as an outcome variable in order to determine which factors influence patients' compliance with a dosage regimen, and treating compliance as a predictor variable on health outcomes using the fact that compliance with a medication regimen is a key factor in determining whether or not the therapeutic drugs have a positive impact on chronic diseases. Although the latter has had more consistent results than the former, most of the research was focused on compliance as a fixed predictor, such as the proportion of prescribed doses taken or the proportion of days with the correct number of doses taken [46]. Melnikow and Kiefe suggest that since compliance may vary over time, it may be more appropriately studied as a dependent (outcome) variable. In order to treat compliance as an outcome however, we would require more information over time on the patients themselves, rather than their medication-taking behaviour. Nonetheless, whether compliance is treated as a predictor or as an outcome, a fixed compliance over the dosage regimen does not capture patient's medication-intake behaviour over time. This is why we can improve the modeling of patient's drug-intake using time-dependent compliance that can be produced using the information provided by an *Electronic Compliance Monitor* (ECM), which is the medication package with the electronic tag programmed to record the dose removal time, rather than a constant (mean) compliance over the treatment regimen. These ECMs will be further

described in Section 1.1.1.

There are very few studies that address compliance as a time varying predictor variable. Among these, Huybrechts *et al.* published a study assessing compliance with osteoporosis treatment as a predictor to bone fracture rates in a managed care population. A patient was defined as (time-dependent) compliant over a given period if their medication possession ratio (MPR) was $\geq 80\%$, where the MPR was calculated as the cumulative number of days during which the patient had medication available divided by the length of follow-up [35]. The patients' compliance profile was determined by the time of each prescription following the osteoporosis diagnosis. Assessment of the effect of compliance over of time was then done by fitting Cox proportional hazards models that included compliance as a time-dependent factor. Although Huybrechts *et al.* consider compliance as time-dependent, the patients' profile were determined at each *prescription*, which are given at each month. This compliance profiling thus does not capture daily medication intake behaviour, which can significantly vary within a same month.

The purpose of this thesis is to analyze the effect of a compliance predictor varying on a smaller, more continuous time scale on hypothetical health outcomes by investigating the impact of noncompliance on the estimation of treatment effects in RCTs. In order to accomplish this, we will analyze the available information provided by the ECMs to obtain patient's medication-taking habits as a time-dependent variable on a more refined scale rather than simply based on the time of each prescription. We can then utilize this variable in simulation studies in order to compare the many ways in which the information provided by the ECMs can be introduced in statistical analyses related to clinical trials.

In the next section, we start by providing the motivation behind the work in this thesis, including the history, the characteristics, and much more information on the ECM devices, and continue with a thorough description of various aspects of compliance and explain why they are of interest. We also present the data and its

initial processing. In Chapter 2, we will review basic pharmacokinetic principles and we will focus on the creation of a more sensible time-dependent compliance predictor based on real compliance data from the ECMs using pharmacokinetic models based on a "standard individual". In Chapter 3, we will review the theory of survival analysis. In Chapter 4, we will first consider basic exploratory data analysis in order to get familiar with the compliance data. We then follow with the Kaplan-Meier estimates to analyze the trend of compliance of all subjects across different medication variables, defining the "noncompliance events" as either an extra dose or a missed dose. We will continue with Chapter 5 by using the framework from Chapter 2 and by simulating a time-to-event study in order to compare different measures of compliance and discussing the results of the Cox's proportional hazards model that includes the proposed compliance indicator as a time-dependent factor. We will conclude the thesis with Chapter 6 with final thoughts and further possible studies.

1.1 Motivation

The topic of this thesis was based on work done by Information Mediary Corp. (IMC), a company that conducts research and development in the areas of pharmaceutical, logistics and supply chain compliance technologies. As IMC had presented the problem, patients using their electronic compliance monitoring devices, called *Med-ic*®, would be registered as having taken their medication on time if it fell within a compliance window of a given number of hours every day. The devices themselves are blister packs that record every time a pill comes out of the package as it activates a sensor. Compliance data generated by the ECM device are downloaded with a CertiScan® Reader and Software or any NFC- or Bluetooth LE-enabled device [73]. An example of the Med-ic ® "Smart" blister pack is shown in Figure 1.1.

For the analysis, IMC had initially proposed to utilize the data provided by the ECM devices to represent patients compliance based on a pharmacokinetic model.

Figure 1.1: Example of the Med-ic® Smart Blister Pack.
© 2014 Information Mediary Corporation, by permission

That is, they wanted to consider compliant patients as those patients who kept their medication concentration within the clinically effective level by taking their medication in the allowed therapeutic window. For this thesis, we decided to focus on the time dependence aspect, which has had relatively little discussion in the current literature. We explored several approches (such as graphical ways of displaying compliance and non-compliance based on ECM data, which are presented in Chapter 2 at the subject level and in Chapter 4 at the group level), to figure out how to best use this type of data in clinical trials. The data produced by the ECMs were seen as potential covariates in models where the outcome is clinically meaningful. The main goal initially was to see how different ways of coding the compliance would affect estimators of treatment effect (e.g. time-dependent versus fixed rate versus no compliance data), in models that could take time-dependent covariates.

The possibilities of analysis were narrowed down while keeping them within the

expectations of the company. The only data shared by IMC were device data (e.g. date and time of dose administration). The absence of any clinical data such as basic patient covariates or trial design restricted options to use compliance as a response, limiting us to using the variables that were available which could be related to compliance to test for potential relationships. Additionally, without the response of interest or standard clinical covariates, we could not use compliance as a covariate in real data analysis, which was the original intended purpose of this work. The only reasonable analysis was using compliance to produce a simulated clinical response (which can be seen in Chapter 5). It was then decided to do what was possible with the temporal aspect of the data by creating a time-dependent compliance covariate, which ultimately somewhat dictated our choice of model to survival analysis, where this time varying compliance variable would be easy to incorporate into the model and to interpret its significance. This motivates our choice of analyzing the first noncompliance event in Chapter 4, even though compliance is not an "all-or-nothing" phenomenon. It was the simplest type of event that could be defined in order to consider the data as independent observations while using standard survival analysis tools. Note that it is possible to define other events such as first time when a patient misses 2 (or more) consecutive doses, or conversely the first time a patient takes 2 or more doses for several consecutive days and carry out a similar time-to-event analysis. Although these events would be of interest, this exercise would not have been more informative with our current data, but in a longer clinical study it would be possible to better detect major noncompliance events based on ECM data.

Recurrent event analysis was considered until some exploratory data analysis was performed, revealing features such as a large number of ties in dose intake that would make defining recurrent events based on both "positive" (time to dose intake) and "negative" (time to missed or extra dose) events difficult to adapt for standard methods. In light of these issues, data organization, analysis and interpretation in the recurrent event context appeared impractical [12, 60, 69]. Note that the same

arguments can be made about a longitudinal model approach, although a longitudinal model with before-and-after measurements could have been fitted. This latter model however would not have regular assessments to go in accordance with the compliance measurements, and we would then drop the time varying aspect that we actually want to take advantage of. This is the reason why only the simpler models and analyses found in Chapters 4 and 5 were taken into consideration for this thesis, in accordance with IMCs wishes as well as what was computationally feasible within the given timeframe.

We now provide more detail on electronic compliance monitoring devices.

1.1.1 Electronic Compliance Monitors

Electronic compliance monitoring devices can take many forms. The underlying premise of these devices is that it is possible to monitor how often and when participants take their medications with a microprocessor embedded in a device such as a medication cap, a pill box, or by communication in real-time through transmission over a cellular network [72]. Each time the device is opened or activated, a record of that event is recorded, stored, and later uploaded to a computer for analysis. Electronic monitoring devices are considered as a possible gold-standard to which other compliance measures should be compared to [1]. The evidence for this claim can be seen in multiple published reviews comparing electronic monitoring device data to other forms of compliance measurement [2, 17, 61]. Other types for measuring compliance are discussed in Section 1.2.1.

Advantages of ECMs

As previously mentioned ECMs monitor the opening of a blister pack (as shown in Figure 1.1) containing medication and thus directly monitor a behaviour (i.e., the act of releasing a dose) that is thought to be a necessary precursor to dose intake

most of the time. To the extent that dose intake occurs in consistent and close proximity to device openings, this kind of assessment can provide the necessary data to closely mirror compliance. The association between ECM data and study drug intake provides considerable support for this assumption [72].

The main advantage of ECMs is that it can measure compliance *in real time*, which in turn creates the ability to track the timing of missed doses, to avoid error due to memory, and produces data that can capture patient's pill-taking behaviours [30]. Another key advantage is the granularity of data collected, which is unmatched by other assessment strategies and can be used to estimate rates of compliance across long periods of time and can estimate compliance within specific dose times as well as persistence with the treatment regimen.

Disadvantages of ECMs

Problems associated with ECMs include: their high cost, the possibility of malfunction, the possible interference with routine compliance activities, the inability to confirm ingestion of the medication, the inconsistent use of the electronic monitors, and the need to censor data [6, 23, 56]. Two key assumptions sustaining this method are that each recorded opening equals one dose of medication consumed and that the device is activated once and only once when each dose of medication is taken. These assumption are violated when subjects remove more than one dose per opening, sometimes known as "pocket dosing", and when subjects open the device multiple times without removing medication, referred to as "curiosity checks". Bangsberg *et al.* [4] describe a procedure for adjusting ECM data based on these phenomenons when electronic devices are paired with other types of compliance measurements; this is merely one of the reasons where it would have been interesting to include other compliance measurements in this thesis had such data been available to us.

Remark 1.1.1. Note that pocket dosing and curiosity checks do not affect our data

due to how the Med-ic® blister pack is constructed. Since each dose has its own sensor, any dose released will be captured, thus eliminating pocket dosing. Also, since the sensor is activated if and only if a dose is released, then curiosity checks (triggering the sensor without releasing a dose) are not possible with a functioning device. This is why this particular ECM has a promising future.

One other noteworthy problem to mention is the potential intervention effect associated with these devices, which is reflected in increased compliance immediately after participants begin using ECMs, which can take 35 to 40 days of continued use to normalize [16, 18]. For this reason, a run-in or practice period with the ECMs is recommended prior to starting data collection. This allows reactivity to the device, sensitization to compliance, and increased openings/releases not related to drug intake to dissipate. Note that a run-in period may or may not have been utilized before the data used in this thesis were collected, but it is impossible for us to confirm given the state of the received data.

In the next section, we proceed with an in-depth description of compliance with its medical, economical and statistical implications.

1.2 Compliance

The lack of compliance with medical regimens has been recognized as one of the most serious problems facing medical practice today. Noncompliance is of concern to health care professionals because of its detrimental effects on not only the quality of medical care, through unsuccessful therapies and interference with therapeutic relationships, but also on advances in medical science by threatening the validity of clinical trials of new treatments [27]. In fact, the efficacy of medication is confounded when patients do not adhere to prescribed regimens. Taking extra doses, missing doses and erratic dosing intervals can diminish drug action or cause adverse effects [13].

There have been several definitions of compliance over the past few decades; we will however use the definition provided by [67] described below.

Definition 1.2.1. ***Compliance*** *is defined as a dimensionless, blanket term encompassing the extent to which patients' drug dosing histories conform, or not, to their corresponding prescribed drug dosing regimen.*

Compliance can also be viewed in terms of the results of taking medication, as stated in [66].

Definition 1.2.2. ***Medication compliance*** *is defined as the number of doses not taken or taken incorrectly during a dosage regimen that jeopardize the therapeutic outcome or the point below which the desired therapeutic result is unlikely to be achieved.*

Definition 1.2.3. *A* ***dosage (medication, therapeutic) regimen*** *is defined as the decision of drug administration regarding formulation, route of administration, drug dose, dosing interval and treatment duration.*

Compliance can be broken into two main components: persistence and execution. Persistence is the length of time during which the medication is taken, i.e. the time from the first dose taken to the last dose taken. Execution is the multidimensional outcome of the comparison of the prescribed drug dosing regimen and the patient's drug dosing history while still engaged with treatment [67]. Since the length of the treatment regimen is inherently fixed by the treatment format (the number of doses to take is fixed, and therefore the number of days is as well), persistence should be the same for all perfectly compliant patients. However, persistence of patients can vary according to their execution. We can think of it in the sense that missing a dose can lead to missing more doses, which can potentially lead to treatment discontinuation (or an extension of time to the next prescription being needed).

Remark 1.2.4. The distinction between these two aspects of the patients' compliance to a prescribed regimen is crucial because the dynamics as well as the clinical

and economic consequences of poor quality of execution and short persistence can differ markedly [67].

Although a patient's persistence is not the main focus here, it will be discussed in Chapter 4 when discussing real world applications while analyzing "first missed dose" events. Our primary focus will however be on patient's execution towards a treatment regimen in order to analyze their compliance with respect to both noncompliance events and clinical events due to noncompliance.

1.2.1 Clinical aspect

History of compliance

The concept of clinical noncompliance was brought by Sackett and Haynes in the 1970s and reflected their role in the field of evidence-based medicine when attempting to explain disappointing responses to hypertension treatment. Their idea of noncompliance was conceptualized as a tragic problem potentially solvable by clinical research, inferring that noncompliant patients are seen as deviant [42]. In their initial study, they explored factors associated with noncompliance, focusing on its understanding, measurement and resolution [20]. As the number of effective medications grew, healthcare providers were inclined to focus their efforts on the problem of noncompliance. By the end of the 1970s, it was clear that determinants of compliance were complex and poorly understood, which lead to very few improvements despite continuing research since the 1980s. Compliance research has focused on the extent and determinants of noncompliance, and strategies to improve compliance. One of the most striking reasons for the lack of progress in compliance research is the absence of taking into consideration the patient's perspective [66].

Patient perceptions of compliance

Many theories about compliance locate the source of noncompliance in the provider-patient relationship, patient knowledge or beliefs about treatment. An alternative perspective was proposed by Conrad [11] stating that one should consider the patient's experience of illness, and the meaning of medication in people's everyday lives in order to comprehend the patient's thought process during a therapeutic regimen. Patients define compliance in terms of apparent good health and seek treatment approaches that are manageable, tolerable and effective. The main issue in compliance analysis is that the patient *has a right* to noncompliance. In other words, intelligent noncompliance exists in a clinical situation where a prescribed medication is intentionally not taken and the patient's reason for noncompliance appears rational when analyzed dispassionately [66]. Therefore, one non-analytic way to potentially improve compliance can be based on a closer understanding of the patients' experience of their illness and medication, rather than the perceptions and expectations of healthcare professionals.

Compliance impact on healthcare costs

The economic benefits of drug treatment are often demonstrated in the controlled setting of a clinical trial. These benefits may not be realized in day-to-day practice, especially for patients who are, to a certain extent, noncompliant with their prescribed therapy. In fact, when medical conditions are treated suboptimally (i.e. when a patient is noncompliant), symptoms and complications may worsen, leading to increased use of medical care [62]. This suggests that higher levels of medication compliance may have positive economic value for some conditions. In fact, increased compliance may generate medical savings that more than offset the associated increased in medication costs [40].

Measuring compliance

Although the idea of compliance has been around for several decades now, there still does not exist a true "gold standard" for measuring compliance with medication (as mentioned in Section 1.1.1, electronic compliance monitoring devices have been *proposed* to act as the leading, golden-standard of compliance measurement, however there is still much debate about this claim), although the list of methods is quite lengthy already. The first widely used method, qualified as an *indirect measurement*, consists of the simple patient interview, also known as self-reports. Interviews are inexpensive and on the surface appear to be a straightforward method of monitoring compliance. However, studies [8] and [15] have shown that patient-reported compliance tend to substantially overestimate the true compliance. Nonetheless, it was concluded that interviews in general are probably a fairly specific method for confirming noncompliance but are not sensitive enough to rule it out.

Another indirect measurement consists of pill counting, also considered as pharmacy refills; this is because the pharmacist will count the number of pills returned when the prescription needs to be renewed, and cross-reference with the number of days passed since the previous refill. In order for pill counting to be a reasonably reliable compliance measurement, studies have shown that pill counts conducted unobtrusively on home visits announced on short notice may give more accurate results, because patients have less opportunity to discard unused medication [46]. While pill counts potentially give a fairly accurate representation of overall compliance during the dosage regimen, they cannot distinguish a perfectly compliant individual from a patient who misses some doses and makes up for it the following day.

The next method is a *direct measurement* consisting of monitoring the serum or urine levels of a drug, which is objective and not readily subject to intentional dissimulation by patients. This measuring method, however, is not popular mainly due to high cost and the logistics of obtaining samples. Another con of this method

is that variation in individual absorption or metabolism may be confused with under- or over-compliance. Additionally, per visit enhancement of compliance (also known as the toothbrush effect, where people are more likely to brush their teeth before visiting the dentist) tend to falsify a serum level evaluation, and thus compliance, on a scheduled measurement visit. Cramer *et al.* suggested in their study that serum levels plotted over time must consist of some measurements obtained unexpectedly at home visits in order to create an accurate picture of compliance [14].

The last method mentioned here is the one we are interested in and which is described in Section 1.1.1: the use of electronic medication dispensers in order to record the date and time of each opening. Although this is also an indirect measurement, it is far less subjective than the previous indirect methods. Assuming that at each such opening one dose is removed and ingested, this method yields a complete dose history. As previously mentioned, the main issue is that this assumption is critical: a patient could open the container but not take the drug or take a different dose than the one prescribed. While the opening of a dispenser does not guarantee the use of the medication inside, the information provided by the pattern of dispenser openings over time is more detailed than that obtained from a pill count.

These are the four main compliance measurement methods; for a complete list, see [46] and [36].

In this thesis, the initial compliance index was provided using electronic medication dispensers. In Chapter 2, we will discuss how to use this fixed compliance index in order to create a time-dependent compliance variable based on pharmacokinetics.

1.2.2 Statistical aspect

Compliance as an outcome

The initial research and the most common approach to the study of non-conforming health behavior have been to identify individual factors that are associated with var-

ious forms of noncompliance. The assumption was that "noncompliant patients possess a unique set of characteristics that differentiate them from compliant patients" [36]. Based on this assumption, compliance research was first focused on discovering individual factors that identify compliant versus noncompliant patients, factors such as social, demographic, education, religion, socioeconomic status, income, medical, and personality variables. Although no single factor can predict whether a patient will comply with medical advice, there have been some significant findings, but they are greatly outnumbered by studies where there was no statistical association between these factors and compliance [32]. The problem with this individualistic model approach is its failure to take into consideration the possible interactions between behavioral, psychological, environmental, structural and physical variables.

One approach to deal with this problem was the use of multivariable models, namely the popular *Health-Belief model* (HBM). The application of the HBM has been first and foremost to explore the relationship between compliance and the patients, stating that compliance is thought to be determined by the knowledge and attitudes of the patients [66]. Patients must believe that they are vulnerable or susceptible to the disease or its consequences, that they actually have it, and that the consequences of the disease on their well-being could be serious. They must believe that by following a particular set of health recommendations the threat or severity of the condition will be abolished or reduced [29]. According to the HBM, the patient's perception of compliance is the key factor that can potentially explain compliance, as mentioned earlier, and that the possibility of a rational noncompliance is real. The Health-Belief model however tends to neglect the possibility that noncompliance can be unintentional [36].

Although many studies have investigated causal relationships between patient and doctor factors and compliance, the findings are not consistent and thus compliance as an outcome variable yields no significant conclusions as of yet.

Compliance as a predictor

While research resulted in inconsistent findings when analyzing compliance as an outcome variable, more focus was put on adressing compliance as a predictor variable. One of the first and probably most impactful articles discussing compliance as a predictor variable was Efron and Feldman's paper on compliance as an explanatory variable in clinical trials [22]. They point out that the main difficulty in treating compliance as a predictor variable is that compliance (and hence amount of active dose) is not assigned in a randomized fashion by the investigators, as it should be in a genuine dose-response experiment. Due to this special nature of compliance, Efron and Feldman focus on three important characteristics not shared by "standard" covariates.

1. Compliance is an *adjustable* variable. In other words, one could change patients' compliance to whatever they wish without changing the sample pool of patients (unlike a variable like age).

2. Compliance has a different meaning in the treatment of placebo control groups. Compliance determines the amount of active drug taken for treatment group patients and also indicates something about the patient's psychological status (as mentioned in the HBM). In the placebo control group, since there is no active drug taken, only the psychological component of compliance applies.

3. The treatment and control groups should behave identically if patients in the treatment group are 100% noncompliant, since zero compliance indicates no active drug.

Remark 1.2.5. Note that if a treatment already exists for the condition, a clinical trial will be conducted using the existing treatment as a control. Only then will compliance have the same meaning in both control and treatment groups.

Since patients' compliance in this thesis is already recorded, only points 1 and 3 are of primary interest for us when treating compliance as a predictor variable. Point 2 can be used when simulating a scenario where we are comparing compliance between a treatment group and a placebo controlled group.

There have been several studies since that have examined the statistical main effect of compliance on outcome, as well as the potential interaction between compliance and assignment to treatment or placebo groups. In the majority of the studies ($> 80\%$), compliant patients had better outcomes, even if assigned to the placebo group [46]. This can, in part, be explained by the placebo effect, those effects of a treatment that are not attributable to the mechanics of the treatment itself, but rather to the circumstances surrounding it [55]. In other words, compliant patients are more likely to follow other instructions provided by their physicians compared to noncompliant patients, which in turn should produce a beneficial effect (medical or other) on the patients health outcome. This means that the patients subjective judgement of effectiveness is still important (as mentioned in Section 1.1.1), because it influences the patients compliance with the physicians prescription [63]. Furthermore, subjectively expected effectiveness is the crucial factor behind the placebo effect, which can sometimes lead to a significant objective improvement in patients health outcome [45]. This is where patient compliance and placebo effects become critical: if a patient is told a plausible story about the underlying mechanism of his complaint and how it is affected by the treatment, expectations and consequently drug compliance and placebo effect should be enhanced. This can be tested by integrating placebo effects and drug compliance as dependent variables into clinical trials. Since we do not have placebo effect data, this thesis is trying to adequately use compliance as a dependent variable into clinical trials; but this is something that has been tried and criticized [52]. The main issue is that compliance may (and most likely does) vary over time in a way that is dependent on other variables, such as side-effects of the active drug or, as just mentioned, expected effectiveness of the active drug. Not much

research has been done using compliance as a time dependent predictor, but many studies conclude their discussion with the proposal of repeating the analyses with a time dependent compliance variable, which further motivates this thesis.

The next section discusses compliance in clinical trials.

1.2.3 Compliance in Randomized Clinical Trials

Definition 1.2.6 ([25]). *We define a* ***Clinical Trial*** *as a prospective study comparing the effect and value of intervention(s) against a control in human beings. We define a* ***Randomized Clinical Trial (RCT)*** *as a Clinical Trial utilizing random allocation of participants between the intervention group and the control group.*

RCTs are comparative studies with an intervention group and a control group; the assignment of the subject to a group is determined by the formal procedure of randomization. Randomization, in the simplest case, is a process by which all participants are equally likely to be assigned to either the intervention group or the control group. There are three main advantages with using RCTs compared to other methods of selected controls:

- Randomization removes the potential of bias in the allocation of participants to the intervention group or to the control group.
- Randomization tends to produce comparable groups; that is, measured as well as unknown or unmeasured prognostic factors and other characteristics of the participants at the time of randomization will be, on average, evenly balanced between the intervention and control groups.
- Randomization guarantees the validity of statistical tests of significance.

Taking these into account, RCTs are considered the most powerful design for evaluating health care treatment effectiveness since they use a systematic methodol-

ogy that allows comparisons between groups [3]. However, in RCTs, noncompliance can be one of the major barriers to achieving statistical power to detect intervention effects [37]. Noncompliance is very common in clinical practice and it is unrealistic to expect perfect compliance in a trial. Non-compliance in RCTs occurs when subjects do not remain in the assigned randomized group, or when subjects do not receive their assigned treatment in full. Subjects can still provide study measurements (laboratory parameters, or efficacy parameters for example) without receiving the assigned intervention. In fact, measurements obtained from subjects not compliant with the trial can add valuable information for those subjects and can provide a more accurate estimation to clinicians of the treatment effect as it happens in a real life situation. This is the ideal for a full "Intention-to-Treat" (ITT) analysis.

