BIOMETRICS 71, 1196–1199
December 2015
DOI: 10.1002/biom.12440

BOOK REVIEWS

EDITOR:
TAESUNG PARK

ALESSANDRO B. ANTOGNINI AND ALESSANDRA GIOVAGNOLI, **Adaptive Designs for Sequential Treatment Allocation.** Boca Raton, FL: Chapman and Hall/CRC

In medical and clinical research, innovative new trial designs are on demand for obvious reasons, especially when the attrition rate is large. Attrition rate is on the rise and is a real practical issue of concern to most, if not all, clinical researchers. As one such innovative approach to trial designs, there is a growing interest in the use of adaptive designs especially in pharmaceutical research where one desires to use the information gathered along the way. Adaptive or sequential designs are often confused with sequential analysis. While these are highly related, however, the analysis phase of an experiment is dealt with separately from the design phase. This book presents a mathematical foundation on adaptive design theory and discusses issues with designing sequential-randomized experiments involving two or more treatments. Such a design approach readily adapts the present treatment administration to past history. The authors stress that their primary intention of the book is in making it a research reference book and as such, they review and reorganize the existing results with a particular attention on mathematical foundation, to complement other references available with little overlap. Further, they stress that “it is not a book on clinical trials” and they chose “to dwell on general aspects more than on individual trials.”

To achieve these goals, they begin by introducing various fundamental mathematical and statistical terminologies and preliminary results. In their first chapter, they motivate inferential considerations and give new conditions for the convergence of a sequential experiment to a given target allocation of the treatments, with asymptotic inference playing the central role. Then, the remainder of the book is consisted of five chapters, each discussing (i) specific randomization procedures that are functions of the past allocations, (ii) those that depend on the responses, (iii) those that depend on the covariates, (iv) compromised step-by-step multipurpose adaptive designs, and (v) constrained and combined optimality of multipurpose adaptive designs.

Chapter 2 illustrates designs whose assignments rule is based on past treatment experiences only. Included are biased coin and urn designs and some extensions. Chapter 3 brings the past data in its adaptive strategy as well, such as sequential maximum likelihood and various types of doubly adaptive designs, and Up-and-Down designs for binary, binomial, and normal responses. The book also covers multiple objective adaptive experiments involving utilitarian choices and ethical issues. Chapter 4 introduces the step-by-step procedure, compromising among different objectives. Included are play-the-winner, drop-the-loser type, link-based designs, and the compound probability approach, followed by asymptotic inference for multipurpose designs and some extensions to several treatments. Chapter 5 discusses an overall strategy optimizing a compound multiple objective optimal targets. Both constrained and combined approaches are discussed along with some extension to the case of several treatments. These are updated to include covariates in Chapter 6.

While the authors primarily used a frequentist approach in developing the book, I like the fact that they also included Bayesian adaptive designs in the appendix. The appendix also presents basic classical optimal design theory.

In summary, the book illustrates theoretical properties of adaptive designs so that researchers can choose the best design for the experiment, covering impressively diverse approaches. There is no short answer to what is the best design and the authors illustrated how the choice depends on the objectives for the experiment. I find the book quite appealing in that the authors believed on the theoretical properties of the designs and mathematical foundations more than how and what of adaptive designs, consistent with their aim for the book as I alluded to above. The book is not very mathematical, and uses minimal mathematical sophistication to get through most of the content. However, while it is rather easy for mathematically prepared readers to follow, it is not quite challenging enough for these readers. On the other hand, it may be too mathematical for clinicians. Nevertheless, I think it would be highly useful as a reference book in a graduate-level course on designs with nearly exhaustive approaches to adaptive design construction.

K.C. Carriere
Department of Mathematical and Statistical Sciences
University of Alberta
Edmonton, AB, Canada
kccarrie@ualberta.ca

RONALD B. GESKUS, **Data Analysis with Competing Risks and Intermediate States**. New York: Chapman and Hall/CRC

This book focused on analyzing data with competing risks and intermediate states. It has advantages over previous works dealt with the two topics. Among them, the last four sections of each chapter consist of the following: a broader discussion on the concepts that have been introduced, exercises for reflecting on the concepts including a few technical exercises, R-based implementation of the methods covered in each chapter along with a brief explanation of options in SAS and Stata, and computer practicals on real datasets asked to practice the concepts in R. This book is divided into five chapters and an epilogue. The first five chapters can also be grouped into three: an introduction and statistical inference in the absence or presence of covariates.

The first chapter introduced the examples that will be used in this book, the common forms of incomplete information such as right censoring and left truncation, the concepts of hazard and cumulative probability with an extension to the competing risks settings, and a more formal overview of the concepts and results from classical survival analysis.

