Preface

Despite its poor statistical properties, the 3+3 algorithm remains the most commonly used dose finding method in phase I clinical trials today. However, as clinicians begin to realize the important role of dose finding in the drug development process, there is an increasing openness to "novel" methods proposed in the past two decades. In particular, the continual reassessment method (CRM) and its variations have drawn much attention in the medical community. To ride on this momentum and overcome the status quo in the phase I practice, it is critical for us (statisticians) to be able to design a trial using the CRM in a timely and reproducible manner. This is the impetus to writing a detailed exposition on the calibration of the CRM for applied statisticians who need to deal with dose finding in phase I trials while having many other duties to attend to.

A natural approach to such a writing project is to write a *how-to* book. By the time I started this book project in the summer of 2008, I had helped design half a dozen CRM trials (three of which are included as examples in this book). In retrospect, I found some general patterns of how I calibrated the CRM parameters in these trials. These patterns, characterized collectively as a trial-and-error approach in Chapter 7, worked well in the sense that they gave reasonable operating characteristics to a design. However, it was time-consuming (weeks of simulation) and would require an intimate understanding of the CRM (I wrote a PhD dissertation on the CRM). I realized that some automation and step-by-step guidelines in this calibration process would be crucial and appreciated if the CRM was to be used on a regular basis by a wide group of statisticians. Chapters 7–10 try to address this need by breaking a CRM design into a list of design parameters, each of which is to be calibrated in a prescribed manner.

Despite my pragmatic approach, I hope this book is not only a cookbook. I intend to provide a full coverage of the CRM. This book includes a comprehensive review of the CRM (Chapter 3) and elaborate properties of the CRM (Chapters 5 and 6). While this book is based on my previous publications on the CRM, I have introduced new material so as to present the CRM under a unified framework (Chapter 4). These chapters serve as the theoretical foundation of the calibration techniques presented in the later chapters. I also reflect on what *not* to do with the CRM (Chapter 12) and when *not* to use the CRM (Chapter 13). From a practical viewpoint, these *not-to* chapters are as important as, if not more important than, the *how-to* chapters, because they avoid abuses and pitfalls in applying the CRM. I believe that using the CRM in a wrong way or in the wrong trial is no better, or arguably worse, than falling back to the 3+3 algorithm. The time-to-event aspect of the toxicity endpoint has been a

recurring concern in my previous CRM trials, and so is included as an extension of the CRM (Chapter 11). All in all, while this is not intended to be a cookbook, the inclusion of materials is based on their practical relevance.

This book does not cover dose finding in all possible clinical settings. In fact, it has a singular focus on the simplest and the most common phase I trial setting, where the study endpoint is defined as a binary outcome and the subjects are assumed to come from a homogeneous population. I make no mention of the concerns with multiple toxicity and the gradation of severe toxicities. The topic of individualized dosing is omitted. While some basic ideas of dose finding using both efficacy and toxicity are outlined in Chapter 13, the discussion is brief and does not do full justice to this fast-growing area. All these are important topics in which I am intellectually interested. Their omission, however, is mainly due to my limited *practical* experience in dealing with these "nonstandard" situations in real dose finding studies; dealing with these issues simply from a *methodological* and *theoretical* viewpoint does not fit my intent of writing a practical book (although I think such a book is interesting in its own right and hope someone more qualified than I will deliver it). I do have a word or two to add from a methodological and theoretical viewpoint here, if not already alluded to in the book's final chapter (Section 14.4, to be precise). First, a complete theoretical framework is crucial for these nonstandard methods to be successfully translated into actual practice. In this book, I try to explicate possible pathological behaviors (e.g., incoherence and rigidity) of some CRM modifications for the simplest setting; it is reasonable to infer that these pathologies will multiply for methods more complex than the CRM for the more complicated clinical settings. Solid theoretical investigation will help us navigate the potential pitfalls. I also hope the theoretical framework developed in this book for the simplest case will prove useful when extended to the complicated settings. Second, and more specifically, I think stochastic approximation offers partial solutions (albeit mostly theoretical) to many of these nonstandard dose finding settings. This is why I close this book with a chapter where I try to connect and compare the CRM with the rich stochastic approximation literature.