One of the most accepted methods used in clinical research to address the problem of noncompliance is the ITT analysis. ITT is a strategy used to analyze the results of an RCT that considers the subjects in the way they were randomized at the beginning of the trial regardless of whether they completed the intervention given to their group, their compliance with the entry criteria, the treatment they actually received, or whether they withdrew from treatment or deviated from the experimental protocol [24]. According to Pocock and Abdallah [52], incorporating compliance data in the analysis of a trial is a valuable complement to the ITT analysis. Many investigators use the Per-Protocol (PP) analysis, which only analyzes the patients who comply with the treatment protocol. However, problems with this analysis arise when the reasons for noncompliance to the protocol may be related to the prognosis. It had been shown that patients who generally adhere to treatment intervention tend to do better than those who do not, even after adjusting for all known prognostic factors and irrespective of the group to which they were assigned (i.e. active treatment of placebo) [33].

According to the clinical literature, generally, compliance is associated with background health characteristics related to the outcome and response. Thus, when one

only considers the compliers in the analysis (following a PP analysis), one is analyzing only those who generally have better outcomes [38]. Another way of looking at it is that ITT provides an unbiased assessment of the effectiveness of the intervention at the level of compliance observed in the trial. However, ITT could also underestimate the treatment effect in compliant patients when treatment is effective but noncompliance is substantial. Moreover, an ITT analysis does not take into account the actual time of subjects on-study. As stated by Tillmann *et al.* [65], in the conventional ITT analysis, the results are independent of the actual time a subject complies with the treatment since every participant will contribute data at the moment of the evaluation. Tillmann *et al.* analyzed long-term mortality in a large-scale clinical trial performed in Australia. They added data regarding compliance and actual time on trial medication to the traditional ITT analysis using an ITT-based Cox proportional hazard model (resulting in a *modified Intention-to-Treat* (mITT) analysis). Thus, compliance and actual trial medication were accounted for in the results obtained in the trial. They found that incorporating compliance data into the conventional ITT analysis resulted in a more accurate effect size estimation. According to these authors, this approach may strengthen the validity of the conventional ITT analysis and may be applicable to other similar situations.

Based on this information, we will later see in Chapter 5 how we can use compliance data in a hypothetical RCT simulation study.

1.3 Methods

1.3.1 The data

The data for this thesis were provided by IMC to not only quantify and model patient compliance, but to analyze the effect of compliance (or rather, the effect of noncompliance) on a hypothetical treatment outcome by simulating clinical event times as

well. The data were first received in two samples containing observations of 27 and 28 different ECM IDs, respectively. Once initial analyses were done, two larger data sets were made available, the first containing 173 different ECM IDs and the second containing 908 distinct ECM IDs. Each patient possessing an ECM ID was instructed to take a total of three oral doses (sometimes referred to herein as pills) per day (2 morning doses and one evening dose) during a 4 week period, partitioned by four packages each containing 7 pills of all three doses. In other words, patients were given 4 Smart Med-ic® packages each containing 21 pills (14 morning pills, 7 evening pills) at the beginning of the study for the duration of the treatment. Additionally, each patient assigned to an ECM ID was given a "spare" package (i.e. 5 packages in total, constituting a patient's full medication or treatment kit) in order to replace unusable pills (i.e. if a pill is lost) or to continue the treatment when the following package was unavailable - this spare package was also associated with the same unique ECM ID as the other 4 associated with that patient. During this period, every time a dose was released from its package, a date and time for the appropriate dose was registered in the IMC database. This then produced multiple observations for each patient over the treatment regimen. It is important to note that the given data cannot in any way divulge any patient personal information that could lead to identification of an individual due to its nature. In other words, patients cannot be traced back from the data, which allows us to use the data without any appeal to an ethics committee. Also, we are assuming here that the 4 samples are independent from each other, but we have no way to determine if this is actually the case.

Remark 1.3.1. It is worth clarifying that a Smart Med-ic® package *is considered as an ECM*, meaning that each package is a device that records the date and time at which a pill is released. The patients receive a total of 5 (4 + 1 spare) packages associated with the patient's unique ECM ID.

The raw data from IMC included 9 different variables:

1. **ECM Index**: Variable indicating the ECM package number which increments across all patients.
2. **ECM ID**: Variable indicating ECM ID number and current associated package number.
3. **Dose Index**: Variable indicating the current dose number with the associated ECM ID.
4. **Dose Date**: Variable indicating the date at which the associated dose is taken.
5. **Dose Timestamp**: Variable indicating the local time at which the associated dose is taken.
6. **Dose Timestamp UTC**: Variable indicating date and coordinated universal time at which the associated dose is taken.
7. **Dose Group**: Variable indicating whether the associated dose is a morning dose or an evening dose.
8. **Dose Label**: Variable indicating the day and the dose number for the associated dose.
9. **Compliant**: Variable indicating whether the associated dose is taken at an acceptable time, i.e. dose compliance.

Now, the only variable that is not inherently defined above is the compliance. The requirement for dose compliance is that the dose must be taken at a certain time and within an allowable tolerance window. This tolerance window can be determined by either "first dose taken", where the compliance window is fixed by the first dose taken by the patient, or "fixed time", where the compliance window is a pre-specified time interval (e.g. between 8am and 10am) attributed to the patient. The binary compliance variable was then defined by whether or not the patient administered the dose in the allowed tolerance window. With this compliance variable, IMC has been able to monitor patient compliance during the period measured by the ECMs

in several pill counting ways. They define, amongst others, the *dose compliance score against packed doses*, which is the ratio of compliant doses taken over the total number of doses in the ECM, and the *dose compliance score against taken doses*, which is the ratio of compliant doses taken over the total number of doses taken. Both these variables contain information about the overall compliance of patients for a given ECM, but none contain information on the **evolution** of patient compliance within the period covered by the ECM. In order to discover information about the latter, some data processing is needed.

1.3.2 Initial data processing

The raw data from IMC were given in a comma-separated values (.csv) file. First, the ECM ID variable was separated into two; one retaining the ECM ID itself and the other indicating the package ID. The universal timestamp variable was removed to avoid redundancy since local time represents already the habits of pill taking of each patient (although this reasonably assumes that the patients remain in the same time zone during the entire regimen). The local timestamp is also a better indicator of what pills were taken in the morning versus what pills were taken in the evening, although this information in our cases is of little importance because of the first-dose-taken scenario. Local time can however be useful if we had a clinical question, for example if the patient is better off taking a dose in the morning or in the evening.

The local timestamp variable was "cleaned" of unnecessary information (i.e. the date, since there already is a date variable). The dose label variable was also separated into two; one retaining the dose number associated with the pill (morning 1, morning 2, or evening) and the other indicating the day number of the associated dose within the current package. Finally, a "type of noncompliance" variable was added, reading 0 if the dose is compliant, 1 if it is an extra dose, and -1 if it is a missing dose. A dose was considered as an extra dose if it was administered outside of the allowed

therapeutic window, or if it was administered within the same window as the previous dose (i.e. taking multiple doses per window). A dose was considered as missing if there was no dosing time registered for that day. Other definitions of extra doses and missed doses are possible but were not implemented. This left the processed data sets with 11 variables: ECM Index, ECM ID, Package ID, Dose Index, Dose Date, Dose Group, Day Label, Dose Label, Dose Timestamp, Compliant, and Type.

Remark 1.3.2. For missing doses, we mention that there is no dosing time registered for that day (or window). This means that there *should* be at least one observation each day containing the date, **regardless of whether a pill is taken or not**. If the pill is taken, the observation contains all the variables mentioned in Section 1.3.1. If no pill is taken, only one dose date appears, and all other variables are left empty within the observation.

The next step was to remove any outlying observations. In this case, outliers consisted of ECM IDs with a single observation. The reasoning behind this is simple: as mentioned above, there should be an observation associated with every day during the treatment period, whether the dose is taken or missed, which means that if there are no doses taken after the first one, we will have observations with the dates, but with no other covariates, so we know if and when there are missing days. If there is only a single observation for a given ECM ID, a likely explanation is a device malfunction, or device failure. Furthermore, since the compliance index variable proposed by IMC is on a "first dose taken" basis, the dose is automatically taken at a compliant time since it is the first and only one. A single dose is thus irrelevant for the purpose of analyzing compliance over time. There were a total of two unique ECM IDs omitted from the analyses across all four samples (to be more specific, both were from the sample with 908 subjects, reducing the sample size to 906).

The final step in the initial data processing was creating numerical time values of each dose timestamp. In order to do this, the package ID and the day label vari-

ables were needed to maintain continuity of days within a same ECM ID. The dose timestamp variable was then used as the fraction of the day at which the dose was administered, creating a numerical value with hours as units. If a dose was missed, then the assigned time of the missed dose was determined by the time of the last non-missing dose of the same type (morning or evening), and then adding the number of hours proportional to the number of days missed, for each day. For two consecutive days missed, a missed dose timestamp would be added 24 hours after the last non-missed dose, and a second missed dose timestamp added 48 hours after the last non-missed dose.

We now have the numerical times at which patients release their doses from the packages, or when a dose is considered as missing. Assuming that releasing a dose from a package is followed directly by the administration of the dose (which will be referred to henceforth as the *instant administration assumption*), we can then estimate, using pharmacokinetic principles, when the medication will fall below its therapeutic level. Using the compliance information of patients, we can explore when noncompliance can be clinically linked to a loss of the drug's therapeutic effect.

Remark 1.3.3. Note that even without direct measurements of pharmacokinetic aspects, it would still be possible to create a time-dependent compliance variable based on missed or late doses. This variable would, of course, be less time-sensitive than the one we define in Chapter 2, but we would be able to construct the former solely from dosing times. This model based on "gap times" between doses would however not be robust enough to capture the effects of taking extra doses compared to the model presented in Chapter 2.

Remark 1.3.4. It is important to realize that a drop in medication level below its therapeutic threshold is not always due to issues with patient compliance. Patient's metabolism is often a key factor is maintaining the beneficial effect of a drug. This

will be discussed in Section 2.1.2.

In the next Chapter, we will review pharmacokinetic principles and investigate how compliance by pill count can be associated with a medication's therapeutic effect.

Chapter 2

Pharmacokinetics and time-dependent compliance

The nature of pharmacokinetics can be understood as a fundamental scientific discipline that gives substantial support to applied therapeutics. Patients are prescribed appropriate medication regimens based on evidence-based medicine in order to treat a clinical condition. The designs of these medication (or dosage) regimens are created based on the understanding of the drug use process (DUP). When a patient is showing specific clinical signs and symptoms, the pharmacist applies the DUP to ensure that the patient is prescribed an appropriate medication regimen. Once a particular medicine is chosen, the principles of clinical pharmacokinetics are required to ensure the appropriate formulation of drug is chosen for an appropriate route of administration. However, since each patient handles a drug differently than another, the development of a dosage regimen for a particular patient requires an understanding of the pharmacokinetic processes of absorption, distribution, metabolism and excretion, referred to herein as the ADME processes.

Definition 2.0.5. ***Pharmacokinetics*** *is defined as the kinetics of drug ADME processes and their relationship with the pharmacological, therapeutic or toxicological*

response in humans. ***Clinical pharmacokinetics*** *is the application of pharmacokinetic principles to the safe and effective therapeutic management of drugs in an individual patient.*

Definition 2.0.6. ***Absorption*** *is defined as the process of movement of unchanged drug from the site of administration to systemic circulation (or to the site of measurement, i.e. plasma).*

Definition 2.0.7. ***Distribution*** *is the reversible transfer of a drug between the blood and the extra vascular fluids and tissues.*

Definition 2.0.8. ***Metabolism (or Biotransformation)*** *of drugs is defined as the chemical conversion of one form to another.*

Definition 2.0.9. ***Excretion*** *is defined as the process whereby drugs and/or their metabolites are irreversibly transferred from internal to external environment.*

Note that the process of **elimination** is the combined effects of metabolism and excretion, and can be defined as the major process of removal of a drug from the body and termination of its action.

A fundamental understanding of the ADME processes paired with the concentration of the therapeutic drug in the body determines the effectiveness of a dosage regimen. Drug concentrations are normally measured in whole blood from which serum or plasma is generated due to the inaccessibility of the site of action of the drug. It is however assumed that drug concentrations in the plasma are in equilibrium with the drug concentration at the site of action. It is also important to mention that the measured drug concentrations in plasma are often referred to as drug levels. There exists a variety of techniques used to represent the pharmacokinetics of a drug. The first part of this Chapter covers the basic theory in order to create a pharmacokinetic model.

2.1 Basic pharmacokinetics

The ultimate goal of drug therapy is to achieve efficacy without toxicity. The development of strong correlations between drug concentrations and their pharmacologic responses has enabled clinicians to apply pharmacokinetic principles to actual patient situations. This involves achieving a plasma concentration level, denoted as C_p within the "therapeutic window". This window consists of two thresholds: the minimal effective concentration (MEC) and the minimal toxic concentration (MTC). Within this therapeutic range, the desired effects of the drug are seen. Below it, there is greater probability that the therapeutic benefits are not realized; above it, toxic effects may occur. The use of clinical pharmacokinetics in this matter consists of the all the factors that determine variability in the C_p and its time-course. Figures 2.1 and 2.2 depict ideal and problematic drug therapeutics.

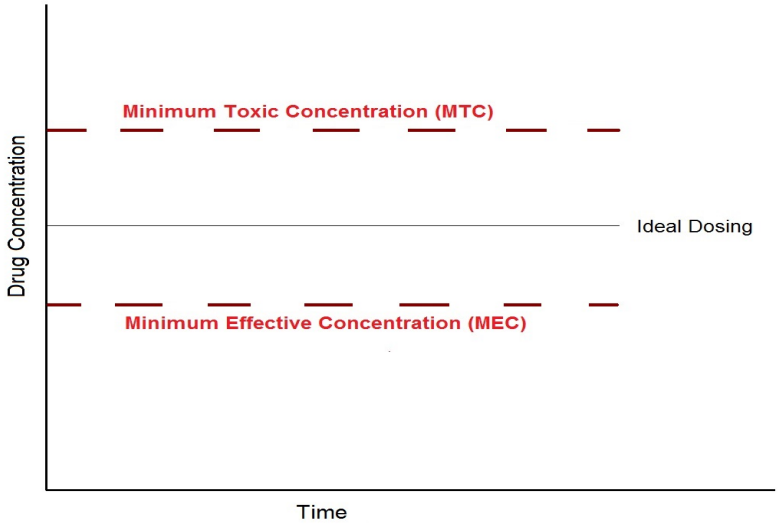

Figure 2.1: Ideal therapeutics: efficacy without toxicity.

Definition 2.1.1. *The minimum concentration of drug in the plasma required to produce the desired therapeutic effect is called the* ***Minimum Effective Concentration (MEC)****. It reflects the minimum concentration of drug at the site of action*

to elicit the desired pharmacological response. The concentration of drug below the MEC is said to be in the sub-therapeutic level.

Definition 2.1.2. *The concentration of drug in the plasma above which adverse or unwanted effects are precipitated is called the* ***Minimum Toxic Concentration (MTC)****. It is also known as maximum safe concentration.*

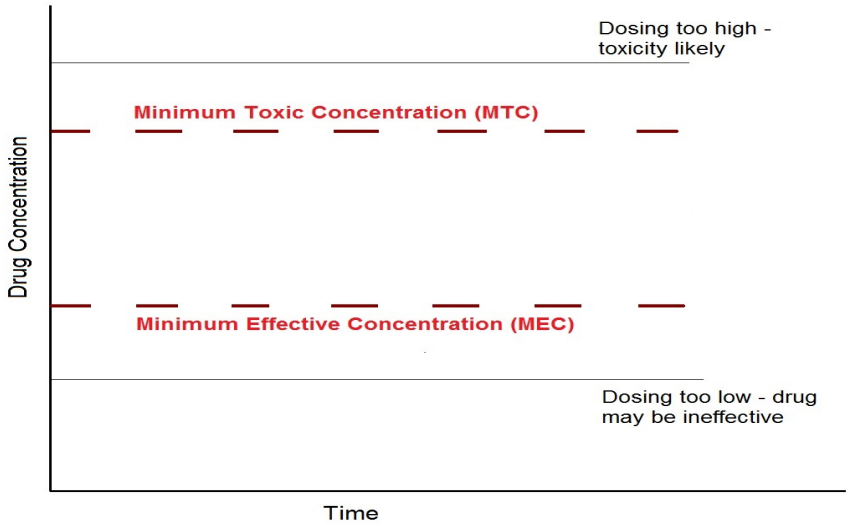

Figure 2.2: Problematic therapeutics: inappropriate dosing.

In many cases, the MTC might not exist, or more likely it might be so high that it becomes practically irrelevant. For the purpose of the definitions in section 2.2 and the analyses found in Chapter 4, it is assumed that the MTC is never attained. In order to analyze the relationship between the C_p and time, several parameters must be defined, as discussed in the next section.

2.1.1 Pharmacokinetic parameters

To consider the ADME processes is to consider their *rates*, which can be characterized by two basic underlying concepts. The rate of a reaction or process is defined as the

velocity at which it proceeds and can be described by n^{th} order reaction rates. Among them are the *zero-order* or *first-order* reaction rates, the two simplest kinetics with the most relaxed assumptions [19].

Consider a zero-order reaction. Then the C_p would decline linearly with time, i.e. the reaction proceeds at a constant rate and is independent of the concentration of the drug present in the body.

Proposition 2.1.3. *Let drug A have a zero-order reaction from the body. If the amount of drug A is decreasing at a constant rate, then the rate of elimination of A can be described by the differential equation*

$$\frac{dA}{dt} = -k^* \tag{2.1.1}$$

where k^* is defined as the zero-order elimination rate constant.

This situation does occur (for example, the elimination of alcohol from the body), but only rarely. The more common situation is first-order reaction, in which the decline in C_p is not constant with time, but varies with the concentration. The reaction would then proceed at a rate that is dependent on the concentration of the drug present in the body. It is assumed that the ADME processes follow first-order eliminations and that most drugs are eliminated in this manner.

Proposition 2.1.4. *Let drug A have a first-order reaction from the body. If the amount of drug A is decreasing at a rate that is proportional to A, then the rate of elimination of A can be described by the differential equation*

$$\frac{dA}{dt} = -kA \tag{2.1.2}$$

where k is defined as the first-order elimination rate constant.

Most drugs used in clinical practice at therapeutic dosages will show first-order rate processes. In fact, as the amount of drug administered increases, the body is

able to eliminate the drug accordingly and accumulation will not occur.

The handling of a drug by the body can however be very complex, as the ADME processes work to alter drug concentrations in the plasma. Simplifications of body processes are necessary to predict a drug's behaviour in the body. The most common way to make these simplifications is to apply mathematical principles to the various processes.

2.1.2 The single compartment model

In order to apply mathematical principles, a model of the body must be selected. The most basic type of model used in pharmacokinetics is the compartment model, which are categorized by the compartments needed to describe the drug's behaviour in the body. The compartment models are said to be deterministic, in the sense that the observed drug concentrations determine the type of compartment model required to describe the pharmacokinetics of the drug.

The one-compartment model is the most frequently used hypothetical structure in clinical practice to describe the course of a drug in a biological system following its administration. Figure 2.3 illustrates the structure of the one-compartment model, a model that depicts the body as a kinetically homogeneous unit. It assumes that after the drug is administered it distributes instantaneously to all body areas, i.e. that the drug equilibrates instantaneously between tissues. Note that this does not imply that the C_p is equal to the drug concentration in the tissues, but that changes in the C_p quantitatively reflect changes in the tissues.

Definition 2.1.5. *Consider a dose D administered as an ester or salt, i.e. a combination of two or more drugs. Suppose that D consists of drugs A and B. Suppose that drug A is the parent drug. Then the proportion of drug A contained in D is defined as the* ***salt factor****, denoted as S.*

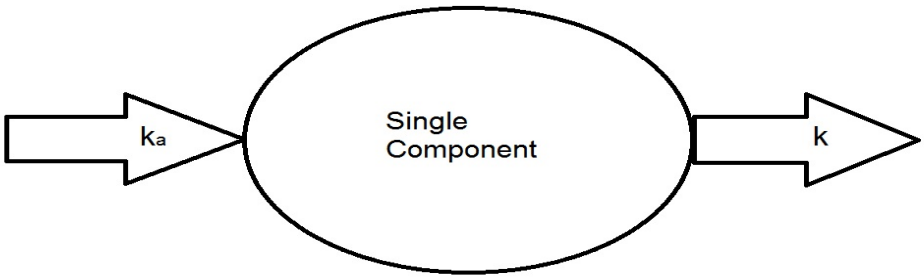

Figure 2.3: One-compartment model, where k_a is the absorption rate constant and k is the elimination rate constant.

Definition 2.1.6. *The proportion of an administered dose that is absorbed chemically unchanged into the systemic blood circulation is defined as the* ***fractional bioavailability*** *(or simply bioavailability), denoted as* F*.*

Definition 2.1.7. *The volume of plasma in which the total amount of drug in the body would be required to be dissolved in order to reflect the* C_p *is defined as the* ***volume of distribution****, denoted as* V*. In the context of oral dose administration, only the* ***apparent volume of distribution*** *is identified, denoted as* V_d*, defined as*

$$V_d = \frac{V}{SF} \tag{2.1.3}$$

where S *and* F *are the salt and bioavailability factors, respectively.*

Although the body can be modeled by a one-compartment structure, it is not by all means a homogeneous unit. It is important to note that the C_p is not necessarily

the same as the drug concentration in the liver, kidneys or other tissues for example. However, the changes in C_p are proportional to the changes in the amount of drug in the tissues. This results in

$$V_d = \frac{X}{C_p} \qquad \Longrightarrow \qquad C_p = \frac{X}{V_d} \tag{2.1.4}$$

where X is the amount of drug in the tissues.

Although two-compartment models and multi-compartment models exist, the latter do not have an explicit solution to the differential equations which solve the rate of change of the amount of drug over time [28], unlike the one-compartment model for oral doses found in Section 2.1.3. Although it is true that one cannot develop an acceptable model without knowing the details of the metabolism and distribution of the drug under study, we had no access to this type of information for this thesis. More complex models were considered but implementing them to obtain meaningful results was beyond the scope of this work and the expectations of IMC.

In any case, the plasma concentration-time profile of a large number of drugs can be described by a one-compartment model with first-order absorption and elimination rates. Admittedly, the model is best described when administering a drug through intravenous therapy [50], although a generalization for drugs that are taken orally can be defined, as described below.

2.1.3 Single and multiple oral doses models

We will now consider a one-compartment model where the drug is administered orally. Note that the one-compartment model can also be used when the drug is administered by injection (eg. insulin for treating diabetes), and its solution to the differential equation representing the change in drug concentration is simpler than the one we provide for an oral dose [19].

Now, consider the concentration versus time profile following a single oral dose. Assuming first-order absorption and elimination rates, the rate of change of amount of drug in the body can be described as

$$\frac{dX}{dt} = k_a D - kX \tag{2.1.5}$$

where D is the amount of drug at the absorption site (i.e. the administered dose), X is the amount of drug in the body, and k_a and k are the absorption rate and elimination rate, respectively, where $k_a > k > 0$. Solving this differential equation yields

$$X = \frac{Dk_a[\exp(-kt) - \exp(-k_a t)]}{k_a - k} \tag{2.1.6}$$

Next, using equations (2.1.3) and (2.1.4), we can convert X to C_p as follows

$$\begin{aligned} C_p(t) = \frac{X}{V_d} &= \frac{Dk_a}{(k_a - k)V_d}[\exp(-kt) - \exp(-k_a t)] \\ &= \frac{SFDk_a}{(k_a - k)V}[\exp(-kt) - \exp(-k_a t)] \end{aligned} \tag{2.1.7}$$

Equation (2.1.7) is the closed form equation representing the concentration versus time profile of an orally administered dose. However, it is quite rare that a medication regimen consists of one single dose. The single oral dose profile can be extended to a multiple oral dose regimen with the application of the superposition principle, as shown in [71].