The second and third chapters dealt with nonparametric estimation of the relevant quantities in competing risks and multi-state models, respectively. In Chapter 2, two hazards, namely, the cause-specific hazard and the sub-distribution hazard were defined in the competing risks settings, and their nonparametric estimation procedures were explained. In addition, it was shown how the two hazards relate to the cause-specific cumulative incidence. The estimators and confidence intervals of the cause-specific cumulative incidence were provided in three different forms: the Aalen–Johansen form, the product-limit form, and the empirical cumulative distribution form. Even though they look different, they are algebraically the same. Finally, the competing risks analogues to the log-rank test were introduced to test equality of the cause-specific hazards. In Chapter 3, the structure of the multi-state models was described and the Markov model structure was also defined formally through the transition hazard or the transition probabilities, as well as the state occupation probabilities. For the nonparametric estimation, data representation formats were introduced, such as a transition-based format and a stacked format, and the Nelson–Aalen estimator of the cumulative transition hazard and the Aalen–Johansen estimator of transition probabilities were formulated. Finally, the use of Markov models was illustrated with an example of the disease course from HIV (Human Immunodeficiency Virus) infection to death, with SI (Syncytium Inducing) phenotype and AIDS (Acquired Immune Deficiency Syndrome) as intermediate events.

In the previous Chapters 2 and 3, the transition probabilities are quantified at a population level, respectively, under the competing risks settings and multi-state models. Often, we are also curious about whether and how these depend on characteristics that vary within the population. In Chapter 4, the basic structure of the cause-specific proportional hazards model was described such as the standard estimation of the regression coefficients and the cause-specific baseline hazard rates. For more powerful and flexible analysis rather than using separate analysis, the ways of combining all competing events or transitions into one analysis were addressed, and the reasons this approach gives valid estimates were explained. The ideas were also extended to the multi-state models. Chapter 4 concentrated on modeling and estimation of effects on the cause-specific hazard and the transition hazard, while Chapter 5 introduced three ways to quantify effects on the cumulative probability for prediction. In the classical survival settings, the effect of a covariate on the hazard is interpreted as a qualitatively similar effect on the cumulative probability. However, in the competing risks settings, effects on the cumulative scale may differ from those on the cause-specific hazard. This is why the cause-specific cumulative incidence is determined by all cause-specific hazards, that is, from all competing events. In Chapter 5, alternative regression models have been suggested in order that the parameter estimates in the models have a direct interpretation on the cumulative probability. Basically, each of the three approaches is an extension to the regression setting of the representation of the nonparametric estimator of the cause-specific cumulative incidence. In addition, if the value of a covariate varies over time, a landmark model or data representation in the counting process format, such as pseudo-individual approach or interval approach can be employed.

The nice tutorial article of competing risks and multi-state models has been published by Putter, Fiocco, and Geskus (Putter et al., 2007) in Statistics in Medicine. Nevertheless, due to lack of detailed procedure for analyzing real examples, the researchers who are not familiar with these two topics have had difficulty of implementing their own problems with R, SAS, and so on. This book may be very useful for them in the sense that it is written with implementation-oriented purpose rather than theory-based only.

References

Putter, H., Fiocco, M., and Geskus, R. B. (2007). Tutorial in biostatistics: Competing risks and multi-state models. *Statistics in Medicine* **26**, 2389–2430.

Jinheum Kim
Department of Applied Statistics
University of Suwon
Gyeonggi-do, Korea
jkimdt65@gmail.com

NIELS G. BECKER, **Modeling to Inform Infectious Disease Control**. Boca Raton, FL: Chapman and Hall/CRC

Infectious disease transmission modeling has been popular to those who study statistics, mathematics, bioengineering, or epidemiology with an interest in infectious diseases, and for those with experiences in user-friendly programs such as Berkley Madonna. "Modeling to Inform Infectious Disease Control" aims to provide knowledge of transmission modeling to readers from all kinds of academic background.

This book categorizes transmission between susceptibles and infectives into two categories: individual contact and community households. Then, it considers homogeneous and heterogeneous mixing among individuals for each transmission type. Further, it describes complicated models in a simple manner, e.g., two household sizes or two individual types, to focus on conveying the meaning of measurements.

In this book, vaccination is considered as an intervention in infectious disease transmission models. From Chapter 1–4, this book explains the effective reproduction number R. R is one of the most important measurements in transmission modeling and any corresponding critical vaccination coverage models that satisfy the transmission threshold property for various transmission types. These include, R, household R ($\mathrm{R_H}$), type-specific R ($\mathrm{R_T}$), and corresponding critical vaccination coverages. Different estimations and interpretations are also given according to different social mixing structures. Chapters 5 and 6 explain the use of transmission intensity function and the infectivity profile according to the status of vaccinee, then assort and compare the differences in impact. According to the book, the transmission intensity function is difficult to estimate because it is influenced by biological and social mixing components, which are difficult to quantify.