The last points I just made give a hint about my methodological and theoretical interests. I hope that this book will in some way simulate research in the CRM and general dose finding methods, despite its practical nature. As I try to present the CRM and the dose finding criteria at a rigorous level, and to cover the CRM literature as comprehensively as possible, I also hope this book can serve as an introduction for those interested in doing research in this area. I taught a course on sequential experimentation at Columbia University from an early unpublished version of this book. This final manuscript is, in turn, adapted from the course notes, and is suitable for use in a course on sequential experimentation or clinical trials.

There are several statistics books on dose finding. The two most popular ones are the edited volumes by Chevret [26] and by Ting [105]. Both give surveys of dose finding methods and are good introductions to the dose finding literature. By comparison, this book is a single-authored work on a specific dose finding method, which I think is necessary if we are to get down to the nuts and bolts of the method.

By writing a book on the CRM, I do not imply that it is the best method out there.

In fact, for the dose finding objective considered here, it is unlikely that there is one method that is best or optimal in a *uniform* sense. While some methods may work best under certain scenarios according to some criterion, the others are optimal under a different criterion. There have been numerous proposals in the last two decades. These proposals can be good alternatives against the 3+3 algorithm as long as they are calibrated properly. And, the CRM is one of these methods. Furthermore, the CRM has been worked out and discussed in the statistical and medical literature so extensively that I believe we are getting close to translating this method into practice. This book hopefully will be a catalyst in this translational process.

I owe a debt of gratitude to Tom Cook, Bin Cheng, and an anonymous reviewer who have been generous with their time and given detailed comments on earlier versions of the book. I am grateful for Jimmy Duong for his help to maintain the R package 'dfcrm' (a companion software with this book). I would also like to thank Rick Chappell who introduced me to the CRM and clinical trials when I was a student at University of Wisconsin–Madison. This book would not be possible without his mentoring. Finally, my most heartfelt thanks go to my wife, Amy, for her support and enthusiasm during this writing process.

New York
October 2010

Contents

Part I

Fundamentals

Chapter 1

Introduction

The clinical development of a new drug or a new treatment proceeds through three phases of testing in human subjects. Phase I trials are small studies that evaluate safety and identify a safe dose range of the treatment. Once a dose range is chosen, its therapeutic efficacy will be examined in a phase II trial. Regimens that are shown promising in phase II trials will be moved to multi-institutional phase III clinical trials for randomized comparison to standard treatments. The ultimate goal of this entire process is to translate promising discoveries in the laboratory into new medical procedures that can be used in the general clinical settings. This division of clinical trials, however, may give an oversimplified picture of the actual drug development process. Often, several phase I-II trial sequels of a drug, possibly with minor variations in the treatment schedule and patient populations, are needed before a phase III trial is warranted. This process is necessarily iterative rather than linear, as the phase I-II-III paradigm appears to suggest. In addition, the taxonomy of trials is not universal across disciplines, and may include finer divisions such as phase IA, IB, IIA, and IIB. The recent trend to combine phases of trials, the so-called combined phase I/II trials and seamless phase II/III trials, renders further refinement of the drug development process.

This having been said, the phase I-II-III paradigm provides a conceptual framework for in-depth study of statistical methods. The subject matter of this book is dose finding using the continual reassessment method (CRM). The CRM [78] is among the first model-based designs for phase I cancer trials in which toxicity is the primary study endpoint. The role of toxicity in early-phase cancer trials had long been a subject for discussion in the medical literature [93, 85]. In particular, for cytotoxic drugs, toxicity serves as evidence that the drug has reached a level that does harm not only to the cancer cells but also to a patient's normal organs. In other words, a therapeutic dose is expected to cause a significant amount of severe but reversible toxicities in the cancer patient population. Therefore, a primary goal of phase I cancer trials is to identify the so-called maximum tolerated dose (MTD). For other disorders such as acute stroke and HIV, identifying the MTD is also a primary objective of early-phase safety studies (usually called phase IB trials). In addition, dose finding is important in phase II proof-of-concept trials where the goal is to identify a dose range with demonstrated biological activity. This objective is usually achieved through the estimation of the minimum effective dose (MED) [106, 27]. From a statistical viewpoint,

the MTD in safety studies and the MED in efficacy studies can be formulated in an analogous way. Therefore, this book is relevant to the design of phase I and II dose finding trials. Under the modernized paradigm, the dose finding principles discussed here also address the design issues in the combined phase I/II trials, in which both the safety and the efficacy endpoints are considered as co-primary (cf. Section 13.3).