Proposition 2.1.8 (The Superposition Principle)**.** *Consider $C_p(t)$, the current single-dose concentration versus time profile, and $C_p^{(n)}(t)$, the concentration versus time profile if the patient takes exactly n doses. Suppose that the dosing size is constant over time. Then the superposition principle is described by the following recursive formula*

$$C_p^{(n)}(t) = C_p^{(n-1)}(t) + C_p(t - t_n) \tag{2.1.8}$$

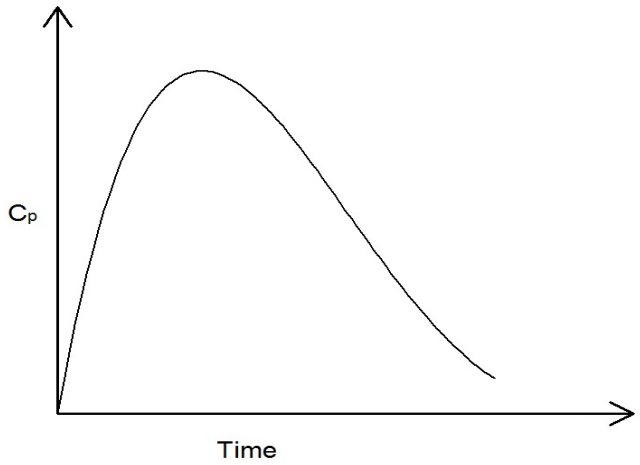

Figure 2.4: Single oral dose concentration versus time profile.

where t_n is the time of the n^{th} dose where $t_1 = 0$. By equation (2.1.8), the concentration-time curve after n doses $C_p^{(n)}(t)$ is then

$$C_p^{(n)}(t) = \sum_{i=1}^{n} C_p(t - t_i) \tag{2.1.9}$$

Remark 2.1.9. By definition, $C_p^{(1)}(t) = C_p(t)$.

From a pharmacokinetic point of view, the principle of superposition states that the total concentration of drug in the body is the sum of the remaining concentrations from each administered dose at that point in time when a measurement is made. It assumes that subsequent dosing events will not be impeded by the drug that is already circulating in the blood stream. In other words, each dose can be considered as an independent event, and the sum of all these dosing events provide the aggregate concentration of drug in circulation at that point in time. Now, in order to derive a closed form for the concentration-time profile of $C_p^{(n)}$, consider the most general case.

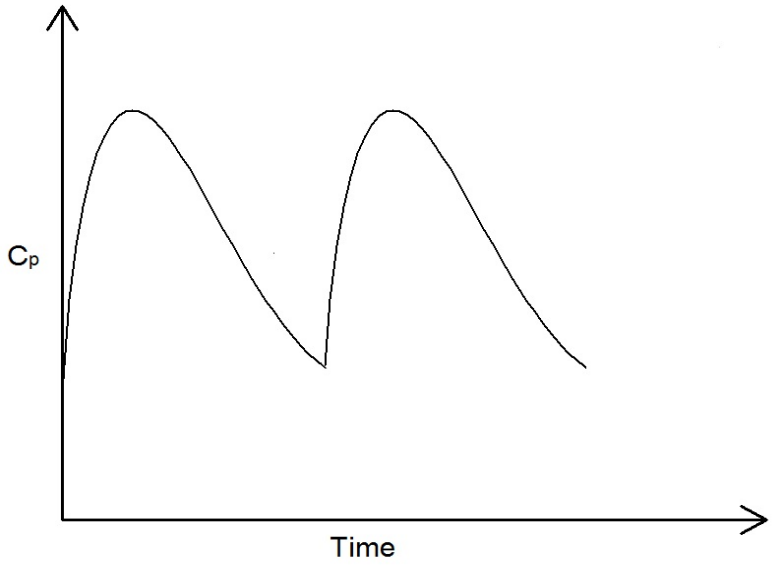

Figure 2.5: Multiple oral doses concentration versus time profile at equilibrium.

Proposition 2.1.10. *Suppose that we are trying to find $C_p^{(n)}(t)$, that is, the concentration in the plasma for the n^{th} dose at time t. Suppose that a patient receives a total of m doses at time $t_1, t_2, ..., t_m$ respectively, where $n \leq m$. Denote the dose amount at times t_i as D_i, for $i = 1, ..., m$. Define the dosing intervals as $\tau_i = t_{i+1} - t_i$ for $i = 1, ..., m$ and let $\tau_0 = t_1$. Then, by applying equation (2.1.9) to equation (2.1.7), we obtain the general multi-dose model for the n^{th} dose:*

$$C_p^{(n)}(t) = \sum_{i=1}^{n} \frac{SFD_i k_a}{(k_a - k)V} \left(\exp[-k(t - \sum_{j=0}^{i-1} \tau_j)] - \exp[-k_a(t - \sum_{j=0}^{i-1} \tau_j)] \right) \quad (2.1.10)$$

where $n \in \{1, ..., m-1\}$ is such that $t_n < t \leq t_{n+1}$ and $n = m$ if $t > t_m$.

A special, yet most common case is when $D_i = D$ for $i = 1, ..., n$, i.e that the

patient is prescribed a constant dose at each dose administration. This would be the case for our data; however, using a constant dose at each administration ignores the fact that in practice, it is possible that patients may miss one or more doses. To handle this noncompliance, we shall use a variant of the ordinal compliance index proposed by Wang *et al.* [70] as follows.

Definition 2.1.11. *Let Y_i denote the dose taken by a patient at the i^{th} dosing time. Then the* ***ordinal compliance index*** *can be written as*

$$Y_i = \begin{cases} 0 \text{ if the patient took 0 doses} \\ 1 \text{ if the patient took 1 dose} \\ 2 \text{ if the patient took 2 doses} \\ \vdots \qquad\qquad \vdots \\ m \text{ if the patient took } m \text{ doses} \end{cases} \tag{2.1.11}$$

In other words, the ordinal compliance index simply treats the missed doses as to have dose amount equal to 0 once "administered", and therefore for the missed dose only the past doses have residual effect on the plasma level. As shown in figure 2.6, the distribution of pills taken correctly over the dosage regimen is far from constant. This results in modifying the $C_p^{(n)}(t)$ taking Y_i into consideration as follows

$$C_p^{(n)}(t) = \sum_{i=1}^{n} Y_i C_p(t - t_i) \tag{2.1.12}$$

Although Y_n takes into consideration potential extra doses as well as missed doses, multiple dosing is not the main interest for this thesis, and therefore will have no direct impact on the creation of a time-dependent compliance variable discussed in the next section of this Chapter; multiple dosing will however be discussed in Chapter 4.

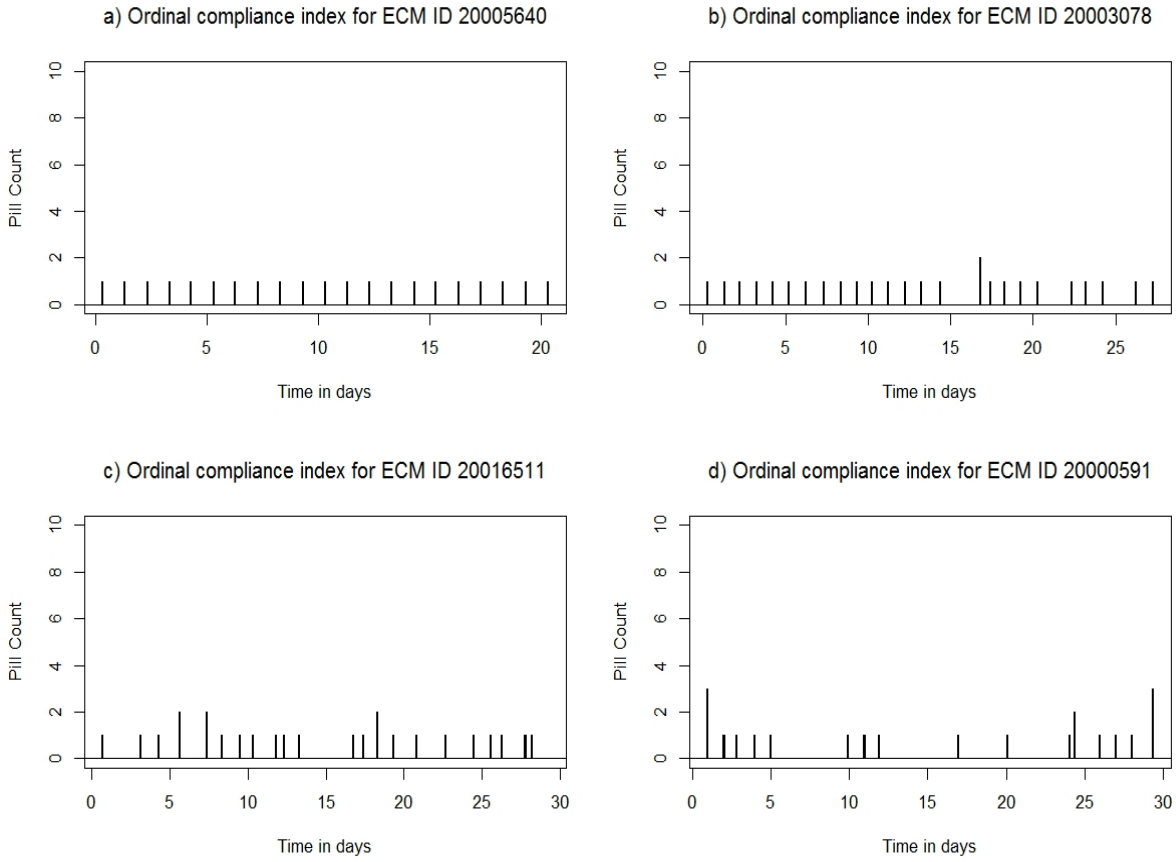

Figure 2.6: Example of 4 different compliance indices. a) Perfect (100%) compliance. b) 90% compliance. c) 60% compliance. d) 30% compliance.

2.2 Time-dependent compliance

As mentioned in the previous section, the concentration level of a drug should lie below the MTC and above the MEC; this is considered to be the steady state of the concentration level while following a treatment. Assuming that the MTC is never attained, the C_p must only stay above the MEC in order for the drug to have a real therapeutic effect. However, this does not violate medication compliance as defined in (1.2.2) since the drop below the therapeutic threshold arises from the plasma

concentration level and not directly from the number of doses not taken or taken incorrectly, even though these two aspects are closely related. Another reason why the plasma concentration can drop below the therapeutic threshold is because of patient's metabolism, i.e. how a patient's body reacts to the drug. It is clear that the prescription given by a physician should in theory be the correct dosage in order for the plasma concentration to be above the therapeutic level. Suppose that this is the case for the first year of a patient's treatment, and then suddenly the C_p drops below the MTC. It is possible that the metabolism suddenly reacts a different way to the treatment, rendering it non-therapeutic at that dosage. It is also possible that after a year, this patient stops taking his medication in accordance with his prescription. In the former case, it is clear that this is not a compliance issue, but rather a metabolic issue. These are two separate entities, and as mentioned before, there were no drug- nor subject-specific variables available to us in order to model these two phenomenons separately, leading to a problem of identifiability between the two. This is why we strictly focus on drops in C_p **caused by noncompliance with the treatment regimen**, which motivates a new approach to capture noncompliance by the C_p, **assuming no metabolic issue**.

Since the concentration level varies over time with respect to a patient's medication intake behaviour, it is only natural that compliance based on the C_p will vary with time as well. This will be the basis of the time-dependent compliance variable discussed in this section.

2.2.1 Defining a time-dependent compliance variable

Definition 2.2.1. *Assume that the plasma concentration level at steady state occurs after the first dose, i.e. that the first dose determines the therapeutic lower bound of the C_p. Suppose that each dose is prescribed to be taken once a day, that is at every 24 hours. Suppose that the timespan of the dosage regimen starts ($t = 0$) when the*

first dose is taken. Then the MEC plasma concentration level at steady state $\boldsymbol{C_{MEC}}$ *occurs just before the second dose, when* $t_{MEC} = 24$ *hours. Thus*

$$\begin{aligned} C_{MEC} &= \frac{SFDk_a}{(k_a - k)V}[\exp(-kt_{MEC}) - \exp(-k_a t_{MEC})] \\ &= \frac{SFDk_a}{(k_a - k)V}[\exp(-24k) - \exp(-24k_a)] \end{aligned} \tag{2.2.1}$$

A critical assumption is that the MEC is *fixed over the duration of the dosage regimen*. However, since our data consists of constant doses throughout the regimen, the assumption is not outrageous [19], and therefore definition (2.2.1) holds.

Definition 2.2.2. *Consider the therapeutic window allowed by IMC, i.e.* ± 2 *hours based on the first dose taken; this allows for a maximum of 28 hours between daily doses, hence* $t_{COMP} = 28$ *hours. A patient is considered* ***compliant*** *with respect to the* $C_p(t)$ *at time* t *for the* i^{th} *dose of the dosage regimen if*

$$C_p^{(i)}(t) \geq C_{COMP} \tag{2.2.2}$$

where $C_{COMP} = \frac{SFDk_a}{(k_a - k)V}[\exp(-t_{COMP}k) - \exp(-t_{COMP}k_a)]$ *and* $C_{COMP} < C_{MEC}$.

The use of C_{COMP} instead of C_{MEC} is motivated by the potential robustness of the threshold when taking biological variation such as metabolism into account. C_{COMP} would then better represent the variation in C_p with respect to actual patient compliance. The choice of t_{COMP} is arbitrary and any other suitable gap time in dose intake that would be considered noncompliant can be used to define C_{COMP}.

Now, notice that there are several subject-specific and drug-specific constants in the closed formula for the $C_p^{(n)}(t)$. As mentioned in the first Chapter, in order to produce concentration vs time profiling, we need to assign values to S, F, D, V, k and k_a with respect to a *standard individual* since we do not have any information on their actual values in the given data. Note that a discussion about this issue can be found the methodology section of Chapter 5.

Example 2.2.3. Suppose the subject- and drug-specific variables are such that

$$k_a = 0.4\ h^{-1}, \quad k = 0.125\ h^{-1}, \quad V = 0.4\ \text{L/kg}, \quad D = 500\ \text{mg}, \quad S = F = 1$$

Then the plasma concentration level vs time graphs associated with the ordinal compliance indices represented in figure 2.6 are

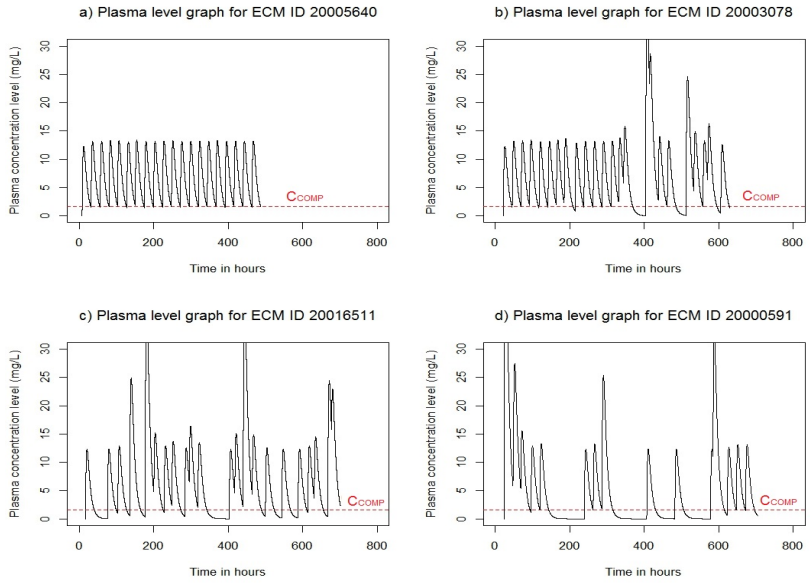

Figure 2.7: Example of 4 different plasma concentration vs time profiles. a) Shows perfect compliance, where $C_p(t) \geq C_{COMP}\ \forall\ t$ b) Shows mild non-compliance. c) Shows serious noncompliance. d) Shows severe noncompliance.

Remark 2.2.4. These parameters will be the ones chosen for the analysis in Chapter 5, i.e. they represent a *standard individual*. They are based on a case study from [19].

Thus, we can see from figure 2.7 that in the graph in b) for example, long periods of noncompliance occur roughly between hours 370 and 400, then again between 480 and 510, and then lastly roughly between 590 and 600. These are the time intervals

in which ECM ID 20003078 is noncompliant in the sense of equation (2.2.2). We can produce these types of intervals for each ECM ID, thus determining the time at which subjects are noncompliant with their treatment regimen. This leads us to

Definition 2.2.5. ***The Indicator of Plasma Concentration ($C_p(t)$) compliance****, denoted as $Z(t)$, during a dosage regimen at time t for the i^{th} dose can be represented as*

$$Z^{(i)}(t) = \begin{cases} 1 & \text{if } C_p^{(i)}(t) \geq C_{COMP} \\ 0 & \text{if } C_p^{(i)}(t) < C_{COMP} \end{cases} \tag{2.2.3}$$

Note that this definition is true regardless of the dose, and so $Z(t) = Z^{(i)}(t)$.

From this definition, we can plot the compliance variables for each case from figure 2.7, which can be found on figure 2.8. We can thus capture time-varying compliance by using a step function which equals 1 for the time intervals where subjects are $C_p(t)$ compliant and 0 during the time of noncompliance. This creates a "gap time" compliance indicator at time i as $Z^{(i)}(t) = \mathbf{1}[t - t_i \leq t_{COMP}]$.

It is important to realize that although ECM ID 20005640 has perfect compliance with respect to definition (1.2.2), there might be some small (almost tiny) moments of noncompliance with respect to definition (2.2.2) depending on the choice of C_{COMP} close to C_{MEC} as shown in figure 2.8a. That is, although the patient administers the doses in the allowed therapeutic window, this window can potentially be so large that the drug effect can wear off before the next dose is taken. This minor variation in the plasma concentration could also potentially, however loosely, be translated as a metabolic effect; the body may handle the drug slightly differently one day to next, resulting in a drop of the C_p below the MEC. This explanation is of course farfetched in our case, since we are not actually modelling against metabolic issues - but it can be incorporated within the model; see [54] for more information.

Now for example, suppose that the second dose is administered at $t_{COMP} = 28$ hours and 1 minute after the first dose is taken. Then, according to our model,

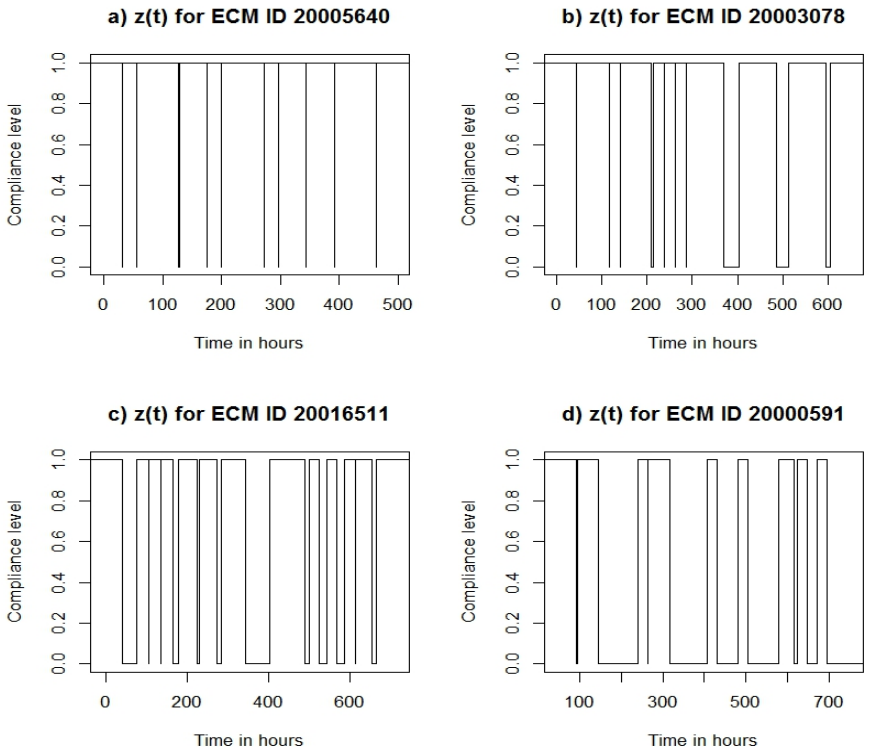

Figure 2.8: Example of 4 different compliance indicators a) Shows quasi-perfect (99%) compliance. b) Shows mild noncompliance. c) Shows above average noncompliance. d) Shows severe noncompliance.

the second dose will be taken when the C_p is below the C_{COMP}, which produces roughly 1 minute of noncompliance with respect to definition (2.2.5). However, after the second dose, the Superposition Principle complicates the calculation of the next interval where $C_p^{(n)} < C_{COMP}$, $n > 2$, due to equation (2.1.12). This results in capturing potential noncompliance within compliance by pill counting when utilizing the compliance indicator with respect to the $C_p(t)$.

Finding critical clinical times from past treatment intake based on the MEC

Recall the one-dose model from equation (2.1.7) and the plasma concentration at time $t > t_{(n)}$ at the n^{th} dose from equation (2.1.12). Define t_{MEC} as the time at which the $C_p(t)$ falls below C_{MEC}, the minimum effective concentration of the medication. The derivative $C_p'(t_{MEC}) < 0$. Let $Q = \frac{SFD}{V_d}$, $k_r = \frac{k_a}{k_a - k} > 1$ and $\delta_a = k_a - k$. As Q and k_r are fixed with respect to time t we have

$$C_{MEC} = Qk_r \left(\exp(-kt_{MEC}) - \exp(-k_a t_{MEC})\right) \tag{2.2.4}$$

so that we can define

$$C^* = \frac{C_{MEC}}{Qk_r} = \left(\exp(-kt_{MEC}) - \exp(-k_a t_{MEC})\right) \tag{2.2.5}$$

While $C_p^{(n)}(t)$ is a non-monotone function, its local minima occurs at t_i^- and we can derive the next maximum after $t_{(n)}$ using basic calculus.

$$\begin{aligned} & \tfrac{d}{dt} C_n(t) && \overset{set}{=} 0 && (2.2.6) \\ \iff \quad & k_a \sum_{i=1}^{n} \exp\left(-k_a(t - t_i)\right) && = k \sum_{i=1}^{n} \exp\left(-k(t - t_i)\right) \\ \iff \quad & \frac{k_a \sum_{i=1}^{n} \exp(k_a t_i)}{k \sum_{i=1}^{n} \exp(k t_i)} && = \exp\left((k_a - k)t\right) \\ \iff \quad & t_{\max} && = \frac{\log\left(\frac{k_a \sum_{i=1}^{n} \exp(k_a t_i)}{k \sum_{i=1}^{n} \exp(k t_i)}\right)}{k_a - k} && (2.2.7) \end{aligned}$$

Thus, from the time of last dose $t_{(n)}$, the next maximum concentration occurs at $t = t_{\max} - t_{(n)}$.

We can use $t_{\max}$ to find the next time at which $C_p^{(n)}(t) < C_{MEC}$ if no new dose is taken. For this, we need to solve

$$C^* = \sum_{i=1}^{n} \left(\exp\left(-k(t - t_i)\right) - \exp\left(-k_a(t - t_i)\right)\right) \tag{2.2.8}$$

for t subject to $t > t_{\max}$, ensuring $\frac{d}{dt}C_p^{(n)}(t) < 0$. There is no explicit solution to equation (2.2.8) but one can set up a fixed point iteration by using

$$C^* = e^{-kt}\left(\sum_{i=1}^{n} e^{kt_i} - e^{-\delta_a t}\sum_{i=1}^{n} e^{k_a t_i}\right) \tag{2.2.9}$$

$$\iff t = -\frac{1}{k}\log\left(\frac{C^*}{\sum_{i=1}^{n} e^{kt_i} - e^{-\delta_a t}\sum_{i=1}^{n} e^{k_a t_i}}\right) \tag{2.2.10}$$

Starting the iteration at $t_{(n)} + \tau$, where τ is the assigned time interval between two doses, leads to the next threshold time t^*, which, from the time of the n^{th} dose is $t^* - t_{(n)}$.

Note that it is not guaranteed that the fixed point iteration algorithm converges (or any fixed point iteration algorithm for that matter). It does however converge for the proposed pharmacokinetic values in example (2.2.3).

We now have all the pharmacokinetic background and the information about the time-dependent compliance that we will use in our simulation study analyses in Chapter 5. Before the analyses however, it is important to review the theory behind the methods. In the next Chapter, we will review basic survival analysis theory, including the Kaplan-Meier estimate for the survival function and Cox's proportional hazards model.

Chapter 3

Survival analysis: the theory

In order to model a time-to-noncompliance event as well as a proportional hazards model comparing different compliance variables, some survival analysis framework is required. Survival analysis at its core examines and models the time it takes for events to occur. The prototypical such event is death, from which the name "survival analysis" and much of its terminology derives, but in reality the ambit of application of survival analysis is much broader (eg. predicting cross-selling in the insurance industry [31]). Survival analysis focuses on the distribution of survival times, or in general, event times (note that these two terms will be used interchangeably throughout this thesis). Although there are well known methods for estimating unconditional survival distributions, most interesting survival modeling examines the relationship between survival and one or more predictors (or covariates). In this Chapter, we will review the core theory of survival analysis, including basic definitions, how to model the survival function when dealing with covariates, the construction of likelihood function, the Kaplan-Meier estimator and Cox's proportional hazards model.

The nature of survival data

Survival data have two important special characteristics:

1. The survival times are non-negative, and consequently they are usually positively skewed. This makes the naive analysis of untransformed event times unpromising, since the majority of basic methods have strict assumptions, commonly requiring the data to be normally distributed.

2. Typically, some subjects (i.e., units of observation) have *censored* survival times. This means that the event times of these subjects are not fully observed, for example, because the event of interest does not take place for these subjects before the termination of the study. Failure to take censoring into account can produce serious bias in estimates of the distribution of survival time and related quantities.

The cause of censoring must be independent of the event of interest (i.e. must be independent of the future value of the *hazard* for the individual) in order to use standard methods of analysis. The most common type of censoring is right censoring, where the event of interest has yet to occur at the end of follow-up (either predetermined or random) for that subject. There also exist several other types of censoring. Among them we can find left censoring, where the event of interest occurs before the time of entry in the study of that subject, and interval censoring, which occurs when subjects have periodic follow-ups and the event time is only known to fall within an interval $[L, R]$. Since the latter censoring types do not appear in the data however, they will have no contribution when constructing the likelihood function when comes time to estimate parameters **in the remainder of this thesis**.

Definition 3.0.6. *Let T be the continuous random variable representing survival time with cumulative distribution function $F(t) = P(T \leq t)$ and probability density function (pdf) $f(t) = \frac{dF(t)}{dt}$. Let y_i be the i^{th} time-to-event from onset from pdf f with survival function S, where the term onset is defined as the start of the dosage regimen.*

Let c_i be the i^{th} time-to-censoring from onset, independent of y_i. Let $t_i = \min(y_i, c_i)$ and define

$$\delta_i = \mathbf{1}[y_i \leq c_i] = \begin{cases} 1 & \text{if } y_i \leq c_i \\ 0 & \text{if } y_i > c_i \end{cases}$$

Then the sample consisting of ***right-censored data*** *takes the form of the pairs (t_i,δ_i) for $i = 1, ..., n$.*

Definition 3.0.7. ***Truncation*** *occurs when the incomplete nature of the observation is due to a systematic selection process inherent to the study design.*

Remark 3.0.8 (Censoring vs. Truncation)**.** Although both are partial observations, truncation is a condition for a subject to be in the sample, while censoring is a subject-specific incomplete observation.

There are two types of truncation. The first and most common one is left truncation (also known as staggered entry), which occurs when subjects enter the study at a time which is not necessarily the origin for the event of interest and are followed from this delayed-entry time until either the event occurs or the subject is censored. In our case, there is no explicit left-truncation. However, the framework of left truncation is necessary when representing a time-dependent covariate in R. The other is right truncation, which occurs when only individuals who have experienced the event of interest are observable. The main impact of truncation on the analysis is that the investigator must use a conditional distribution in constructing the likelihood or employ a statistical method which adjusts for the effect of truncation. There is no right truncation in our data.

3.1 Survival and hazard functions

Definition 3.1.1. *Let T be the continuous random variable representing survival time with cumulative distribution function $F(t) = P(T \leq t)$ and probability density*

function (pdf) $f(t) = \frac{dF(t)}{dt}$. *The* ***survival function*** $S(t)$ *is the probability of an individual surviving beyond time* t*, defined as*

$$S(t) = P(T > t) = 1 - P(T \leq t) = 1 - F(t) = \int_t^\infty f(x)dx \tag{3.1.1}$$

Definition 3.1.2. *The function which assesses the instantaneous risk of demise at time* t *conditional on survival up to that time is the* ***hazard function*** $h(t)$*, also called* ***hazard rate****, defined as*

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T > t)}{\Delta t} \tag{3.1.2}$$

With these two definitions, we can observe the relationship between survival and hazard.

Proposition 3.1.3. *Let* $f(t)$ *be the pdf of survival time* T*. Let* $S(t)$ *be the survival function of* T *and suppose that* T *is continuous. Then*

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt} \log S(t) \tag{3.1.3}$$

Solving equation (3.1.3) with respect to $S(t)$ yields

$$S(t) = e^{-H(t)} \tag{3.1.4}$$

where

$$H(t) = \int_0^t h(x)dx$$

is known as the **cumulative hazard function**.

3.2 Modeling the survival function with covariates

To model the relationship between covariates and event times, we present here two of the most popular models: Accelerated Failure Time model and Cox's proportional hazards model. Consider an ordinary regression model for the log of the survival time T of the form

$$Y = \log T = \mu + \boldsymbol{\beta}'\mathbf{z} + \sigma\epsilon \tag{3.2.1}$$

where $\boldsymbol{\beta}' = (\beta_1, \beta_2, ..., \beta_p)$ is a vector of regression coefficients, $\mathbf{z} = (z_1, z_2, ..., z_p)'$ is a vector of covariates and where the error term ϵ has a suitable distribution, e.g. extreme value, generalized extreme value, normal or logistic. This model is called the Accelerated Failure Time (AFT) since it has an accelerated life interpretation. In this formulation, we view $\mu + \sigma\epsilon$ as a standard or reference distribution that applies when $\mathbf{z} = \mathbf{0}$. This results in the baseline survival function being $S_0(t) = \exp(\mu + \sigma\epsilon)$. If we consider the effect of the covariates $\mathbf{z}$ to be nonzero, we obtain

$$P(T > t|\mathbf{z}) = S_0(te^{-\boldsymbol{\beta}'\mathbf{z}})$$

In words, the probability that a subject with covariates $\mathbf{z}$ will survive up until time t is the same as the probability that a reference subject will survive up until time $te^{-\beta'\mathbf{z}}$. This may be interpreted as time passing more rapidly (or more slowly) by a factor of $\exp(\boldsymbol{\beta}'\mathbf{z})$.

Although this first approach provides a useful framework, for a number of cases in real applications its use is restricted by the error distribution that one can assume. An alternative approach to modeling survival data is to assume, without any parametric assumptions about the form of the distribution, that the effect of the covariates is to increase or decrease the *hazard* by a proportionate amount at all durations. The multiplicative hazard models thus states that the conditional hazard rate of an individual with covariates $\mathbf{z}$ is a product of a baseline hazard rate $h_0(t)$ and a non-negative function of the covariates $\mathcal{C}(\boldsymbol{\beta}'\mathbf{z})$, that is

$$h(t|\mathbf{z}) = h_0(t)\mathcal{C}(\boldsymbol{\beta}'\mathbf{z})$$

where $\mathcal{C}$ is called the link function and $h_0(t)$ is the hazard function at baseline. The most popular link function is $\mathcal{C}(\boldsymbol{\beta}'\mathbf{z}) = e^{\boldsymbol{\beta}'\mathbf{z}}$, known as the Cox model. We will return

to the Cox model in a further section; for now, let us consider how to construct the likelihood function when dealing with our survival data.

3.3 Construction of the likelihood function

Suppose we have a random sample of individuals of size n from a population with survival time T and let $f(t)$ and $S(t)$ be the density and survival functions of T, respectively. Due to right censoring, the opportunity of observing the true event times of individuals does not always arise. Let C be the right-censoring random variable denoting censoring times $(C_1, ..., C_n)$. Thus, the observed data are the minimum of the survival time and censoring time for each subject in the sample and the indication whether or not the subject is censored, as described by (3.0.6). Each type of observation contributes a part of the desired likelihood function. In our case, only observed events (survival times) and right-censored times contribute to the likelihood function.

Let $\mathcal{D}$ be the set of subjects who have failed and $\mathcal{C}$ be the set of right-censored subjects. Consider first the case where the true event times are observed. Let δ_i be defined as in (3.0.6), and, in an abuse of notation, let $P(T = t_i)$ denote density for simplicity. Then the probability of observing an event at time t_i when the i^{th} subject fails is

$$\begin{aligned} P(T_i = t_i, \delta_i = 1) &= P(T_i = t_i, T_i \leq C_i) \\ &= P(T_i = t_i | T_i \leq C_i) P(T_i \leq C_i) \\ &= P(t_i \leq C_i | T_i = t_i) P(T_i = t_i) \\ &= P(C_i \geq t_i) f_T(t_i) \\ &= S_C(t_i) f_T(t_i) \end{aligned} \tag{3.3.1}$$

where $S_C(\cdot)$ denotes the survival function of the censoring times and $f_T(\cdot)$ denotes the density function of survival time.

Next, consider the case where the observation is censored.

$$P(T_i = c_i, \delta_i = 0) = S_T(c_i) g_C(c_i) \tag{3.3.2}$$

by the same process as the first case, where $g_C(\cdot)$ denoted the density function of the censoring times. Therefore, combining (3.3.1) and (3.3.2) while conditioning on Y_i, the left truncations, we obtain that the likelihood function can be written as

$$\begin{aligned}
\mathcal{L} &= \prod_{i \in \mathcal{D}} \frac{S_C(t_i) f_T(t_i)}{S_T(y_i) g_C(y_i)} \prod_{j \in \mathcal{C}} \frac{S_T(t_j) g_C(t_j)}{S_T(y_j) g_C(y_j)} \\
&\propto \prod_{i \in \mathcal{D}} \frac{f_T(t_i)}{S_T(y_i)} \prod_{j \in \mathcal{C}} \frac{S_T(t_j)}{S_T(y_j)} \quad \text{since censoring is non-informative} \\
&= \prod_{i=1}^{n} \left(\frac{f_T(t_i)}{S_T(y_i)} \right)^{\delta_i} \left(\frac{S_T(t_i)}{S_T(y_i)} \right)^{1-\delta_i} \\
&= \prod_{i=1}^{n} \frac{1}{S_T(y_i)} [f_T(t_i)]^{\delta_i} [S_T(t_i)]^{1-\delta_i} \\
&= \prod_{i=1}^{n} e^{H(y_i)} [h(t_i) e^{-H(t_i)}]^{\delta_i} [e^{-H(t_i)}]^{1-\delta_i} \quad \text{by equation (3.1.3)} \\
&= \prod_{i=1}^{n} e^{H(y_i) - H(t_i)} [h(t_i)]^{\delta_i}
\end{aligned} \tag{3.3.3}$$

where $h(\cdot)$ and $H(\cdot)$ are the hazard and cumulative hazard functions of T. In reality, when a parametric distribution is assumed, $f(\cdot)$, $S(\cdot)$, $h(\cdot)$, and $H(\cdot)$ should be written as $f(\cdot|\boldsymbol{\theta})$, $S(\cdot|\boldsymbol{\theta})$, $h(\cdot|\boldsymbol{\theta})$, and $H(\cdot|\boldsymbol{\theta})$, as we are using the likelihood function in order to estimate the underlying parameters $\boldsymbol{\theta} = (\theta_1, ..., \theta_p)$ of the functions in order to obtain their maximum likelihood estimators (MLE). In order to obtain the Kaplan-Meier (KM) estimator, one must maximize $\mathcal{L}$ with respect to $S(\cdot)$ such that it only assigns positive probability to the observed failure times *without assuming a parametric distribution*, as discussed in the next section.

3.4 Kaplan-Meier analysis

3.4.1 The Kaplan-Meier estimator

It is easy to see that if the data is not censored, the obvious estimate is the empirical survival function

$$\hat{S}(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}[t_i > t] \tag{3.4.1}$$

where t_i represents distinct event times. The estimator is simply the proportion alive at time t. Due to the nature of survival data however, the empirical estimate cannot be used when censoring is present. Kaplan and Meier [48] extended this nonparametric estimate to censored data. It is extremely popular as it requires only very weak assumptions and yet utilizes the information content of both fully observed and censored data. The following is a semi-heuristic walkthrough on how to obtain the KM estimator $\hat{S}_{KM}(t)$.

For simplicity, consider the case without truncation (taking truncation into account is not much harder). Let $t_{(1)} < t_{(2)} < ... < t_{(m)}$ denote the distinct ordered times of failure, not counting censoring times. Let d_i be the number of failures at time $t_{(i)}$, and let n_i be the number of subjects who have not yet failed just before $t_{(i)}$, which can be considered as the number exposed to risk at time $t_{(i)}$. Consider the contribution to the likelihood function of cases that fail or are censored at time t. As seen in the previous section, if a subject is censored at time t its contribution to $\mathcal{L}$ is $S(t)$. In order to maximize the likelihood we would like to make this as large as possible. Because a survival function must be non-increasing, the best we can do is keep it constant at t. In other words, the estimated survival function does not vary at censoring times. Next, if a subject fails at time t then this is one of the distinct times of failure, say $t_{(i)}$. In this case, in order to maximize the likelihood, the survival function just before time $t_{(i)}$ needs to be as large as possible. The largest it can be is

the value at the previous time of failure or 1, whichever is less. There is also a need to make the survival at $t_{(i)}$ itself as small as possible, meaning discontinuity at time $t_{(i)}$.

Let c_i denote the number of cases censored between $t_{(i)}$ and $t_{(i+1)}$. Then the likelihood function takes the form

$$\mathcal{L} = \prod_{i=1}^{m} [S(t_{(i-1)}) - S(t_{(i)})]^{d_i} S(t_{(i)})^{c_i} \tag{3.4.2}$$

where the product is over the m distinct times of failure, and letting $t_{(0)} = 0$ with $S(t_{(0)}) = 1$. We are now looking to estimate m parameters representing the values of the survival function at the m failure times. Let $\pi_i = \frac{S(t_{(i)})}{S(t_{(i-1)})}$ be the conditional probability of survival from $S(t_{(i-1)})$ to $S(t_{(i)})$. Thus we can write

$$S(t_{(i)}) = \pi_1 \pi_2 \cdots \pi_i$$

since to survive up to time $t_{(i)}$, you must first survive up to time $t_{(1)}$, and then survive from time $t_{(1)}$ to time $t_{(2)}$ given that you have already survived up to time $t_{(1)}$, and so on. The likelihood then becomes

$$\mathcal{L} = \prod_{i=1}^{m} (1 - \pi_i)^{d_i} \pi_i^{c_i} (\pi_1 \pi_2 \cdots \pi_{i-1})^{d_i + c_i} \tag{3.4.3}$$

Note that all cases who fail at time $t_{(i)}$ or are censored between $t_{(i)}$ and $t_{(i+1)}$ contribute a term π_j to each of the previous failure times from $t_{(1)}$ to $t_{(i-1)}$. Next, we can define $n_i = \sum_{j \geq i} (d_j + c_j)$ (recall that n_i is the total number of subjects exposed to risk at time $t_{(i)}$). By then collecting the terms on each π_i, we can rewrite equation (3.4.3) as

$$\mathcal{L} = \prod_{i=1}^{m} (1 - \pi_i)^{d_i} \pi_i^{n_i - d_i} \tag{3.4.4}$$

which is a binomial likelihood function. The MLE of π_i for the time of failure $t_{(i)}$ is then

$$\hat{\pi}_i = \frac{n_i - d_i}{n_i} = 1 - \frac{d_i}{n_i} \tag{3.4.5}$$

We then obtain the KM estimator by multiplying all conditional probabilities $\hat{\pi}_j$'s so that

$$\begin{aligned}
\hat{S}_{KM}(t) &= \begin{cases} 1 \text{ if } t < t_{(1)} \\ \prod_{i=1}^{m} \hat{\pi}_i \text{ if } t_{(1)} \leq t \leq t_{(m)} \end{cases} \\
&= \begin{cases} 1 \text{ if } t < t_{(1)} \\ \prod_{i=1}^{m} \left(1 - \frac{d_i}{n_i}\right) \text{ if } t_{(1)} \leq t \leq t_{(m)} \end{cases} \\
&= \begin{cases} 1 \text{ if } t < t_{(1)} \\ \prod_{i:t_{(i)} \leq t} \left(1 - \frac{d_i}{n_i}\right) \text{ if } t_{(1)} \leq t \leq t_{(m)} \end{cases}
\end{aligned} \tag{3.4.6}$$

The KM estimate results in a step function with discontinuities at the observed failure times. It can be shown that $\hat{S}_{KM}(t)$ coincides with the empirical survival function in (3.4.1) if there is no censoring.

Remark 3.4.1. $\hat{S}_{KM}(t)$ is well-defined for $t < t_{max}$, where t_{max} is the largest observed time. If t_{max} corresponds to a failure time, then the $\hat{S}_{KM}(t) = 0$ for $t \geq t_{max}$. If it corresponds to a censoring time, then the value of $\hat{S}_{KM}(t)$ beyond t_{max} is undetermined due to the fact that there is no information on whether the last subject would have failed if the observation had not been censored. The survival curve ends at t_{max} with $\hat{S}_{KM}(t_{eventmax})$, where $t_{eventmax}$ is the largest observed event time.

3.4.2 Greenwood's formula

From the likelihood function obtained in the previous section, it follows that the large sample variance of $\hat{\pi}_i$ conditional on the data n_i and d_i is given by the usual binomial

variance formula

$$\mathrm{Var}(\hat{\pi}_i) = \frac{\pi_i(1-\pi_i)}{n_i}$$

Remark 3.4.2. Note that $\mathrm{Cov}(\hat{\pi}_i, \hat{\pi}_j) = 0$ for $i \neq j$, so the covariances of the contributions from different failure times are all zero. This can be verified by taking logs and then first and second derivatives of the log-likelihood function.

In order to obtain the large sample variance of $\hat{S}_{KM}(t)$, one has to apply the *delta method* (see [9]) twice. First we take the log of $\hat{S}_{KM}(t_{(i)})$ so that instead of the variance of a product we can find the variance of a sum, working with

$$\log \hat{S}_{KM}(t_{(i)}) = \sum_{j=1}^{i} \log \hat{\pi}_j$$

We then apply the delta method to find the variance of $\log(\hat{\pi}_i)$, which results in

$$\mathrm{Var}(\log \hat{\pi}_i) = \left(\frac{1}{\hat{\pi}_i}\right) \mathrm{Var}(\hat{\pi}_i) = \frac{1-\pi_i}{n_i \pi_i}$$

Now, because $\log \hat{S}_{KM}(t_{(i)})$ is a sum and the covariances of the π_i's (and hence of the $\log \pi_i$'s) are zero, we obtain

$$\mathrm{Var}(\log \hat{S}_{KM}(t_{(i)})) = \sum_{j=1}^{i} \frac{1-\pi_j}{n_j \pi_j} = \sum_{j=1}^{i} \frac{d_j}{n_j(n_j - d_j)}$$

The second application of the delta method consists of obtaining the variance of the KM survival function estimate from the variance of its log, resulting in

$$\begin{aligned} \mathrm{Var}(\hat{S}_{KM}(t)) &= [\hat{S}_{KM}(t)]^2 \sum_{i:t_{(i)} \leq t} \frac{1-\hat{\pi}_i}{n_i \hat{\pi}_i} \\ &= [\hat{S}_{KM}(t)]^2 \sum_{i:t_{(i)} \leq t} \frac{d_i}{n_i(n_i - d_i)} \end{aligned} \tag{3.4.7}$$

which is known as Greenwood's estimator of the variance of the survival function (see [57]).

3.4.3 A note on estimating and smoothing the hazard rate

Survival data can be generally described and modelling in terms of either *survival* or *hazard*. Although the work in this thesis is concentrated on a survival approach, which reflects the cumulative non-occurrence of failures, the hazard rate can be especially useful because it can provide insight into the conditional failure rates and can provide a vehicle for specifying a survival model.

Unlike $S(t)$, there is no simple way to estimate $h(t)$. Instead, we estimate the cumulative hazard function $H(t)$, which is used as an intermediary measure for estimating $h(t)$ and as a diagnostic tool in assessing model validity. Plotting the raw hazard rate results in a curve that is ragged, indicating high variability. However, since the **true** hazard rate is typically assumed to be a smooth function, smoothing the failure rates can provide an aesthetically improved estimate. A smoothing procedure, when applied properly, also improves the statistical performance of the resulting hazard rate estimator [53]. For continuously observed data, hazard estimators are ordinarily obtained by applying a kernel smoother to the increments of the Nelson-Aalen estimator [34] for the cumulative hazard function $H(t)$. For more information on the smoothing process, see [68].

3.4.4 Addressing the effect of covariates

With a nonparametric estimate such as $\hat{S}_{KM}(t)$, in order to take covariates into account, one must consider comparing subpopulations. In other words, categorical variables can define several subpopulations (e.g. gender) on which to test differences in the behaviour of survival. Various tests have been proposed for testing for differences in survival between categorical covariates. Among these tests, let us consider the case where the variables have only two categories for simplicity. The hypothesis test in this case can be written as

$$H_0 : S_1(t) = S_2(t) \quad vs \quad H_1 : S_1(t) \neq S_2(t) \tag{3.4.8}$$

Let $d_{k,i}$ be the number of failures in group k at ordered time $t_{(i)}$, where the ordering is over all categories. Let $d_i = \sum_{k=1}^{K} d_{k,i}$ be the total number of failures at time $t_{(i)}$, where $K = 2$ when there are two categories. Let $n_{k,i}$ be the number of subjects of group k at risk an instant before $t_{(i)}$ and n_i. If H_0 is true, then the expected number of deaths in group k at time $t_{(i)}$ is

$$\hat{e}_{k,i} = \frac{n_{k,i} d_i}{n_i} \tag{3.4.9}$$

The $n_{k,i}$ term is the number at risk in category k. The ratio $\frac{d_i}{n_i}$ is the overall proportion in both populations failing at time $t_{(i)}$. The variance in $d_{k,i}$ is given by the variance of the hypergeometric distribution such as

$$\hat{v}_{1,i} = \hat{v}_{2,i} = \frac{n_{1,i} n_{2,i} d_i (n_i - d_i)}{n_i^2 (n_i - 1)} \tag{3.4.10}$$

The test statistic for testing (3.4.8) is then

$$q = \frac{[\sum_{i=1}^{m} w_i (d_{1,i} - \hat{e}_{1,i})]^2}{\sum_{i=1}^{m} w_i^2 \hat{v}_{1,i}} \tag{3.4.11}$$

where m is the number of distinct failure times and w_i are pre-determined weights. The most popular weights are $w_i = 1$ for the log-rank test and $w_i = n_i$ for the Wilcoxon test. If H_0 is true, then $q \sim \chi_1^2$ asymptotically. Note that the statistic is based on one subpopulation's sample moments only, as the other is deterministic conditional on the first. Although it is omitted here, note that if there are $K \geq 3$ subpopulations of interest, similar tests can be constructed by generalizing notation to use matrix algebra (for more information, see [57]).

The existing tests described above are very useful in assessing whether a covariate affects survival. However, they do not allow us to say *how* survival is affected. Ideally, we would like to be able to say how much more at risk one group is than another.