Another advantage of dividing the drug development process into phases is that by doing so, we can set a clear and manageable benchmark to achieve in a particular study. This entails a clearly defined set of study endpoints and an interpretable study objective. Since clinical trials are conducted in human subjects, each benchmark is to be reached within certain ethical constraints. In particular, in phase I dose finding studies, randomization is not entirely appropriate because it may expose subjects to excessively high and toxic doses without sufficiently testing the lower doses. (Some would also argue randomization exposes subjects to low and inefficacious doses, although this aspect is apparently not as alarming.)

We illustrate these points using a bortezomib dose finding trial [62]. Bortezomib is a proteasome inhibitor with proven activity in lymphoma. In the trial, bortezomib was given in combination with the standard chemotherapy as a first-line treatment for patients with diffuse large B cell or mantle cell non-Hodgkin's lymphoma. Each patient would receive up to six 21-day cycles of the treatment combination. Table 1.1 describes the five dose schedules of bortezomib tested in the trial. The primary safety concerns related to bortezomib were neuropathy, low platelet count, and symptomatic non-neurologic or non-hematologic toxicity. Toxicity was graded according to the National Cancer Institute Common Terminology Criteria for Adverse Events [71], with grade 3 or higher defined as *dose limiting*. Generally, a grade 3 toxicity is severe but can be resolved by symptomatic treatment, whereas a grade 4 toxicity is irreversible; toxic death due to the treatment is invariably defined as grade 5. The primary endpoint of each patient was the indicator of whether any dose-limiting toxicity (DLT) was experienced at any time during the six cycles. The objective of the trial was to determine the MTD, defined as a dose associated with a 25% DLT rate. Table 1.1 gives the number of patients and the number of DLTs per dose in the bortezomib trial. The data show strong evidence that the highest dose is adequately safe: we pool the observations in dose levels 4 and 5 by assuming an increasing dose–toxicity relationship; based on 1 DLT out of 16 patients, we obtain a 95% confidence upper bound of 0.26 for the DLT probability.

Table 1.1 The bortezomib trial [62]: dose schedules of bortezomib, sample size (n), and the number of DLT (z) at each dose

Level	Dose and schedule within cycle	n	z
1	0.7 mg/m^2 on day 1 of each cycle	0	0
2	0.7 mg/m^2 on days 1 and 8 of each cycle	0	0
3	0.7 mg/m^2 on days 1 and 4 of each cycle	4	0
4	1.0 mg/m^2 on days 1 and 4 of each cycle	9	1
5	1.3 mg/m^2 on days 1 and 4 of each cycle	7	0

While simple analyses are usually adequate to address the primary scientific questions in a phase I study, the summary statistics in Table 1.1 ignore how the data were collected. Figure 1.1 shows the dose assignments of the trial in chronological order. The trial started at level 3, a dose schedule that the investigators believed to be safe to treat patients. Escalation to the next higher dose occurred after four patients had been followed for several weeks without signs of toxicity, and another escalation took place after three following patients. Shortly after the eighth patient entered the trial at the highest dose, patient 7 at dose level 4 experienced a DLT, thus leading to a deescalation for the ninth patient. Subsequent patients were enrolled in a staggered fashion, allowing months to pass before reescalating to the highest level. A central feature of this dose assignment scheme is its outcome adaptiveness. Specifically, in the bortezomib trial, the dose assignments were made in accordance with the time-to-event continual reassessment method (TITE-CRM), an extension of the CRM to be discussed in Chapter 11. For ethical reasons, most dose finding trials are conducted in an outcome-adaptive manner, so that the dose assignment of the current patient depends on those of the previous patients. As such, the focus of this book is the *design* (as opposed to analysis) of a dose finding study using the CRM and its variants.