Furthermore, incorporating non-categorical covariates would also be of interest, since many continuous variables (such as age) are an important factor in modeling survival. Although this can potentially be done using an *ad hoc* approach by categorizing the covariates arbitrarily, a more satisfying approach is to do some semiparametric modeling to investigate the functional relationship between the covariates and survival. This can be done using Cox's proportional hazards model, discussed in the next section.

3.5 Cox's proportional hazards models

3.5.1 Dealing with fixed covariates

Consider the general problem where we have a vector $\mathbf{z} = (z_1, ..., z_p)'$ of covariates fixed in time. Recall from section 3.2 that we wish to model survival data without making any assumptions about the about the form of the distribution (and thus, about the baseline hazard $h_0(t)$) with Cox's model

$$h(t|\mathbf{z}) = h_0(t)e^{\boldsymbol{\beta}'\mathbf{z}}$$

The goal is to estimate $\boldsymbol{\beta}$ in order to determine how the covariates affect survival. Since we will only parametrically model the relationship between the covariates and the time to event, Cox's model is considered *semiparametric*. Let $t_{(1)} < ... < t_{(m)}$ denote the observed distinct ordered times of failure, as before. Let R_i denote the risk set at time $t_{(i)}$ defined as the set of indices of the subjects who have not yet failed just before $t_{(i)}$. Let us first consider the case where one and only one subject, say k_i, fails at time $t_{(i)}$, i.e. that there are no ties in the observation times. Then the conditional probability that this particular subject fails at time $t_{(i)}$ given the risk set R_i and that there are no ties is given by

$$\begin{aligned} p(\mathbf{z}_{k_i}) &= \frac{h(t_{(i)}|\mathbf{z}_{k_i})}{\sum_{k\in R_i} h(t_{(i)}|\mathbf{z}_k)} \\ &= \frac{h_0(t_{(i)})e^{\boldsymbol{\beta}'\mathbf{z}_{k_i}}}{\sum_{k\in R_i} h_0(t_{(i)})e^{\boldsymbol{\beta}'\mathbf{z}_k}} \\ &= \frac{e^{\boldsymbol{\beta}'\mathbf{z}_{k_i}}}{\sum_{k\in R_i} e^{\boldsymbol{\beta}'\mathbf{z}_k}} \end{aligned} \tag{3.5.1}$$

and so $p(\mathbf{z}_{k_i})$ does not depend on the baseline hazard. Next, by multiplying these conditional probabilities together over all distinct failure times we obtain

$$\begin{aligned} \mathcal{L}(\boldsymbol{\beta}) &= \prod_{i=1}^{m} p(\mathbf{z}_{k_i}) \\ &= \prod_{i=1}^{m} \frac{e^{\boldsymbol{\beta}'\mathbf{z}_{k_i}}}{\sum_{k\in R_i} e^{\boldsymbol{\beta}'\mathbf{z}_k}} \end{aligned} \tag{3.5.2}$$

which is known as the Cox partial likelihood. From this, taking the log and then maximizing with respect to $\boldsymbol{\beta}$ one can find the MLE $\hat{\boldsymbol{\beta}}$. This partial likelihood can also be derived as a profile likelihood from the full censored (omitting truncation) data likelihood in (3.3.3) by fixing the $\boldsymbol{\beta}$'s and maximizing the likelihood as a function of $h_0(t_{(i)})$ only. It can be shown that the profile MLE for $h_0(t_{(i)})$ is given by

$$\hat{h}_0(t_{(i)}) = \frac{1}{\sum_{k\in R_i} e^{\boldsymbol{\beta}'\mathbf{z}_k}} \tag{3.5.3}$$

which then gives Breslow's estimator of the baseline cumulative hazard

$$\hat{H}_0(t) = \sum_{i:t_{(i)}\le t} \hat{h}_0(t_{(i)}) = \sum_{i:t_{(i)}\le t} \frac{1}{\sum_{k\in R_i} e^{\boldsymbol{\beta}'\mathbf{z}_k}}$$

Adjusting for ties

The development so far has assumed that only one failure occurs at each distinct time $t_{(i)}$. In practice, we often observe several failures at $t_{(i)}$. Let d_i be the number of events observed at time $t_{(i)}$ and $\mathcal{D}_i$ be the set of all individual who experience the event at $t_{(i)}$. Suppose $\mathcal{S}_i$ is the sum of the vectors $\mathbf{z}_i$ over all individuals who fail at time $t_{(i)}$, that is

$$\mathcal{S}_i = \sum_{j \in \mathcal{D}_i} \mathbf{z}_j$$

and let $\mathcal{R}_i = \mathcal{R}(t_{(i)})$ be the set of all individuals at risk right before $t_{(i)}$. Then the partial likelihood in (3.5.1) can be written as

$$\mathcal{L}_B(\boldsymbol{\beta}) = \prod_{i=1}^{m} \frac{e^{\boldsymbol{\beta}' \mathcal{S}_i}}{[\sum_{k \in \mathcal{R}_i} e^{\boldsymbol{\beta}' \mathbf{z}_k}]^{d_i}}$$

which was proposed by Breslow. There also exists an adjustment from Efron, which uses the average risk of the subjects who fail instead of adding the risks over the complete risk set in $\mathcal{L}_B$, which is more accurate than Breslow's adjustment when there are more than an insignificant number of ties, unless d_i is very large relative to n_i. Cox also proposed an adjustment for ties considering the failure time distribution as discrete, treating the hazard function as a conditional probability.

Estimating the survival function

Once we have found the MLE $\hat{\boldsymbol{\beta}}$ of the regression coefficients, we now consider how to estimate survival function. Using the same framework as above, let $t_{(1)} < ... < t_{(m)}$ be the distinct failure times and d_i be the number of failures at time $t_{(i)}$. Then the estimator of $H_0(t)$ is given by

$$\hat{H}_0(t) = \sum_{i: t_{(i)} \leq t} \frac{d_i}{\sum_{k \in \mathcal{R}_i} e^{\hat{\boldsymbol{\beta}}' \mathbf{z}_k}} \tag{3.5.4}$$

where $\hat{\boldsymbol{\beta}}$ is the partial MLE of $\boldsymbol{\beta}$. Note that this is simply equation (3.5.3) while adjusting for ties. Thus, from equation (3.1.4), we obtain that the estimator of the baseline survival function is

$$\hat{S}_0(t) = e^{-\hat{H}_0(t)} \tag{3.5.5}$$

which then yields the survival function estimate

$$\hat{S}(t|\mathbf{Z} = \mathbf{z}_0) = [\hat{S}_0(t)]^{e^{\hat{\boldsymbol{\beta}}'\mathbf{z}_0}} \tag{3.5.6}$$

for a given vector of covariates $\mathbf{z}_0$.

3.5.2 Dealing with time-dependent covariates

A nice feature of the Cox model and partial likelihood is that it extends easily to the case of time-varying covariates. Note that the partial likelihood is built by considering only what happens at each failure time, so we only need to know the values of the covariates at the distinct times of failure.

Consider $\mathbf{Z}(t) = (Z_1(t), ..., Z_p(t))$', a vector of time-dependent covariates. The proportional hazards model can be expressed as

$$\begin{aligned} h(t|\mathbf{Z}(t)) &= h_0(t)e^{\boldsymbol{\beta}'\mathbf{Z}(t)} \\ &= h_0(t)e^{\sum_{j=1}^{p}\beta_j Z_j(t)} \end{aligned} \tag{3.5.7}$$

Here, the data are triplets $(T_i, \delta_i, [\mathbf{Z}_i(t), 0 \leq t \leq T_i])$ for $i = 1, ..., n$, where T_i is the time on study (failure or censoring) of the i^{th} patient, δ_i is the usual event indicator and $\mathbf{Z}_i(t)$ is the covariate vector. The likelihood function in (3.5.2) then becomes

$$\mathcal{L}(\boldsymbol{\beta}) = \prod_{i=1}^{m} \frac{e^{\boldsymbol{\beta}'\mathbf{Z}_{k_i}(t_{(i)})}}{\sum_{k\in R_i} e^{\boldsymbol{\beta}'\mathbf{Z}_k(t_{(i)})}}$$

Note that $\mathbf{Z}(t)$ may represent the actual value of a variable at time t, or any index based on the individual's history up to time t. Also note that ties can be implemented just as with fixed covariates.

It is important to realize that estimating the survival and cumulative hazard functions is rather difficult with a time-dependent covariate because $H_0(t)$ depends on the random process $\mathbf{Z}(t)$. Thus, unless $\mathbf{Z}(t)$ is a deterministic function, one needs to estimate the distribution of $\mathbf{Z}(t)$. Finally, in practice, $\mathbf{Z}(t)$ with real data will be a step function.

Remark 3.5.1. Although not mentioned here in detail, another use of time-varying covariates is to check the assumption of proportionality of hazards.

We now have the framework for our survival analyses in the next Chapter. Using "first noncompliance events" (either missing doses or extra doses) to determine censored observations, we will be able to analyze compliance over time between several groups of observations. We will also simulate clinical events and use Cox's model with the plasma concentration compliance variable to compare models with different compliance indices.

Chapter 4

Analyzing ECM data with compliance as response

In order to make inference on the data, every analysis is made on a single type of dose in a single data subset, with the exception of a few survival models in section 4.3. This produces eight difference subsets (or subsamples) when considering one morning dose and one evening dose per supplied dataset, including samples. These 8 data subsets will henceforth be referred to as Morning 1, Morning 2, Morning 3, Morning 4, Evening 1, Evening 2, Evening 3, and Evening 4. We will make sense of this separation when discussing the results in section 4.3.

In this Chapter, we start by the preliminary data analysis in order to examine what the compliance data looks like, using the initial compliance variable that was supplied in the original datasets. This includes histograms and exploratory statistics, such as the means, minimums, and maximums of the compliance proportions, i.e. the ratio of the correct number of doses taken over the number of doses prescribed. We then proceed by fitting survival models to the compliance data and comparing the behaviour of compliance with respect to a "first noncompliance event" between data

subsets and pill types. We will then evaluate how these covariates affect survival using a proportional hazards model. But first, we discuss how these "first noncompliance events" can relate to real world problems.

4.1 Real World Applications

It was already mentioned that compliance is not an all-or-nothing phenomenon; that it evolves over time. While this is true, it does not mean that a first event of either an extra dose or a missing dose is unuseful. These first events may act as a precursor for two clinically worrisome circumstances: missed doses can potentially indicate a patients risk of treatment discontinuation, while extra doses may indicate whether a patient is likely or not to abuse the treatment medication.

4.1.1 First missed dose

Recall from Section 1.2 that compliance can be broken down into persistence and execution. While execution is of more interest within the context of this thesis, persistence is what relates to early discontinuation, which is actually known as "short persistence" [67]. Early discontinuation of treatment is a major problem with long term treatment regimens. For example, it was estimated in [7] that the six month persistence with antihypertensive treatment in a population of 22 918 patients was 68%, meaning that 32% of patients discontinued treatment before six months, even though their prescribed treatment term was for much longer. The analysis of the first missed dose may then be a indicator of early discontinuation of treatment, since patients that are prone to withdraw from studies will do so early on [7].

One can also define a noncompliance event as missing a few consecutive days of treatment (known as a "drug holiday" [36]). If missing several consecutive days of medication can be associated to a true compliance issue, the analyses in this Chapter

could be done with these types of major noncompliance events using time to first drug holiday (for some appropriate definition of drug holiday for the given study) in a similar fashion as what is done with the first missed dose event.

Being able to detect potential patient drop-outs would also be extremely useful in the analysis of RCTs; they allow the investigators to exclude them from primary analyses by using either a modified Intention-to-Treat population or even a (less popular) PP population. These populations, as described in Section 1.2.3, can be used to deal with protocol deviations, such as noncompliance. With the information of potential drop-outs, we can use, for example, a mITT population as patients who must have followed the protocol (and hence, the treatment regimen) accordingly. This would then exclude those patients from the analysis, leading to a more accurate result. This approach is of course very strict, but it serves the purpose of illustrating the utility of analyzing first missed doses from a compliance perspective.

4.1.2 First extra dose

Not all drugs can lead to addiction, but there is a specific type that is more likely to induce addiction than others: painkillers, or opioids. Opioids are important, if not essential, agents in treating certain types of chronic pain. However, the prevalence of drug misuse, abuse, and addiction has fostered considerable consternation among physicians, who may hesitate to prescribe these medications both due to concern for patients and fears of prosecution and/or professional sanction [44]. Passik *et al.* [51] suggest that addiction-related assessment tools into a pain management setting is a good start to preventing eventual drug abuse. One of these tools is the electronic compliance monitoring devices, since they can provide physicians insight to patients' patterns of drug use and compliance so as to direct the type and conduct of treatment that can and should be provided. Indeed, compliance monitoring has been shown to be a useful approach to acquiring information from biological, psychological, and

social domains that can assist in identifying and/or predicting patterns of drug use, compliance, misuse, and abuse [43].

Similarly to the first missed dose, the first extra dose may be an indicator of future drug abuse from patients. Indeed, one could define a "first potential addiction event" consisting of, for example, a few consecutive days where a patient takes more doses than prescribed. Again, the analysis of such an event could be done in the same manner as the analyses of this Chapter. However we do not have enough information to do this with the data we possess.

4.2 Exploratory data analysis

Exploratory data analysis (EDA) techniques have been devised as an aid when trying to determine important characteristics of the data while looking at every single variable for every single observation. Most of these techniques work in part by hiding certain aspects of the data while making other aspects more clear.

EDA is generally cross-classified in two ways. First, each method is either non-graphical or graphical, and second, each method is either univariate or multivariate. Non-graphical methods generally involve calculation of summary statistics, while graphical methods obviously summarize the data in a diagrammatic or pictorial way. In our case, we have a univariate scenario, where the variable of interest is the compliance by pill count defined in the initial dataset. The data that come from making a particular measurement (i.e. compliance) on all of the subjects in a sample represent our observations for that characteristic. We should think of these measurements as representing a "sample distribution" of the variable, which in turn more or less represents the "population distribution" of the variable. We start by the non-graphical EDA, which is used to better appreciate the sample distribution and also to make some preliminary assessments about the population distribution of the variable using the data of the observed samples. We then complement the non-graphical EDA with

its counterpart, the histogram.

4.2.1 Summary statistics

We first start by examining basic compliance statistics using the supplied compliance variable in the initial datasets. Note that the characteristics of our samples are not inherently interesting, except to the degree that they represent the population that it came from.

From the summary statistics shown in table 4.1, we can see that the third dataset (comprised of Morning 3 and Evening 3) seems to not have the same trend in overall compliance, with significantly lower means and medians. Since overall compliance is not normally distributed, as shown in figures 4.1 and 4.2, we can verify this hypothesis by using the Kruskal-Wallis Test, which is the nonparametric version of the ANOVA test using ranks [41]. In fact, testing if all samples come from the same population, we obtain $\chi^2 = 22.96$ on 7 degrees of freedom with an associated p-value of 0.0017. This suggests that *all* samples do not come from the same population, meaning that there is at least one sample that comes from a different source population. If fact, repeating this test excluding Morning 3 and Evening 3, we obtain $\chi^2 = 5.56$ on 5 degrees of freedom with an associated p-value of 0.35, which suggests that the 6 remaining samples originate from the same source population. As the Morning 3 and Evening 3 samples contain only $N = 28$ subjects, these tests should be taken with a grain of salt, as the results are most likely due to sampling variance (of a small sample) than to a difference in populations.

Sample	N	Mean (SD)	Median	(Min, Max)
Morning 1	906	0.66 (0.32)	0.78	$(0.00, 1.00)$
Morning 2	173	0.63 (0.33)	0.70	$(0.00, 1.00)$
Morning 3	28	0.55 (0.24)	0.59	$(0.15, 0.96)$
Morning 4	27	0.64 (0.35)	0.81	$(0.07, 1.00)$
Evening 1	906	0.64 (0.34)	0.78	$(0.00, 1.00)$
Evening 2	173	0.61 (0.34)	0.74	$(0.00, 1.00)$
Evening 3	28	0.45 (0.29)	0.41	$(0.00, 0.91)$
Evening 4	27	0.58 (0.38)	0.76	$(0.00, 1.00)$

Table 4.1: Compliance summary statistics for each sample

4.2.2 Graphical compliance distribution

While the non-graphical methods are quantitative and objective, they do not give a full picture of the data. Therefore, graphical methods, which are more qualitative and involve a degree of subjective analysis, are also required. The most basic graph is the histogram, representing the frequency of cases for a range of values.

In the figures 4.1a-b, we can see some sort of bimodal trend in compliance: that most subjects are either very compliant ($> 80\%$), or very noncompliant ($< 20\%$). This can most likely be associated with the fact that people are usually inherently compliant or noncompliant with their medication prescriptions. For the figures 4.1c-d, this pattern is only somewhat respected in d), and we can see that there is no clear pattern in c). This is likely due to the lack of observations in those samples ($N = 28$ and $N = 27$).

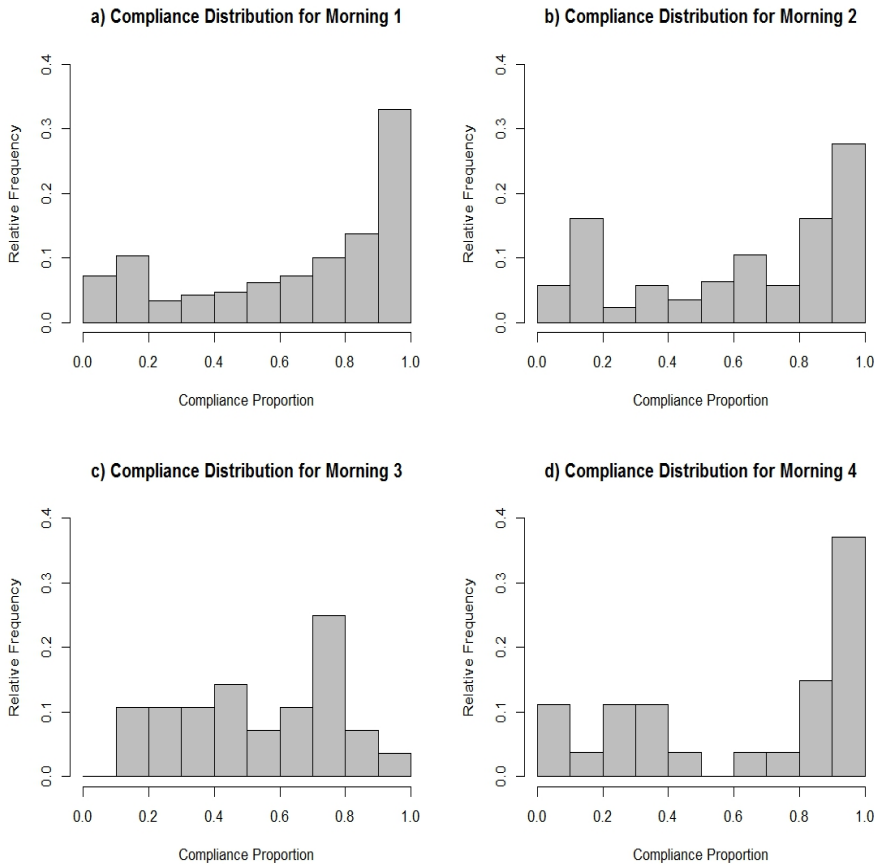

Figure 4.1: Histograms of the four morning samples. a) Morning 1 b) Morning 2 c) Morning 3 d) Morning 4

For the evening doses, the same bimodal pattern is clear in the figures 4.2a-b-d, whereas the figure in c) provides us only with a sense that overall compliance is distributed evenly among subjects.

EDA is most useful when we do not have *a priori* notions of the expected relationships between the variable of interest (i.e. compliance), and other variables that might influence compliance. In the next section, we will use survival models in order to make inference on compliance between variables that might have an effect on compliance.

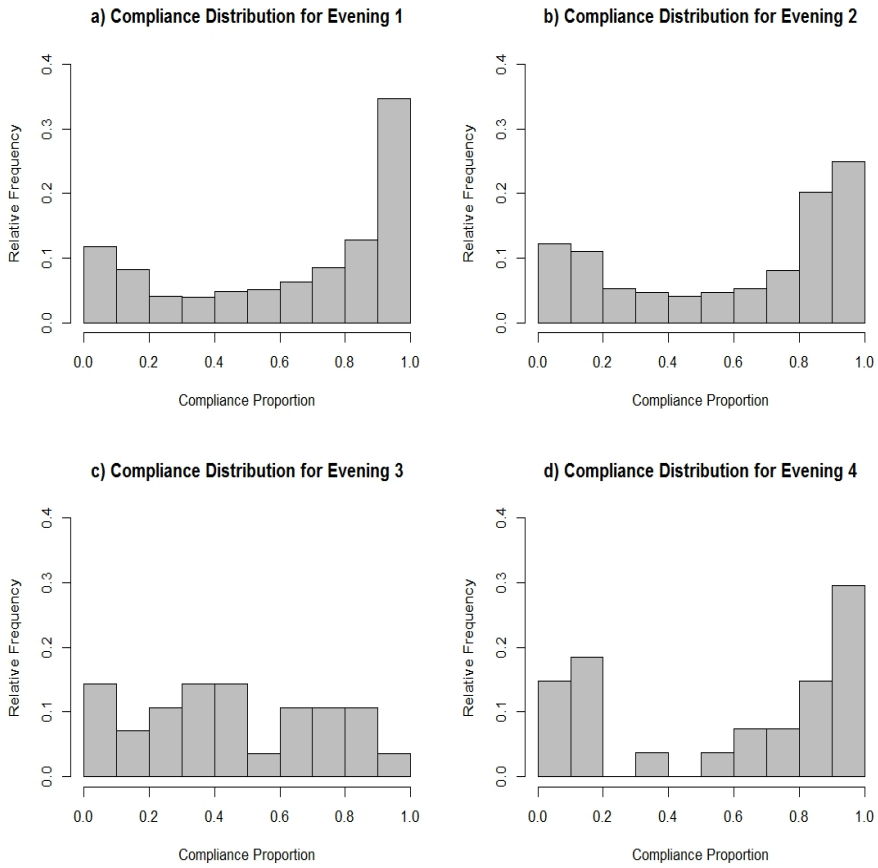

Figure 4.2: Histograms of the four evening samples. a) Evening 1 b) Evening 2 c) Evening 3 d) Evening 4

4.3 Survival analysis for compliance data

The information provided by the ECMs allows us to see whether or not each dose released from a package is administered (assuming instant administration) within the allowed therapeutic window. If a dose is not taken in the allowed time, then this dose administration "event" is noncompliant.

Definition 4.3.1. *We define a* ***noncompliance event*** *as an extra dose or a missed dose during the treatment regimen.*

Note that the missed dose population consists of patients who miss at least a day of dosing. With this information, we can first examine the overall compliance of all 8 samples combined when considering the *first* noncompliant events for each patient.

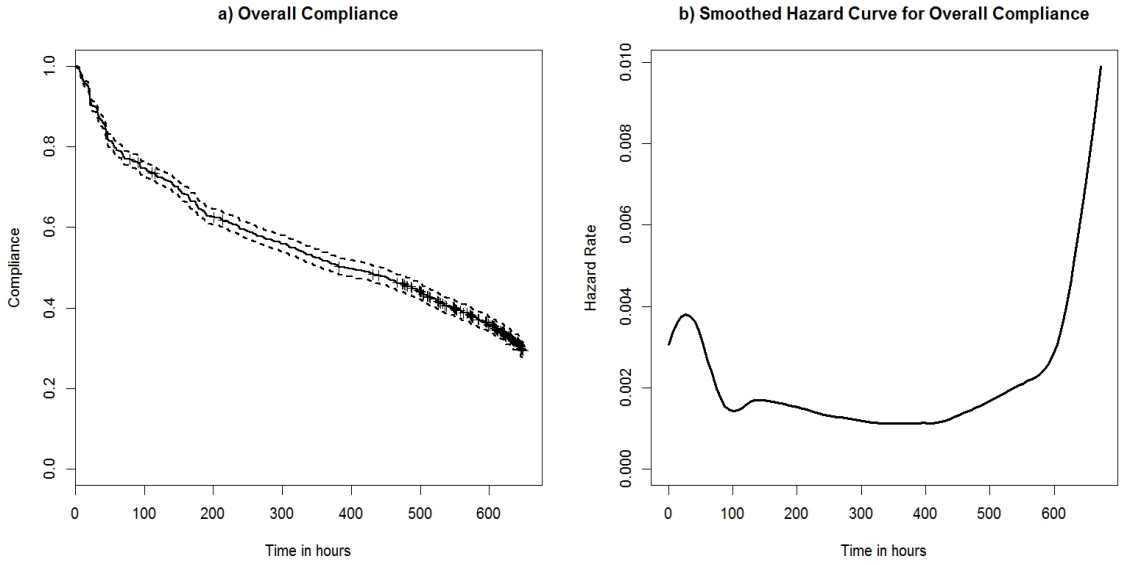

Figure 4.3: a) Overall compliance of all administered doses. b) Smoothed hazard curve for overall compliance.

We can first notice in figure 4.3a that 20% of subjects have a noncompliant event by about 2 days (48 hours) after their first dose. Another 20% of subjects experience a noncompliant event between 50 and 200 hours after the initial dose. This suggests that most first noncompliant events occur before the 8 day mark. This can be explained by human behaviour: if a patient is inherently noncompliant, the first noncompliant event is going to occur more likely than not at the beginning of the study. It is also interesting to note that only about 25% of subjects are fully compliant over the period of observation. This can seem low, but it actually agrees with the literature [9, 11, 12, 15, 19, 31, 32] (most studies define a compliance proportion of $> 80\%$ in order to be compliant, unlike our situation which only captures patients

with 100% compliance rate). Notice also the slight drop in compliance at the end of the study period around 500 hours. One possible explanation for this phenomenon is purely behavioural: patients tend to become less compliant the longer the study runs, especially if the patient does not feel any improvement (or less than expected) in his health [10]. Although we do not have any information on treatment efficacy according to individual patients, it is possible that the treatment on study loses its effectiveness over time. Conversely, the perceived efficacy of the treatment might so good that patients might decide to discontinue the treatment because they think that it is no longer needed. If we look at the smoothed hazard curve in 4.3 b, we can see that the increase in hazard rate matches the sudden drops in compliance, both at the beginning of the study and at the end. The sudden increase in early hazard might be explained by the behaviour of patients that discontinue the treatment early due to "overwhelming" efficacy. One possible explanation (however unlikely) is that the spike in hazard near the end of the study is the result from the toothbrush effect (as described in Section 1.2.1), i.e. patients want to give the illusion of taking all their medication (which results in noncompliance by taking extra doses) right before a visit with a clinician. However, the increase in hazard around 600 hours might also be due to inherent data issues within the IMC datasets, namely the presence of an "empty" record belonging to the fifth package at the end of the four week study period.

The overall compliance graph in 4.3 has a slight issue however: the first non-compliant event consists of either an extra dose or a missed dose, whichever comes first. The problem is that an extra dose is in most cases a deliberate or intentional act of taking an additional pill when it is not required, while a missed dose, again in most cases, is the act of forgetting which is unintentional. This can be seen in figure 4.4, where we can see the steep decline in compliance for extra doses within 24 hours, whereas the decline in compliance for missed doses is steadier throughout the regimen. This rationale however is based on the instant administration assumption; it is possible (and likely) that some of these extra doses are not *real*, but rather a result of,

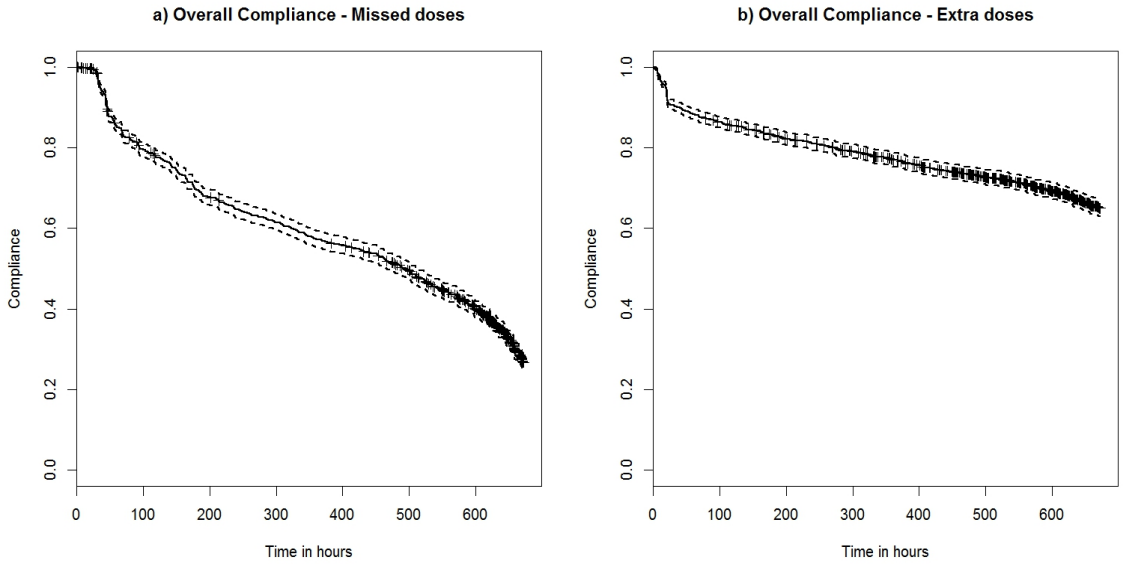

Figure 4.4: Overall compliance of all administered doses for missed doses and extra doses separately.

say, a patient releasing all doses and storing them in another container that is more suitable for their lifestyles. Another possible explanation of extra dose inflation is when a dose becomes "unusable" due to manipulations (i.e. dropping or losing a pill) and then truly administering only a second pill. The ECM devices then record two dose releases, but in reality only the second one is taken, while the first is discarded, and thus the patient remains compliant but our method records otherwise; this is an inherent issue with extra doses. But, regardless of the possible explanations, it is clear that the two curves are quite different from one another. Since they are considered as two different outcomes, no formal hypothesis test can be made, although it is clear that these two events follow two different patterns, and that analyzing them separately is appropriate.

It is important to realize that figure 4.3a is based on the first noncompliance event, whether it is an extra dose or a missing dose. It tells us that extra doses

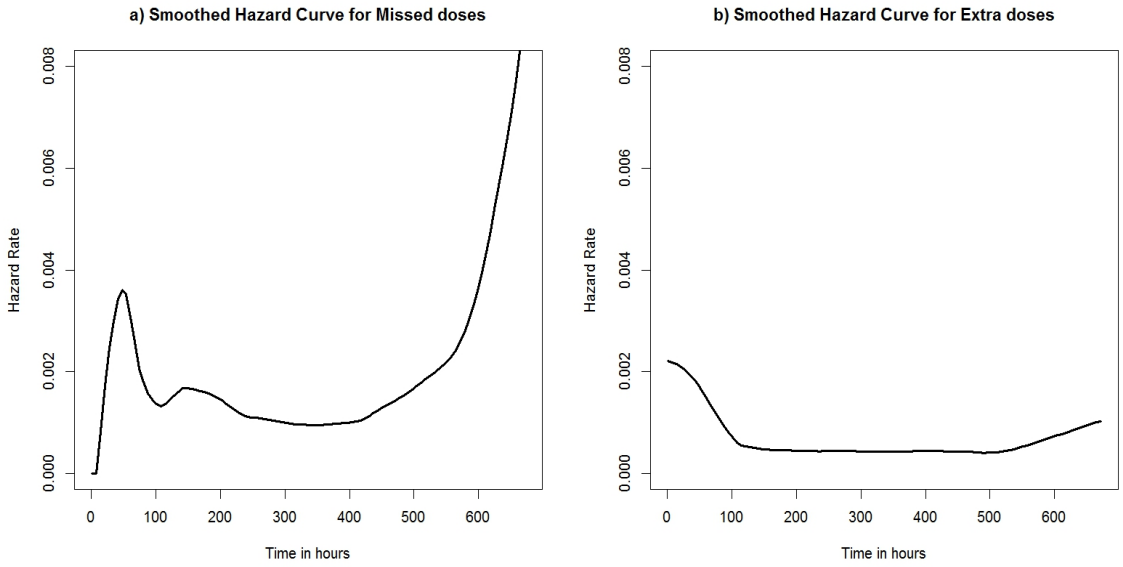

Figure 4.5: Smoothed hazard curves of all administered doses for missed doses and extra doses separately.

are much more prevalent compared to missed doses at the beginning of the regimen; however, this does not mean that extra doses are more prevalent than missed doses over the treatment period when analyzing them separately. In fact, we can see from 4.4 that except for the rapid drop of compliance due to extra doses after 1 day of treatment, compliance has a greater decline due to missing days of treatment. The graph shows that $> 80\%$ of patients experience a missed dose over the period of observation, as opposed to $< 50\%$ that experience an extra dose. To explain this, one can suggest that there are fewer inherently noncompliant patients who take extra doses than there are "normal" patients who sometimes forget to take their medication. Next, if we look at figure 4.5a, we can see that the hazard rate for missing doses is quite similar to the hazard curve for overall compliance 4.3b with the increase in hazards at the beginning and at the end of the study (where the sudden increase in risk at the end might be due to a poor estimation of the tails, or because of the nature

of the data in the structure of the end of follow-up), suggesting that in general, missed doses as first noncompliance events are much more prevalent in general compared to extra doses as first noncompliance events. If we look at figure 4.5b, the hazard is very flat between 100 and 550 hours, which suggests that the risk of taking an extra dose is constant within this time interval, following an exponential distribution. The hazard rate then increases slightly after 550 hours (which can be related to the possible explanations mentioned a little earlier) until the end of the observation period.

Now that we have introduced the missed dose population and the extra dose population separately, we can proceed by modeling their compliance with respect to pill type (Morning vs. Evening), data subsets.

4.3.1 Comparing compliance between morning and evening doses

The rationale behind testing if the medication taking behaviour between the morning pill and evening pill differs arises from human nature. People generally follow a routine in their everyday lives; without habits, people would be doomed to plan, consciously guide, and monitor every action [49]. It is also reasonable to assume that these habits are different between the mornings and evenings, which in turn could result in a different pill taking behaviour at both times of the day. We can thus compare compliance between the two pill types, stratified of course by type of noncompliance.

We can see from figure 4.6a that there does not seem to be a difference in missed dose compliance between morning pills and evening pills. In fact, testing equality between compliance curves with respect to missed doses yields $\chi^2 = 1.4$ on 1 degree of freedom with associated p-value = 0.236, which suggests that we cannot reject the possibility that both compliance curves are equal. If we look at 4.6b, we can see that both compliances behave the same for the first 100 hours of the regimen,

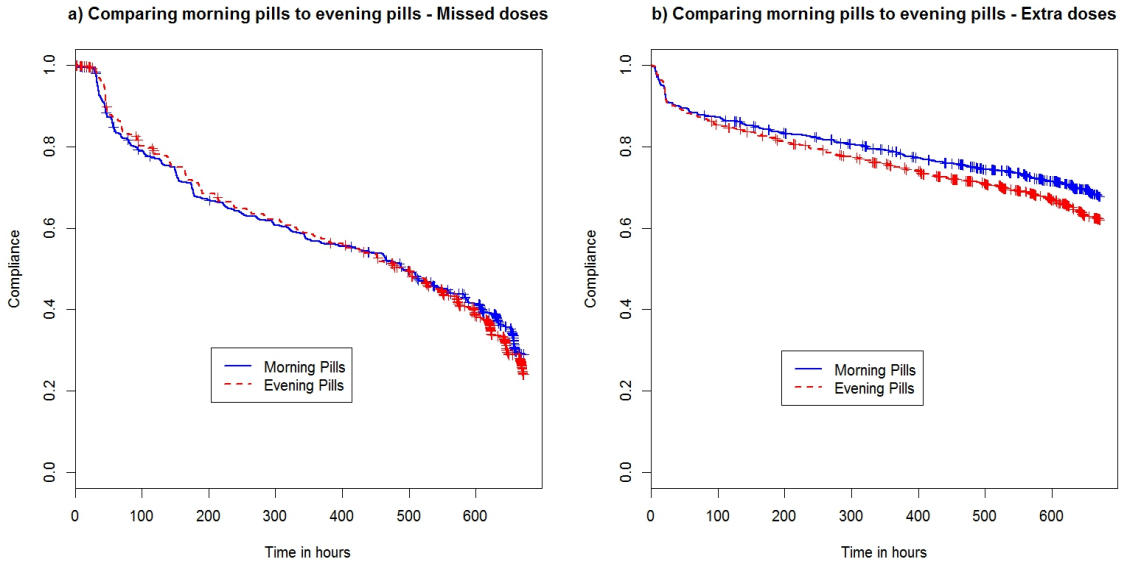

Figure 4.6: Comparing compliance by pill type. a) Morning pill vs Evening pill in missed doses. b) Morning pill vs Evening pill in extra doses.

and then start to separate. Extra dose noncompliance due to morning pills seems to be less severe than noncompliance due to evening pills. Testing equality between compliance curves with respect to extra doses yields $\chi^2 = 6.3$ on 1 degree of freedom with associated p-value $= 0.012$, which indicates that the compliance curves may be different. The results suggest that there exists a difference between the morning dose and the evening dose in noncompliance due to extra doses, resulting in a higher rate of noncompliance due to extra doses in the evening compared to the morning. The results also suggest that there is no difference in noncompliance between morning and evening doses due to missed doses.

4.3.2 Comparing compliance between samples

Once we have compared compliance between the two different doses, it is a good idea to compare the different samples within the same doses in order to determine

if subjects' morning and evening pill-taking behaviours can be generalized across the whole data. Since there was a difference in compliance between both pills for extra doses, it is only natural to separate compliance into either morning or evening samples for both types of noncompliance.

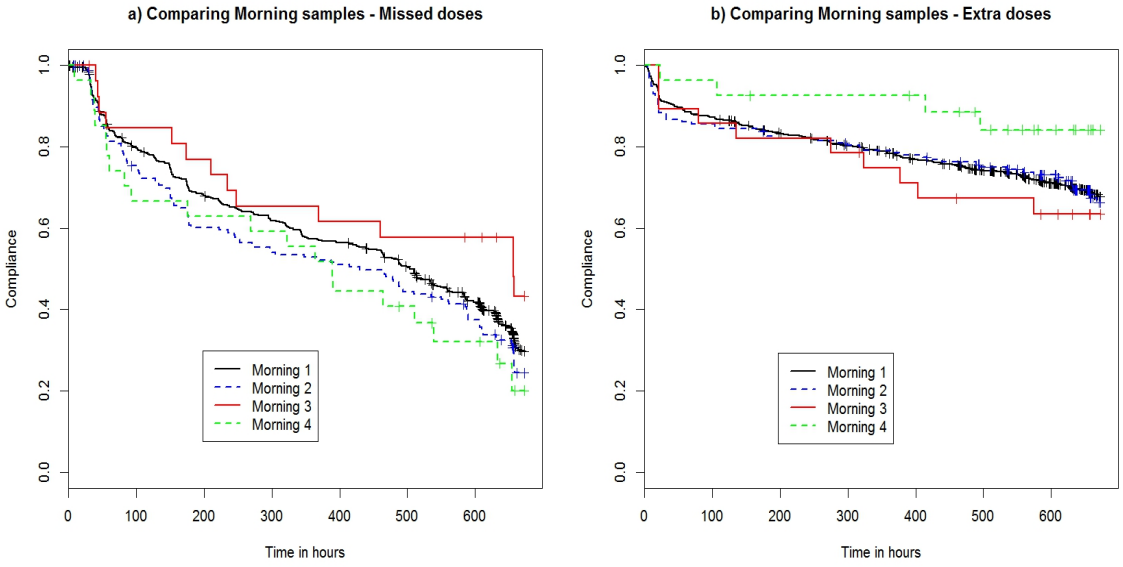

Figure 4.7: Comparing compliance by Morning samples. a) Morning samples in missed doses. b) Morning samples in extra doses.

We can see from figure 4.7a that the noncompliance due to missed doses is slightly different between each morning sample, although they each follow the same pattern. Testing equivalence between compliance curves results in $\chi^2 = 5.8$ on 3 degrees of freedom with associated p-value $= 0.120$, suggesting that we cannot reject the possibility of the curves being the same. If we look at 4.7b, we can notice that the smaller samples (particularly Morning 4) seem to differ in compliance, although both larger samples have practically the same curve. Testing for equality between these yields a $\chi^2 = 3.9$ on 3 degrees of freedom with associated p-value $= 0.274$, again resulting in lack of evidence to reject the hypothesis that the morning pill compliance

curves due to extra doses are the same. It would then seem reasonable to assume that the pill taking behaviours are the same across all morning samples.

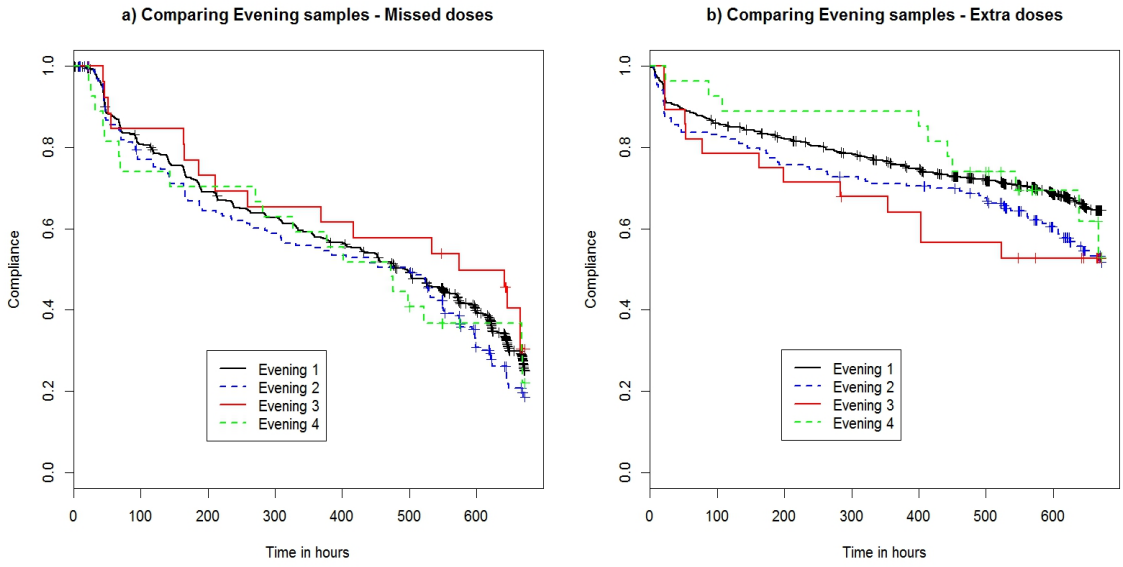

Figure 4.8: Comparing compliance by Evening samples. a) Evening samples in missed doses. b) Evening samples in extra doses.

We now proceed to do the same analysis as above for the evening samples. From figure 4.8a that there doesn't seem to be any difference in noncompliance due to missed doses. Testing for equality between compliances results in $\chi^2 = 2.9$ on 3 degrees of freedom with associated p-value $= 0.412$, suggesting that we cannot reject the possibility of the compliance curves being the same between the 4 evening samples with respect to missed doses. If we look at figure 4.8b, we can see that the difference between compliance curves is rather significant, again particularly between the smaller samples. Testing compliance equivalence yields a $\chi^2 = 8.0$ on 3 degrees of freedom with associated p-value $= 0.046$. We thus have enough evidence to reject the hypothesis that the compliance curves are the same at significance level $\alpha = 0.05$. This result suggests that the pill taking behaviours are different across all evening

samples. However, it is important to note again that the significant differences detected by the previous test are most likely not due to a difference in source population, but rather to the inherent variability of the small samples.

As mentioned in section 3.4.3, these tests are only meant to see *if* the pill types and the different samples affect compliance, and not *how* compliance is affected by them. In the next section, we will use Cox's proportional hazards model to analyze the functional relationship between pill type and samples and compliance.

4.4 Proportional hazards model

4.4.1 Fitting the PH model

As the only significant difference observed with the Kaplan-Meier analysis above is a difference in compliance with respect to extra doses between the morning pills and the evening pills, we will now perform a confirmatory analysis using a proportional hazards model to determine if the pill type produces different compliance rates among extra doses. We start by fitting the model

$$\lambda(w) = e^{\beta w}\lambda_0$$

where $w \in \{1, 2\}$ represents the pill type (1=morning, 2=evening), and λ_0 is the baseline rate of noncompliance. Upon fitting this model, we obtain an estimated $\hat{\beta} = 0.186$ with associated p-value = 0.012. This result suggests that the rate of noncompliance due to extra doses is different between morning doses and evening doses. Exponentiating our estimate gives us $e^{0.186} = 1.204$, with a 95% confidence interval of (1.042, 1.393). This suggests that the rate of noncompliance due to extra doses is 20.4% higher for the evening pills compared to morning pills, agreeing with the KM analysis from section 4.2.1.

A plausible explanation for this difference is simply that patients (or rather,

people in general) have different habits in the morning than in the evening, thus affecting medication-intake behaviour, as suggested in [49]. For what reason do we see a difference in noncompliance due to extra doses and not missed doses? It can possibly be explained by the fact that, as mentioned in section 4.2, the act of taking an extra dose is in most cases deliberate, while missing a dose is generally unintentional, and therefore patients may deliberately be noncompliant with one pill or the other depending on their habits. Another possible reason for this difference may lie in medication posology: the regimen consists of 2 morning pills and 1 evening pill. It is not impossible that patients may sometimes think that they need to take 2 pills in the evening as well.

4.4.2 Testing the proportionality of hazards assumption

Now, in order to test the goodness-of-fit for our model, we much check whether the proportionality assumption is satisfied. This test will be done in two different ways. We will first verify the proportionality assumption by plotting the behaviour of the estimate $\hat{\beta}$ for pill type over time, where a systematic departure from a horizontal line is indicative of non-proportional hazards. Using the *cox.zph* function in R, we can plot the scaled Schoenfeld residuals against transformed time for each covariate in a model; in our case, the scaled residuals are plotted for pill type.

Figure 4.9 shows that the line is close to horizontal where the majority of observations are located (between times 30 and 500), suggesting that proportionality assumption is somewhat satisfied. In fact, testing the null hypothesis of proportional hazards, we obtain a Schoenfeld residuals-based chi-squared test statistic proposed in [75] of $\chi^2 = 2.6$ on 1 degrees of freedom with associated p-value = 0.11. This results in a lack of evidence to reject the null, suggesting proportionality is satisfied for pill type.

The second test consists of fitting the model with an additional covariate $w_p =$

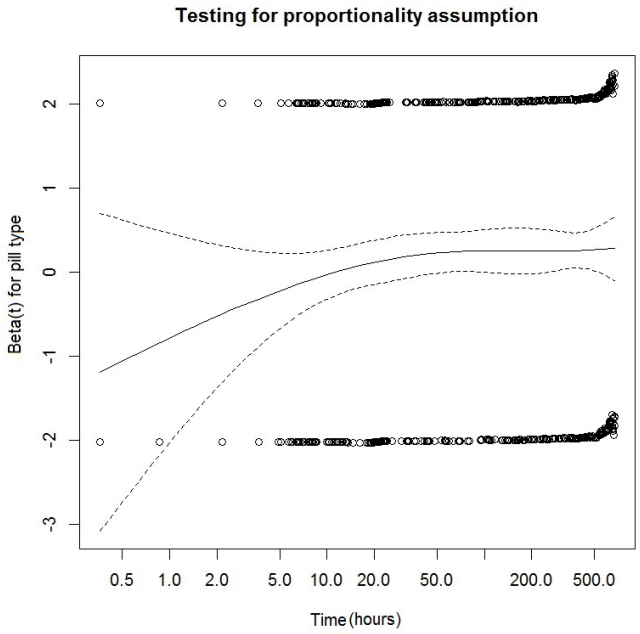

Figure 4.9: Testing for proportionality of hazards assumption for Morning and Evening pills.