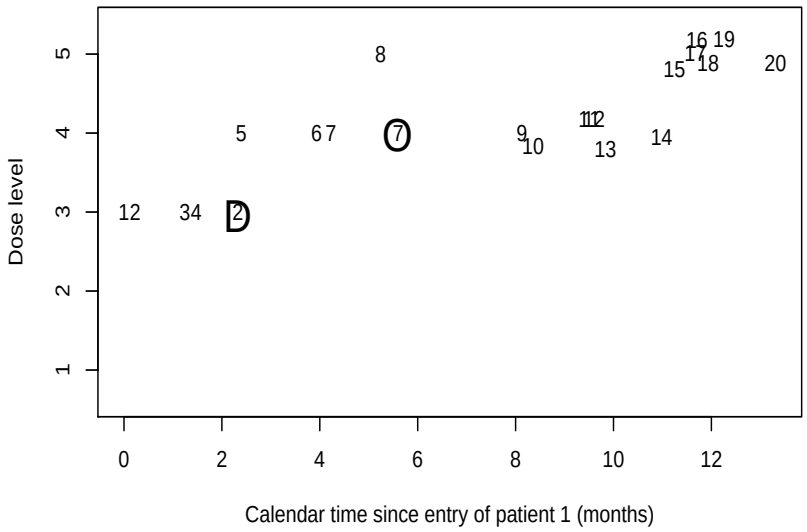

Figure 1.1 Dose assignments in the bortezomib trial. Each number indicates a patient: An unmarked number represents the patient's entry time; a number marked with "O" indicates the time when a DLT occurs, and "D" indicates the time of dropout. Vertical positions of some numbers are jittered for clarification.

This book is organized into three parts. Part I (Chapters 2–6) contains the background and introductory material of the CRM. Specifically, Chapter 2 provides the clinical background, outlines the problem of dose finding in the context of several real trial examples, and reviews the dose finding literature. Chapter 3 introduces the basic approach of the CRM and presents its major modifications. The method will be developed along with a description of an R package 'dfcrm'. Chapter 4 presents a unified framework for dose–toxicity models used in the CRM. Chapters 5 and 6, respectively, discuss the theoretical and empirical properties of the CRM. The objective of Part I is for the readers to develop a basic understanding of the CRM and be able to implement the method using a simple R code. Readers familiar with the basic CRM methodology are also encouraged to review the materials, as they are reorganized and presented in a unified framework in this book.

Part II (Chapters 7–10) details the calibration process of the CRM based on the notation and the theory introduced in Part I. Chapter 7 introduces a system of design parameters involved in the CRM, and classifies them into two categories: *clinical parameters* and *model parameters*. The subsequent chapters then present fine-tuning techniques of the model parameters: the initial guesses of the toxicity probabilities (Chapter 8), the prior distribution of the model parameter (Chapter 9), and the initial design of a two-stage CRM (Chapter 10). The objective of Part II is for the readers to develop the ability to design a "good" CRM trial within a reasonable timeline.

Part III (Chapters 11–14) contains a variety of advanced topics related to the CRM. Chapter 11 presents the TITE-CRM to deal with situations in which the toxicity outcome is defined with respect to a nontrivial duration. Chapter 12 gives a critical review of CRM using multiparameter models. Chapter 13 considers situations where the CRM is an inappropriate design, and puts forward some alternatives. Chapter 14 connects the CRM and modern dose finding trials to the large literature of stochastic approximation. The objective of Part III is to stimulate further research in the CRM and general dose finding methodology.

The materials in this book are presented at a level that requires college algebra and some basic calculus concepts. Sections marked with "†" in the table of contents contain technical details that may be skipped without affecting the reading of the other chapters. Exposition in the book will be supplemented by illustrations of the usage of R functions in the 'dfcrm' package. While some basic knowledge of R will enhance the reading experience, proficiency in R is not required. Interested readers can find out more information about R from the Comprehensive R Archive Network (CRAN) [83] at

`http://www.r-project.org`.