$w \times \log(t)$ (as mentioned in remark 3.5.1) and verifying its statistical significance in the model. Upon fitting this new model, we obtain a statistically significant contribution of w_p ($Z = 3.47$, p-value $= 0.0005$), suggesting that the proportionality assumption is violated, disagreeing with the first test. However, even if there is no consensus on whether proportionality is satisfied or not, the analysis above is more exploratory than anything else, meaning that these results must be taken lightly.

Chapter 5

Simulation of events based on $C_p(t)$ compliance

5.1 Generating event times from a nonhomogeneous Poisson process

In order to make and compare inferences using the information provided by the ECMs, a model is required to simulate events with respect to $Z_i(t)$. These simulated events will be associated with the sequence of intervals where the subjects are noncompliant, i.e. when the drug has no therapeutic effect. The events will then be used in a survival model, where we will investigate and compare the effectiveness of a time-dependent compliance variable in predicting clinical events.

Now, consider a simulation of some system which evolves through time. In our case, the system would be the clinical events of patients under a dosage regimen. There are two possibilities for patients: the clinical event occurs during the regimen, considered as a failure time, or the clinical event does not occur during the dosage regimen, which would entail a time that is censored, as we do not know what would happen after the treatment is complete. This approach is called *discrete event*

simulation.

Definition 5.1.1. *[58]. Suppose that events are occurring at random time points and let $N(t)$ be the counting process denoting the number of events that occur in the time interval $[0, t]$. These events are said to constitute a* ***Poisson process with rate*** *λ, where $\lambda > 0$, if*

- $N(0) = 0$.
- *The numbers of events occurring in disjoint time intervals are independent (*independent increment assumption*).*
- *The distribution of the number of events that occur in a given interval depends only on the length of the interval and not on its location (*stationary increment assumption*).*
- $\lim_{h\to 0} \frac{P[N(h)=1]}{h} = \lambda$.
- $\lim_{h\to 0} \frac{P[N(h)\geq 2]}{h} = 0$.

From a modeling point of view, the major weakness of the Poisson process is its assumption that events are just as likely to occur in all intervals of equal size. A generalization, which relaxes this assumption, leads to the nonhomogeneous process.

Definition 5.1.2. *[58]. Suppose that events are occurring randomly in time, and let $N(t)$ denote the number of events that occur by time t. Then we say that $\{N(t), t \geq 0\}$ constitutes a* ***nonhomogeneous Poisson process with intensity function*** *$\lambda(t)$, $t \geq 0$, if*

- $N(0) = 0$.
- *The numbers of events that occur in disjoint time intervals are independent.*
- $\lim_{h\to 0} \frac{P[\textit{exactly 1 event between } t \textit{ and } t+h]}{h} = \lambda(t)$.

- $\lim_{h\to 0} \frac{P[\textit{2 or more events between } t \textit{ and } t+h]}{h} = 0.$

The quantity $\lambda(t)$, called the intensity at time t, indicates how likely it is that an event will occur around the time t. Note that $\lambda(t) \equiv \lambda$, the nonhomogeneous reverts to the usual Poisson process.

In the simulation of clinical events, we suppose that the patients' failure time arrives in accordance with a non-homogeneous Poisson process with a bounded intensity function $\lambda(t)$. In our case, the intensity function is defined as the treatment effect. The treatment effect varies over time as it is tied to $C_p(t)$ compliance. That is, if a patient is noncompliant, it is equivalent to having no real beneficial therapeutic effect.

Definition 5.1.3. *We define the* ***rate of clinical event*** *at time t for the i^{th} dose as*

$$\lambda(t) = \begin{cases} \lambda_1 & \textit{if } C_p^{(i)}(t) \geq C_{MEC} \\ \lambda_0 & \textit{otherwise} \end{cases} \tag{5.1.1}$$

where λ_1 is the reduced rate from the true beneficial therapeutic effect, and λ_0 is the original (baseline) rate of event. Note that $\lambda_1 < \lambda_0$ if the clinical event is unfavourable (eg. time of relapse), and $\lambda_1 > \lambda_0$ if the clinical event is favourable (eg. time of remission).

Example 5.1.4. Suppose that the true beneficial therapeutic effect reduces the rate of occurrence of the clinical event under study by half, that is, the patients who are compliant are 50% less likely to be assigned a clinical event during the dosage regimen, thus $\lambda_0 = 1$. Then the rate of clinical event can be written in accordance to a Cox model as

$$\lambda_1(t|Z(t)) = e^{\beta_1 Z(t)}\lambda_0 = \begin{cases} 0.5 & \text{if } Z(t) = 1 \\ 1 & \text{if } Z(t) = 0 \end{cases}$$

where $\beta_1 = -\log(2)$ is the true beneficial therapeutic effect and $z(t)$ is the compliance variable defined in the previous section. Now, suppose we want to compare this to a model which considers a fixed compliance variable z, averaged over the treatment regimen. Then we would compare $\lambda_1(t|Z(t))$ with

$$\lambda_2(t|z) = e^{\beta_2 z}\lambda_0$$

In the end, we want to see whether $\hat{\beta}_1$ or $\hat{\beta}_2$ is a better estimator for β.

Subroutine for generating the time of events

This is the subroutine for generating clinical event times proposed by [58], where T_{event} is the random variable defined to equal the time of the first arrival after the starting time s.

- Step 1: Let $t = s$
- Step 2: Generate $U \sim Uniform[0, 1]$
- Step 3: Let $t = t - \frac{1}{\lambda_0}\log U$
- Step 4: Generate $V \sim Uniform[0, 1]$
- Step 5: If $V \leq \frac{\lambda(t)}{\lambda_0}$, set $T_{event} = t$ and stop
- Step 6: Go to Step 2

Thus, the clinical event occurs at the first generated time with probability 1 when patients are noncompliant with respect to the plasma concentration level.

5.2 Simulation methodology

5.2.1 Study design

The underlying study design that was chosen for this simulation is the RCT. The control group can be considered as the group where compliance will be captured as a fixed (average) rate λ_0 for each subject $i = 1, ..., N$ over the course of the study. The control group can then be defined as $Z_i(t) = 0 \ \forall \ t$ for the time-dependent variable using the ECM data and $Z_i = 0$ for the variable without the ECM data and for the fixed compliance rate. The treatment group can be defined as $Z_i(t) = \mathbf{1}[C_p(t) > C_{MEC}]$ for the compliance variable, $Z_i = 1$ without compliance data, and $Z_i = z \in [0, 1]$ for the fixed compliance rate. Note that the fixed compliance rates Z_i are defined as the amount of doses taken within the allowed therapeutic window over the total amount of doses prescribed.

5.2.2 Populations

The simulations were done using the 8 subsamples defined in Chapter 4, where

- $N_{Morning1} = N_{Evening1} = 906$
- $N_{Morning2} = N_{Evening2} = 173$
- $N_{Morning3} = N_{Evening3} = 28$
- $N_{Morning4} = N_{Evening4} = 27$

Each subsample provided the information for both compliance variables (that is, from a population perspective, subjects dosing histories will provide us with both their fixed compliance rate and their time-dependent compliance). Only those subjects who had at least 2 recorded times in the IMC datasets were considered to be part of the simulation analysis. The simulated RCTs thus each have two arms of size $N \in \{27, 28, 173, 906\}$.

5.2.3 Modelling

As mentioned in example (5.1.4), we are going to fit two models to that assess the *rate* of the simulated clinical events occurring:

$$\text{Model 1 } = \lambda_1(t|Z(t)) = e^{\beta_1 Z(t)}\lambda_0(t) \quad vs \quad \text{Model 2 } = \lambda_2(t|Z) = e^{\beta_2 z}\lambda_0(t)$$

where λ_0 is defined as the baseline rate of the event (by baseline, we mean the rate of the event occurring when patients do not take any treatment, or equivalently when patients are never $C_p(t)$ compliant). For our purpose, we will limit ourselves to the case where $\lambda_0(t) = \lambda_0 = \frac{1}{400}$, a constant. The choice of $\lambda_0 = \frac{1}{400}$ was made in order to have roughly 20 to 40% censoring in the outcome over a period of time based on the actual data. If censoring was too high, we would not be able to distinguish between groups for lack of a sufficient number of events in the data. The clinical event times for the treatment group were generated by a nonhomogeneous Poisson process (as described in Section 5.1) based on the **true** $\lambda(t)$ with $Z(t)$. We can then consider Model 1 as the **true model**, where $Z(t)$ is *perfectly observed*, whereas Model 2 is a miss-specified model constructed with a fixed compliance Z defined by IMC (see the histograms in Section 4.2.2), which would be considered as a standard approach had we not had the temporal aspect in our ECM data. Table 5.1 shows how the data was organized for analysis in R.

This modelling allows our control group to have $\lambda_{1,2}(t) = \lambda_0 = \frac{1}{400}$ without the presence of a placebo effect caused by patient behaviour since our model does not allow medical effect of the placebo (which would require additional covariates and parameters). The treatment group can be defined as Model 1 (above) based on the parameters from Remark 2.2.4.

It is important to note that while the clinical events are simulated via discrete event simulation, the dosing times, and thus compliance, are **not** simulated. Com-

ECM ID	Start time	End time	δ	$Z(t)$	Z
40000124	0.00000	22.31983	1	1	0.9642857
40000613	0.00000	53.59461	0	1	0.9230769
40000613	53.59461	63.16862	0	0	0.9230769
40000613	63.16862	124.88746	0	1	0.9230769
40000613	124.88746	125.81446	0	0	0.9230769
40000613	125.81446	177.16174	1	1	0.9230769
40000676	0.00000	37.13444	0	1	0.8148148
40000676	37.13444	37.15006	0	0	0.8148148
40000676	37.15006	80.51025	0	1	0.8148148
⋮	⋮	⋮	⋮	⋮	⋮

Table 5.1: Example of the data organization in R used for the analyses.

pliance was extracted from the real data, and so Z_i and $Z_i(t)$ are also based on the parameters from Remark 2.2.4.

Remark 5.2.1. The simulated outcome generated by the non-homogeneous Poisson process based on the extracted $C_p(t)$ compliance was **based on the C_{MEC} and not on the C_{COMP}**. This is because the C_{MEC} represents the true effective treatment period, even if this causes some small time intervals where patients are under the therapeutic threshold. These time intervals where treatment (in)efficacy contradicts patient compliance should cumulate to a negligible amount of time over the course of the study.

Notice that $\lambda_1(t|Z(t))$ is the rate of a piecewise exponential distribution since $Z(t)$ is a step function, while $\lambda_2(t|Z)$ is the fixed rate of an exponential distribution. This choice was made to simplify the discrete event simulation algorithm. Note that we have dropped the subject specific indices for notation purposes. Our goal in comparing both models is to analyze the estimators of the TBE, to see whether the first model can produce better estimates of the treatment effect compared to the second model in a time-to-event study.

5.2.4 Analysis

We fit the models for $K = 100$ simulations and we calculate $\hat{\beta}_j$, $j = 1, 2$, resulting in $\hat{\beta}_1^{(1)}, ..., \hat{\beta}_1^{(100)}$ and $\hat{\beta}_2^{(1)}, ..., \hat{\beta}_2^{(100)}$ that we compared via the (absolute) empirical bias, the empirical standard error and the empirical mean squared error.

Definition 5.2.2. *We define the* ***empirical bias*** *of $\hat{\beta}_j$ to be given as*

$$EB[\hat{\beta}_j] = \frac{1}{K}\sum_{i=1}^{K}\hat{\beta}_j^{(i)} - \beta$$

over the K simulations, where $j = 1, 2$. Furthermore, we define the ***absolute empirical bias*** *of $\hat{\beta}$ to be*

$$EAB[\hat{\beta}_j] = \frac{1}{K}\sum_{i=1}^{K}|\hat{\beta}_j^{(i)} - \beta| \tag{5.2.1}$$

Definition 5.2.3. *We define the* ***variance*** *of $\hat{\beta}_j$ to be given as $\sigma_j^2 = \frac{1}{K-1}\sum_{i=i}^{K}(\hat{\beta}_j^{(i)} - \bar{\hat{\beta}}_j)^2$ for $j = 1, 2$ at simulation i. We define the* ***empirical standard error*** *of $\hat{\beta}$ over K simulations to be given as*

$$ESE(\hat{\beta}_j) = \sqrt{\frac{1}{K-1}\sum_{i=1}^{K}(\hat{\beta}_j^{(i)} - \bar{\hat{\beta}}_j)^2} \tag{5.2.2}$$

where $\bar{\hat{\beta}}_j = \frac{1}{K}\sum_{i=1}^{K}\hat{\beta}_j^{(i)}$.

Definition 5.2.4. *We define the* ***empirical mean squared error*** *of $\hat{\beta}_j$ such as*

$$EMSE[\hat{\beta}_j] = EB^2[\hat{\beta}_j] + ESE^2[\hat{\beta}_j] \tag{5.2.3}$$

for $j = 1, 2$.

Upon fitting the models will compare the EBs of $\hat{\beta}_1$ and $\hat{\beta}_2$ compared to the true parameter $-\beta = \log(TBE)$ as shown in 5.2.2, the ESEs of both estimates as shown in 5.2.3 and their EMSEs as shown in 5.2.4. For each sample, 4 different beneficial effects are simulated in order to determine if we can accurately estimate β; these are

$TBE_1 = 1.25$, $TBE_2 = 1.50$, $TBE_3 = 2.00$ and $TBE_4 = 4.00$ times lower rate of having the simulated clinical event occur (eg death) while being C_p compliant with the treatment schedule compared to not. We will also calculate the ESE associated with these estimates, to finally calculate the EMSE, which in turn we will use in order to compare $\hat{\beta}_1$ with $\hat{\beta}_2$

5.3 Results

First and foremost, we can point out that the data in the Morning 3 and Evening samples are not distributed the same way as the other samples, as shown in Chapter 4. Therefore, it is reasonable to assume that the results from these samples are not representative of the true population, and therefore will be ignored while discussing results. (the results are in the tables for exploratory purposes only).

We start by analyzing the results of $\hat{\beta}_1$, presented in table 5.2. We can first notice that, for each TBE, the average bias is smaller as the sample size increases, as expected. However, the average biases for the different TBEs within a same sample seem to differ; the largest treatment effect ($\exp(-\beta) = 4$) has the greatest bias among all treatment effects. One explanation for this result is that this TBE is too *high*. This leads to a higher variability in the simulated clinical event times and thus overdispersion might occur. With higher variability in clinical events, we then obtain a larger bias, a larger ESE and thus a larger EMSE. Another explanation lies in the fact that the space where $Z(t) = 1$ is not sufficiently covered, as mentioned above. If the clinical event occurs more rapidly when patients are $C_p(t)$ noncompliant, any change in $Z(t)$ (whether it be during a few seconds or a few hours) from 1 to 0 might cause the event to occur and thus terminate the observation. The accuracy of the estimator $\hat{\beta}$ depends on there being sufficient events where $z(t) = 1$. If there are too many censored observations in the $Z(t) = 1$ group, testing statistical significance of $\hat{\beta}$ in the model may not be accurate.

The average standard errors across all TBEs within samples are fairly similar, resulting in approximately equal variances among all TBEs within samples. The EMSEs follow the same pattern as well, except that it is slightly higher when TBE=4.0 due to larger EB.

Now, let us consider table 5.3 which contains the results for Model 2. The general pattern is the same as in table 5.2, from the EB which decreases as sample size increases, to the behaviour of the EMSEs. We can notice however that within the same pattern, the values are larger. Let us take the results from the Morning 1 sample again. We obtain relative EBs of 59%, 34%, 28% and 21% for TBEs=1.25, 1.5, 2.0 and 4.0 respectively. These relative EBs are overall larger than the ones obtained in Model 1. We can also see from the table that the ESEs are higher in Model 2, leading to larger EMSEs. Again, we reiterate that similar results can be seen across all samples (except of course for Morning 3 and Evening 3).

Note that these results do not incorporate the Morning 3 and Evening 3 samples. In both tables (particularly for Model 1) we can see that the EMSEs are much higher than the other samples. From Chapter 4 we found that there is reason to believe that these samples do not have the same compliance distribution as the others, even if the sample size is comparable to the Morning 4 and Evening 4 samples. The average compliance rates were significantly different than the other 6 samples, being lower by 20 to 30%.In theory, this should not affect the behaviour of our estimates when analyzing simulated events. In practice however, we do see a huge difference, for example where the EB goes up to 691%. Why do we see this? The problem of a small sample size which in turns produces a higher ESE is of course part of the problem. However, the larger issue here is that to efficiently estimate β, we require a sufficiently large coverage of $Z(t)$ where $Z(t) = 1$. If $Z(t) = 0$ for a large part of the study, we simply do not have enough coverage when $Z(t)$ to obtain a consistent estimate of β. In other words, if patients do not adequately follow their prescribed treatment schedule, we will need many more to better estimate the TBE even with

the ECM data, which as we can see does not fully overcome the lack of compliance.

In table 5.4 we can find the comparison between $\hat{\beta}_1$ and $\hat{\beta}_2$, including the proportion where $\hat{\beta}_1$ is less biased than $\hat{\beta}_2$, the proportion of confidence intervals for $\hat{\beta}_1$ which contains the true parameter β_1, and the ratio of the average EMSEs over the $K = 100$ simulations. We can see from table 5.4 that the proportion where $\hat{\beta}_1$ is less biased than $\hat{\beta}_2$ lies roughly between 60% and 70% (the average across all samples excluding Morning 3 and Evening 3 is actually 63%). This means that the bias in Model 1 is less than the bias in Model 2 in a 2:1 ratio, on average. This agrees with the results from the previous tables, noting the differences in bias between both estimates. Next, if we look at the proportion of simulations where the 95% $\text{CI}_{\hat{\beta}_1}$ contains the true TBE, we can see that most proportions respect the theory (meaning that 95% of the time, the $\text{CI}_{\hat{\beta}_1}$ will contain the true parameter), although it can be seen that even the largest samples (Morning 1 and Evening 1) coverage is slightly less than 95%, mostly due to the bias. On the other hand, the ESE is so large for the smallest sample sizes that the produced coverages are larger than nominal. Although they are omitted in the table, this proportion was also computed for 95% $\text{CI}_{\hat{\beta}_2}$, obtaining similar, but slightly smaller, proportions. It would take more simulations and more time to better understand this coverage problem.

We now arrive at the heart of the comparison between the two models: the EMSE ratio. It is clear from the simulations that $\text{EMSE}[\hat{\beta}_1] < \text{EMSE}[\hat{\beta}_2]$. In other words, the estimator $\hat{\beta}_1$ predicts the clinical event times with TBE=$e^{-\beta}$ *better* than $\hat{\beta}_2$. The EMSE ratio allows us to quantify the relative decrease of variation due to randomness (i.e. variance) and bias. We can see from the table that the average EMSE ratios mostly lie between 0.60 and 0.75 (excluding the proportion for TBE=4.0 in Evening 4, which is believed to be due to small sample size). In fact, averaging these ratios across all samples yields 0.672, meaning that the average reduction in EMSE across all samples is 32.8%.

By fitting and comparing two proportional hazards models to illustrate the re-

Sample	TBE $= e^{-\beta}$	$EB[\hat{\beta}_1]$	$ESE[\hat{\beta}_1]$	$EMSE[\hat{\beta}_1]$
Morning 1	1.25	0.102	0.331	0.120
	1.50	0.092	0.330	0.117
	2.00	0.132	0.332	0.128
	4.00	0.288	0.348	0.204
Morning 2	1.25	0.185	0.488	0.272
	1.50	0.192	0.486	0.273
	2.00	0.188	0.490	0.275
	4.00	0.312	0.520	0.368
Morning 3	1.25	1.560	5.685	34.753
	1.50	2.118	9.055	86.479
	2.00	1.414	5.286	29.941
	4.00	1.561	6.874	49.689
Morning 4	1.25	0.564	0.810	0.974
	1.50	0.519	0.811	0.927
	2.00	0.548	0.826	0.983
	4.00	0.597	0.870	1.113
Evening 1	1.25	0.110	0.342	0.129
	1.50	0.088	0.340	0.123
	2.00	0.122	0.341	0.131
	4.00	0.276	0.358	0.204
Evening 2	1.25	0.245	0.540	0.352
	1.50	0.201	0.535	0.327
	2.00	0.242	0.538	0.348
	4.00	0.334	0.564	0.430
Evening 3	1.25	4.132	20.363	431.725
	1.50	2.472	11.093	129.165
	2.00	4.100	20.483	436.363
	4.00	2.661	12.597	165.765
Evening 4	1.25	0.552	0.820	0.977
	1.50	0.548	0.823	0.978
	2.00	0.568	0.829	1.010
	4.00	1.094	2.650	8.219

Table 5.2: Estimating treatment effect when modeling with $Z(t)$, $K = 100$ simulations. TBE = True Beneficial Effect.

Sample	TBE $= e^{-\beta}$	$EB[\hat{\beta}_2]$	$ESE[\hat{\beta}_2]$	$EMSE[\hat{\beta}_2]$
Morning 1	1.25	0.131	0.392	0.171
	1.50	0.139	0.396	0.176
	2.00	0.195	0.403	0.200
	4.00	0.293	0.429	0.270
Morning 2	1.25	0.274	0.591	0.424
	1.50	0.313	0.595	0.452
	2.00	0.279	0.609	0.449
	4.00	0.377	0.653	0.569
Morning 3	1.25	0.980	1.128	2.233
	1.50	1.066	1.162	2.487
	2.00	1.223	1.197	2.929
	4.00	1.350	1.344	3.629
Morning 4	1.25	0.757	0.901	1.385
	1.50	0.663	0.906	1.260
	2.00	0.874	0.947	1.661
	4.00	0.889	1.026	1.843
Evening 1	1.25	0.156	0.389	0.176
	1.50	0.155	0.393	0.178
	2.00	0.210	0.401	0.205
	4.00	0.308	0.429	0.279
Evening 2	1.25	0.290	0.637	0.490
	1.50	0.313	0.641	0.509
	2.00	0.350	0.659	0.557
	4.00	0.456	0.705	0.705
Evening 3	1.25	0.995	1.091	2.180
	1.50	0.948	1.102	2.113
	2.00	1.178	1.150	2.501
	4.00	1.707	1.292	4.583
Evening 4	1.25	0.726	0.913	1.361
	1.50	0.717	0.927	1.373
	2.00	0.707	0.947	1.397
	4.00	0.905	1.028	1.876

Table 5.3: Estimating treatment effect when modeling with Z, $K = 100$ simulations. TBE = True Beneficial Effect.

Sample	TBE $= e^{-\beta}$	$EB[\hat{\beta}_1] < EB[\hat{\beta}_2]$	$\beta \in 95\%$ CI$_{\hat{\beta}_1}$	$\frac{EMSE[\hat{\beta}_1]}{EMSE[\hat{\beta}_2]}$
Morning 1	1.25	0.62	0.88	0.702
	1.50	0.64	0.95	0.665
	2.00	0.69	0.82	0.640
	4.00	0.45	0.36	0.756
Morning 2	1.25	0.66	0.96	0.642
	1.50	0.75	0.96	0.604
	2.00	0.68	0.94	0.612
	4.00	0.54	0.81	0.647
Morning 3	1.25	0.58	0.98	15.563
	1.50	0.54	0.96	34.772
	2.00	0.73	0.99	10.222
	4.00	0.67	1.00	13.692
Morning 4	1.25	0.64	0.92	0.703
	1.50	0.60	0.98	0.736
	2.00	0.74	0.92	0.592
	4.00	0.62	0.97	0.604
Evening 1	1.25	0.65	0.92	0.733
	1.50	0.68	0.98	0.691
	2.00	0.76	0.84	0.639
	4.00	0.49	0.47	0.731
Evening 2	1.25	0.56	0.96	0.718
	1.50	0.62	0.98	0.642
	2.00	0.67	0.94	0.625
	4.00	0.62	0.88	0.610
Evening 3	1.25	0.49	0.95	198.039
	1.50	0.52	0.97	61.129
	2.00	0.59	0.98	174.475
	4.00	0.71	0.99	36.170
Evening 4	1.25	0.61	0.97	0.718
	1.50	0.61	0.97	0.712
	2.00	0.62	0.98	0.723
	4.00	0.61	1.00	4.381

Table 5.4: Comparing model with $Z(t)$ and model with Z, $K = 100$ simulations. TBE = True Beneficial Effect.

lationship between the event times and compliance, we have shown that our $C_p(t)$ compliance variable $Z(t)$ has more accuracy when estimating the true beneficial effect of a treatment than the usual compliance rate Z, and thus $\hat{\beta}_1$ is more efficient (in terms of estimator optimality) than $\hat{\beta}_2$ in predicting the rate of the clinical events. It is important to note that these results are consistent among all samples that were determined to have the same distribution in Chapter 4, with a sample size varying from $N = 27$ to $N = 906$. It would then be reasonable to assume that these results can be generalized to populations where the distribution of compliance (as described in section 4.1.2) follows a type of bimodal distribution.

5.4 Points of Discussion

Some remarks about the results presented in this Chapter. First, we can notice that there exists a relationship between compliance and the estimators. Low compliance implies low coverage of $Z(t)$ which in turn implies that we require a large sample size to accurately estimate the true beneficial therapeutic effect of the treatment. Next, it is worth mentioning that the reason why we can calculate the bias is because we know the real values of beta in the simulations. Even if we (supposedly) start with an unbiased estimator, we see that the estimators in the simulations are still biased. It is understood that compliance is not considered as an additional source of noise, unlike a standard missing covariate, since it is inherently included in a treatment variable. This means that when estimating the treatment effect, compliance may not contribute extra variability of the estimator (i.e. $Z(t) = 0$ implies no treatment effect), except for the non-additive effect of varying compliance rates within samples. This essentially means that the advantage of having compliance data is to reduce the bias of the estimated treatment effect.

In this scenario, the TBE estimator will always be biased if there is noncompliance since $\hat{\beta}$ in our model estimates the average treatment effect over the average

time of compliance. This means that $EB[\hat{\beta}_1]$ increases as the space where $Z(t) = 0$ increases. In reality, a linear model could essentially eliminate this bias when taking compliance into consideration (assuming that the design matrix is well-behaved, or invertible), since a compliance variable would replace the treatment variable, and thus adjusting the TBE estimation. A proportional hazards model would in theory be able to reduce the bias, but the nature of the estimation is difficult to comprehend, especially with a time-dependent covariate. In fact, it is still unclear if the starting estimator is even unbiased, since the estimation is done by partial maximum likelihood, as shown in section 3.5.1.

It is important to realize that repeating these analyses using simulated dose administration times can theoretically be done, although the complexity of recreating human behaviour through simulations is a tedious task that requires much more time than what was available for this thesis. We would then be able to provide a better description of the relationship between the accuracy of the estimator, the TBE and the required sample size in order to obtain the desired power in the analyses.

Finally, note that the analysis in this Chapter could have also been done using a fixed compliance rate *by package*. In other words, instead of using $Z(t)$, we could have had longitudinal data where each patient has a compliance rate for a given package, resulting in 4 observations per patient in a scenario where time would be defined as a categorical variable. We can then determine whether there is a pattern between different time points within a patient's treatment regimen. The ECM devices can then be used to calculate the compliance rate at each time point.

We can thus see that overall, having compliance data in general (it is not restricted to ECMs. However, the monitors are as of yet the best way to capture compliance) is a great source of information when estimating treatment effects. What can be retained from the discussion above is that using compliance data wisely generally yields a reduction in bias, and not necessarily a reduction in variance. In most cases, this bias-variance tradeoff is advantageous in the estimation when looking at

the EMSE, as long as we have a "large enough" sample size and a "high enough" compliance. If this is not the case, then the bias reduction in the estimation will not compensate for the variance part of the EMSE. It is also worth mentioning that if patients are perfectly compliant, then $Z(t) = Z \ \forall \ t$. The ECM devices would then guarantee the quality of the statistical analyses, since with perfect compliance there is no longer variability due to varying compliance rates within samples, and thus compliance is no longer measured as a differential effect.

Chapter 6

Conclusion

This thesis presents some new approaches to analyzing and utilizing information provided by electronic compliance monitors of patients following a treatment regimen. The data gathered by ECM devices can not only provide us, under the instant administration assumption, with the amount of doses taken correctly in accordance to the prescribed treatment schedule, but also allows us to determine approximately the time at which each dose is administered. Previous use of ECM data has been of a discrete and fixed-in-time nature. That is, the time at which the dose is taken is only to determine if the dose is taken in the allowed therapeutic window or not, yielding a binary compliance variable, which is then averaged over the dosage regimen, resulting in a compliance proportion. We demonstrated that such an approach can be suboptimal and that one can address the time-dependent properties of adherence to treatment in various ways in statistical analyses.

The survival analysis of compliance data found in chapter 4 was exploratory in nature, and to our knowledge no such analysis is yet to be found in the literature. It did provide us with some insight on the behaviour of patients' compliance with respect to the time of their first noncompliant doses. It also allowed us to conclude that the medication-taking habits of individuals in the study provided in the morning

are different than in the evening, which led to analyzing both doses separately.

An important contribution of this thesis is the proposal of an alternate, more efficient way to analyze the available information provided by electronic compliance monitors of patients following a treatment regimen. By fitting a one-compartment, multiple oral dose pharmacokinetic model, we can calculate the concentration vs time profiles of each patient with respect to their (approximated) dose administration times. Next, after determining the minimum effective concentration of the therapeutic drug, we can estimate, using pre-specified pharmacokinetic parameters, during which time intervals the concentration falls below the MEC. During these intervals, the treatment is considered to have no real therapeutic effect, and thus this is equivalent as to not taking any medication, or in other words, complete noncompliance with respect to the treatment. Using this as a time-dependent compliance variable, we were then able to compare its effectiveness against the effectiveness of the compliance rate in predicting simulated clinical events using discrete event simulation. In our simulations, this new pharmacokinetic compliance variable yielded a 25 to 40% reduction in EMSE compared to the compliance rate, depending on the true beneficial effect of the treatment. Using the ECM data by analyzing compliance as a time-dependent covariate in time-to-event outcomes is thus a more efficient way of utilizing the ECM technology in time-to-event statistical analyses.

It is of course possible to determine $z(t)$ without the use of pharmacokinetics given the information provided by the ECMs. One can approximate non-compliance intervals with the dosage times only by supposing that $z(t) = 0$ when more than 24 hours have passed without administrating a dose, and then $z(t)$ goes to 1 when the next dose it taken.

ECM devices have without a doubt an advantage over standard administrative data, where one can only find the number of doses obtained at the first prescription and every subsequent refill thereafter. With ECM, we can determine almost exactly the medication-intake behaviour of patients. This results in knowing exactly when

patients start taking their medication and when they stop. In other words, the main advantage of ECM is that the data can be used in many different ways in order to comprehend the temporal behaviour of compliance within a treatment regimen.

Some measuring drawbacks of this thesis include the potentially misleading information provided by the ECM devices. Although the instant administration assumption is not necessarily strict, it does assume that the doses released from the packages are taken immediately, and not kept to be taken at a later time. The other issue with the recorded times is when there are multiple dose releases at the same time, or only moments apart. There is no way for us to determine whether a patient actually administered all released doses, simply transferred the released pills to another compartment that is more suitable for them, or even to replace a released pill that becomes unusable. Some modeling drawbacks include the choice of the pharmacokinetic model and parameters. Although the literature suggests that a one-compartment model with first-order reaction rates can depict reality when taking oral doses, it is still unclear if multiple compartment models can better represent the ADME processes associated with drug intake. Furthermore, the pharmacokinetic parameters that were used for the modeling of the concentration vs time profiling were taken from a specific case-study, and therefore they probably cannot be utilized to model the C_p of the population as whole with respect to any drug. Since the parameters are drug- and patient-specific, the modeling of pharmacokinetic compliance would then represent in reality a much larger task than what is presented in this thesis.

Some further research interest would start by doing pharmacokinetic analyses to determine the drug- and patient-specific parameters of the medication when fitting the one-compartment model presented in this these in order to more accurately determine when the plasma concentration level is below the minimum effective concentration when creating the pharmacokinetic compliance variable, although this may be costly. Another would be to use a different pharmacokinetic model entirely to see

whether the results are comparable with the findings of this thesis. Another interesting addition would be to utilize ECM data for non-regular medication, such as analgesics (painkillers), in a survival analysis for migraines for example.

Additional research also includes the modeling of noncompliance in the manner of recurrent events, as explained in Cook and Lawless's *Statistical analysis of recurrent events*, [12]. One could model the noncompliance events (either extra doses or missed doses) provided by the ECM data as recurrent events over the treatment regimen. It would then be possible to simulate patients based on these models and adjust parameters for different levels of compliance.

Other research interest would also include modeling compliance as a response variable as a function of clinical covariates, as discussed in section 1.1.2. It is reasonable to think that certain patient level characteristics (e.g. education level, gender, level of stress) and certain drug level characteristics (e.g. side effects, method of administration) can affect compliance. In fact, a recent hand washing study showed a significant difference in the number of men who washed their hands compared to women [5]. Clearly taking a pill is generally easier than washing one's hands, however this study illustrates the fact that there might exist major differences in compliance with respect to a treatment within patient-specific characteristics. If we could better understand the factors that determine compliance, we could in turn better control it in further studies. One possibility is to target less compliant groups of individuals and develop strategies to ensure better compliance on their part.

Finally, one could also explore the role of treatment compliance in statistical analyses found in clinical trials using ECM devices where the primary endpoint is not a time-to-event, including exploratory, safety and dose tolerability analyses in phase I and phase II clinical trials, which would lead to an increase in validity of results.

Bibliography

[1] Acosta, F. J., Bosch, E., Sarmiento, G., Juanes, N., Caballero-Hidalgo, A., & Mayans, T. (2009). Evaluation of noncompliance in schizophrenia patients using electronic monitoring and its relationship to sociodemographic, clinical and psychopathological variables. *Schizophrenia Research*, **107**(2), 213-217.

[2] Ailinger, R. L., Black, P. L., & Lima-Garcia, N. (2008). Use of electronic monitoring in clinical nursing research. *Clinical Nursing Research*, **17**(2), 89-97.

[3] Armijo-Olivo, S., Warren, S., & Magee, D. (2009). Intention to treat analysis, compliance, drop-outs and how to deal with missing data in clinical research: a review. *Physical Therapy Reviews*, **14**(1), 36-49.

[4] Bangsberg, D. R., Hecht, F. M., Charlebois, E. D., Chesney, M., & Moss, A. (2001). Comparing objective measures of adherence to HIV antiretroviral therapy: Electronic medication monitors and unannounced pill counts. *AIDS and Behavior*, **5** (3), 275-281.

[5] Borchgrevink, C. P. , Cha J. & Kim, S. (2013). Hand washing practices in a college town environment, *Journal of Environmental Health*, **75**(8), 18-24.

[6] Bova, C. A., Fennie, K. P., Knafl, G. J., Dieckhaus, K. D., Watrous, E., & Williams, A. B. (2005). Use of electronic monitoring devices to measure an-

tiretroviral adherence: practical considerations. *AIDS and Behavior*, **9**(1), 103-110.

[7] Caro, J. J., Speckman, J. L., Salas, M., Raggio, G., & Jackson, J. D. (1999). Effect of initial drug choice on persistence with antihypertensive therapy: the importance of actual practice data. *Canadian Medical Association Journal*, **160**(1), 41-46.

[8] Caron, H. S. (1985). Compliance: the case for objective measurement, *Journal of Hypertension*, **3**(1), 11-17.

[9] Casella G., & Berger, R. L. (2002) *Statistical Inference* (Vol. 2). Pacific Grove, CA: Duxbury Press.

[10] Claxton, A. J., Cramer, J., & Pierce, C. (2001). A systematic review of the associations between dose regimens and medication compliance. *Clinical Therapeutics*, **23**(8), 1296-1310.

[11] Conrad, P. (1985). The meaning of medications: another look at compliance. *Social Science and Medicine*, **20**(1), 29-37.

[12] Cook, R. J. & Lawless, J. F. (2007). *The Statistical Analysis of Recurrent Events*, Springer.

[13] Cramer, J. A. et al. (1989). How Often Is Medication Taken as Prescribed: A Novel Assessment Technique, *Journal of the American Medical Association*, **261**(22), 3273-3277.

[14] Cramer, J. A., Scheyer, R. D. & Mattson, R. H. (1990). Compliance declines between clinic visits, *Archives of Internal Medicine*, **150** (7), 1509-1510.

[15] Cramer, J. A. & Spilker, B. (1991). *Patient Compliance in Medical Practice and Clinical Trials*, New York: Raven Press.

[16] Denhaerynck, K., Schfer-Keller, P., Young, J., Steiger, J., Bock, A., & De Geest, S. (2008). Examining assumptions regarding valid electronic monitoring of medication therapy: development of a validation framework and its application on a European sample of kidney transplant patients. *BMC Medical Research Methodology*, **8** (1), 5.

[17] Deschamps AE, De Geest S, Vandamme AM, Bobbaers H, Peetermans WE, Van Wijngaerden E. (2008). Diagnostic value of different adherence measures using electronic monitoring and virologic failure as reference standards. *AIDS Patient Care and STDs*, **22**(9), 735-743.

[18] Deschamps, A. E., Van Wijngaerden, E., Denhaerynck, K., De Geest, S., & Vandamme, A. M. (2006). Use of electronic monitoring induces a 40-day intervention effect in HIV patients. *Journal of Acquired Immune Deficiency Syndromes*, **43** (2), 247-248.

[19] Dhillon, S. & Kostrzewski, A.J. (2006). *Clinical Pharmacokinetics*, Pharmaceutical Press.

[20] Donovan, J. L. & Blake, D. R. (1992). Patient non-compliance: deviance or reasoned decision making?, *Social Science and Medicine*, **34**(5), 507-513.

[21] Dunbar-Jacob, J. & Mortimer-Stephens, M. K. (2001). Treatment adherence in chronic disease, *Journal of Clinical Epidemiology*, **54**(12), 57-60.

[22] Efron, B. & Feldman, D. (1991). Compliance as an Explanatory Variable in Clinical Trials, *Journal of the American Statistical Association*, **86**(413), 9-17.

[23] Fennie, K. P., Bova, C. A., & Williams, A. B. (2006). Adjusting and censoring electronic monitoring device data: implications for study outcomes. *Journal of Acquired Immune Deficiency Syndromes*, **43**, S88-S95.

[24] Fisher, L. D., Dixon, D. O., Herson, J., Frandowski, R. K., Hearron, M. S., & Peace, K. E. (1990). Intention to Treat in Clinical Trials. In K.E. Peace (Ed.), Statistical Issues in Drug Research and Development. New York: Marcel Dekker.

[25] Friedman, L. M., Furberg, C., & DeMets, D. L. (2010). *Fundamentals of Clinical Trials* (Vol. 4). New York: Springer.

[26] Girard, P., Sheiner, L.B., Kastrissios, H. & Blaschke, T.F. (1996). Do We Need Full Compliance Data for Population Pharmacokinetic Analysis?, *Journal of Pharmacokinetics and Biopharmaceutics*, **24**(3), 265-282.

[27] Glanz, K., Fiel, S. B., Swartz, M. A. & Francis, M. E. (1984). Compliance With an Experimental Drug Regimen for Treatment of Asthma: Its Magnitude, Importance, and Correlates, *Journal of Chronic Diseases*, **37**(11), 815-824.

[28] Godfrey, K. (1983). **Compartment Models and Their Application**, Academic Press.

[29] Griffith, S. (1990). A review of the factors associated with compliance and the taking of prescribed medicines, *British Journal of General Practice*, **40**(332), 114-116.

[30] Haberer, J. E., Kahane, J., Kigozi, I., Emenyonu, N., Hunt, P., Martin, J., & Bangsberg, D. R. (2010). Real-time adherence monitoring for HIV antiretroviral therapy. *AIDS and Behavior*, **14**(6), 1340-1346.

[31] Harrison, T. & Ansell, J. (2002). Customer retention in the insurance industry: using survival analysis to predict cross-selling opportunities. *Journal of Financial Services Marketing*, **6**(3), 229-239.

[32] Haynes, R. B., Taylor, D. W. & Sackett, D. L. (1974). *Compliance in Health Care*, The John Hopkins University Press.

[33] Horwitz, R. I., Viscoli, C. M., Donaldson, R. M., Murray, C. J., Ransohoff, D. F., Horwitz, R. I., ... & Sindelar, J. (1990). Treatment adherence and risk of death after a myocardial infarction. *The Lancet*, **336**(8714), 542-545.

[34] Hosmer, D. W., Lemeshow, S. (1999). *Applied Survival Analysis: Regression Modelling of Time to Event Data.*, New York: Wiley.

[35] Huybrechts, K. F., Ishak, K. J. & Caro, J. J. (2006). Assessment of compliance with osteoporosis treatment and its consequences in a managed care population, *Bone*, **38**(6), 922-928.

[36] Jay, S., Litt, I. F. & Durant, R. H. (1984). Compliance with Therapeutic Regimens, *Journal of Adolescent Health Care*, **5**(2), 124-136.

[37] Jo, B. (2002). Statistical power in randomized intervention studies with noncompliance. *Psychological Methods*, **7**(2), 178.

[38] Montori, V. M., & Guyatt, G. H. (2001). Intention-to-treat principle. *Canadian Medical Association Journal*, **165**(10), 1339-1341.

[39] Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, **53**(282), 457-481.

[40] Kleinke, J. D. (2001). The price of progress: prescription drugs in the health care market, *Health Aff (Millwood)*, **20**(5), 43-60.

[41] Kruskal, W. H. & Wallis, W. A. (1952). Use of Ranks in One-Criterion Variance Analysis, *Journal of the American Statistical Association*, **47**(260), 583-621.

[42] Lerner, B. H. (2010). From Careless Consumptives to Recalcitrant Patients: The Historical Construction of Noncompliance, *Journal of Social Science & Medicine*, **45**(9), 1423-1431.

[43] Manchikanti, L. (2007). National drug control policy and prescription drug abuse: facts and fallacies. *Pain Physician*, **10**(3), 399.

[44] Manchikanti, L., Atluri, S., Trescot, A. M., & Giordano, J. (2008). Monitoring opioid adherence in chronic pain patients: tools, techniques, and utility. *Pain Physician*, **11**(2 Suppl), S155-80.

[45] Meissner, K., Distel, H., & Mitzdorf, U. (2007). Evidence for placebo effects on physical but not on biochemical outcome parameters: a review of clinical trials. *BMC Medicine*, **5**(1), 3.

[46] Melnikow, J. & Kiefe, C. (1994). Patient Compliance and Medical Research: Issues in Methodology, *Journal of General Internal Medicine*, **9**(2), 96-105.

[47] Miller, N. H. (1997). Compliance with treatment regimens in chronic asymptomatic diseases, *The American Journal of Medicine*, **102**(2), 43-49.

[48] Moeschberger, M. L., & Klein, J. (2003). *Survival analysis: Techniques for Censored and Truncated Data: Statistics for Biology and Health*. Springer.

[49] Neal, D. T., Wood, W. & Quinn, J. M. (2006). Habits - A repeat performance. *Current Directions in Psychological Science*, **15**(4), 198-202.

[50] Norman, J. (1992). One-compartment kinetics. *British Journal of Anaesthesia*, **69**(4), 387-396.

[51] Passik, S. D., Kirsh, K. L., & Casper, D. (2008). Addiction-Related Assessment Tools and Pain Management: Instruments for Screening, Treatment Planning, and Monitoring Compliance. *Pain Medicine*, **9**(s2), S145-S166.

[52] Pocock, S. J., & Abdalla, M. (1998). The hope and the hazards of using compliance data in randomized controlled trials. *Statistics in Medicine*, **17**(3), 303-317.

[53] Ramlau-Hansen, H. (1983). Smoothing counting process intensities by means of kernel functions. *The Annals of Statistics*, 453-466.

[54] Reddy, M., Yang, R. S., Andersen, M. E., & Clewell III, H. J. (2005). *Physiologically Based Pharmacokinetic Modeling: Science and Applications*. John Wiley & Sons.

[55] Richardson, P. H. (1994). Placebo effects in pain management. *Pain Reviews*, **1**(1), 15-32.

[56] Riekert, K. A., & Rand, C. S. (2002). Electronic monitoring of medication adherence: when is high-tech best?. *Journal of Clinical Psychology in Medical Settings*, **9**(1), 25-34.

[57] Rodriguez, G. (2005). *Non-parametric Estimation in Survival Models*, Princeton Press.

[58] Ross, S. (1997). *Simulation (Statistical Modeling and Decision Science)*, Harcourt Academic Press.

[59] Rowe, P. (2012). *Pharmacokinetics*, Philip Rowe & Ventus Publishing ApS.

[60] Schaubel, D. E., & Cai, J. (2004). Regression methods for gap time hazard functions of sequentially ordered multivariate failure time data. *Biometrika*, **91**(2), 291-303.

[61] Simoni J.M., Kurth A.E., Pearson C.R., Pantalone D.W., Merrill J.O., & Frick P.A. (2006). Self-report measures of antiretroviral therapy adherence: a review with recommendations for HIV research and clinical management. *AIDS and Behavior*, **10**(3), 227-245.

[62] Sokol, M. C. et al. (2005). Impact of Medication Adherence on Hospitalization Risk and Healthcare Cost. *Medical Care*, **43**(6), 521-530.

[63] Steinkopf, L. (2012). Enhancing drug compliance and the placebo effect by raising subjective expectations. *Medical Hypotheses*, **79**(5), 698-700.

[64] Su, J., Sun, P., Li, X. & Hartford, A. H. (2009).*Fitting Compartment Models to Multiple Dose Pharmacokinetic Data using SAS PROC NLMIXED*, Merck & Co., Inc.

[65] Tillmann, H.C., Sharpe, N., Sponer, G. & Wehling, M. (2001). Does intention-to-treat analysis answer all questions in long-term mortality trials? Considerations on the basis of the ANZ trial. *International Journal of Clinical Pharmacology and Therapeutics*, **39**, 205-212.

[66] Vermeire, E. et al. (2001). Patient adherence to treatment: three decades of research. A comprehensive review. *Journal of Clinical Pharmacy and Therapeutics*, **26**(5), 331-342.

[67] Vrijens, B. et al. (2008). Adherence to prescribed antihypertensive drug treatments: longitudinal study of electronically compiled dosing histories, *BMJ*, **336**(7653), 1114-1117.

[68] Wang, J. L. (2005). Smoothing hazard rates. *Encyclopedia of Biostatistics*, 4.

[69] Wang, M. C., & Chang, S. H. (1999). Nonparametric estimation of a recurrent survival function. *Journal of the American Statistical Association*, **94**(445), 146-153.

[70] Wang, W., Husan, F. & Chow, S-C. (1996). The Impact of Patient Compliance on Drug Concentration Profile in Multiple Doses, *Statistics in Medicine*, **15**(6), 659-669.

[71] Wang, W & Ouyang, S.P. (1998). The Formulation of the Principle of Superposition in the Presence of Non-Compliance and Its Applications in Multiple Dose

Pharmacokinetics, *Journal of Pharmacokinetics and Biopharmaceutics*, **26**(4), 457-469.

[72] Williams, A. B., Amico, K. R., Bova, C., & Womack, J. A. (2013). A proposal for quality standards for measuring medication adherence in research. *AIDS and Behavior*, **17**(1), 284-297.

[73] Wilson, A. (n. d.). 21 CFR Part 11. *Rx for the third Millennium*. Retrieved December 2, 2014, from http://www.informationmediary.com/news-a-media/compliance-papers.

[74] Sabate, E. (Ed.). (2003). *Adherence to long-term therapies: Evidence for action*. World Health Organization.

[75] Xue, X., Xie, X. et al. (2013). Testing the proportional hazards assumption in case-cohort analysis. *BMC Medical Research Methodology*, **13**(1), 88.