Chapters 7–10 describe the estimation of relevant model inputs and provide guidance for an effective planning of interventions. Chapter 7 provides different methods of targeting high transmission intensities to mitigate transmission of infectious diseases by measuring social distances. Such method was applied to the study of human swine influenza epidemic in 2009. Chapter 8 discusses the current issues in vaccine research. It describes the difficulty in calculating epidemic sizes in large community settings, the deterministic model (which is suited to explore the effectiveness of intervention with large populations), and the comparison between the deterministic model and the stochastic model. Also, Chapter 8 illustrates the changes in type-specific attack rates over time and compares them according to various proportions of immunized individuals through mass vaccination campaigns. For further discussion of vaccinations, the impact on herd immunity, individual type, and full/partial vaccine protection is included. Chapter 9 explains the applications and limitations of deterministic models to predict endemic and epidemic transmission in large populations. Chapter 10 describes model-free comparison by using goodness-of-fit test and chains of transmission.

In summary, the consistency among the chapters (estimation of effective reproduction number and critical vaccine coverage) allows the readers to understand the transmission of infectious diseases under different assumptions. Also, the exercises and supplementary material presented at the end of each chapter will benefit the readers and the lecturers.

Jin Kyung Park
Department of Biostatistics and Data management
International Vaccine Institute
Seoul, Korea
jkpark@ivi.int

CHRISTOPHER R. BILDER AND THOMAS M. LOUGHIN, **Analysis of Categorical Data with R.** Boca Raton, FL: Chapman and Hall/CRC

There have been many excellent textbooks published for categorical data analysis (CDA) such as Agresti (2012). Most of these books focus on methodological works. All programs and codes of R software or SAS are usually provided in the Appendix. As the authors declared "The focus on this book is on the analysis of data, rather than on the mathematical development of methods." Unlike other CDA books, this book focuses on how to analyze categorical data properly, and it is a modern account using the vastly popular R software. The authors use R not only as a data analysis method but also as a learning tool. All R codes are given in the main body of the book. While putting emphasis on R, this book does not require any prior experience with R. Appendix A summarizes an introduction to the essential features and functions of R. Introductory details on the use of R are provided in the earlier chapters in order to help inexperienced R users. I think this book is a well-organized self-teaching book.

Chapter 1 is an introduction chapter focusing on the binary response. It discusses how to estimate and make inferences about a single probability of success and generalizes this discussion to two success probabilities. First, it reviews on distributional theory including Bernoulli binomial distributions and inference on confidence intervals and testing. In addition, the authors provide the detailed R code explanations for the readers who have little experience with R. Although this chapter introduces some basic materials in the first course of elementary statistics, it provides more advanced topics that are rarely covered in the course of elementary statistics. Such topics include Wilson confidence interval (CI), Agresti–Coull CI, and Clopper–Pearson CI for confidence interval for one success probability, and Agresti–Caffo CI for difference of two success probabilities.

Chapter 2 is a generalization of Chapter 1 to a situation where there are many different possible probabilities of success to infer. It maintains the view of analyzing a binary response in the framework of regression models. It starts with linear regression and moves quickly to logistic regression. All issues on logistic regression are comprehensively covered, such as estimation and hypothesis testing. For example, odds ratio interpretation interaction, even convergence issue, and Monte Carlo simulation are well described. The generalized linear model (GLM) is briefly handled at the end. One minor change I want to point out is the order of presentation. Note that R function for the logistic regression is glm( ) function. The use of glm( ) function requires specification of family and link function. I think it would be better to introduce the GLM first along with glm( ) function and then treat the logistic regression model as one of GLMs, which might provide a smoother reading.

Chapter 3 is for analyzing a multicategory response, where the response variable is chosen from a fixed set of more than two category choices. For an IxK contingency table, analyses for nominal response regression and ordinal response regression models are covered. For the nominal responses, the general logit model is introduced. For the ordinal responses, the cumulative logit model and proportional odds model are described. I found out that this chapter is the most similarly described chapter compared to other textbooks for categorical data analysis.

Chapter 4 is for analyzing a count response. It starts with Poisson regression for count responses and then introduces log-linear models as special case of Poisson regression for contingency table, where all column and row variables are indicator variables for nominal response and some score variables for ordinal response. Next, Poisson rate regression is described to analyze rate response. Finally, it covers special cases that are used to describe situations that do not fit clearly into the standard Poisson regression framework. That is "Zero-inflation," where the responses come from a mixture of distributions, and one of distributions places all of its mass on zero. I like this last part "Zero inflation" the most because it provides a good real example for clear conceptual illustration and provides a perfect solution to demonstrate how to handle it properly.

Chapter 5 is for model selection and evaluation. The first part of this chapter presents model selection techniques that can be used to select an appropriate set of explanatory variables from among a larger pool of candidate variables. Model selection is one of hottest topics in statistics these days. The traditional model selection approaches and then a modern variable selection method, least absolute shrinkage and selection operator (LASSO) are described. The second part of this chapter is for model evaluation which checks whether the model assumptions are well satisfied. Residual analysis, goodness-of-fit tests, and influence analysis are provided. Over-dispersion, which is one of most common phenomena in CDA, is then covered. It is very impressive that its causes, interpretations, detection, and solutions are completely described. The final Section 5.4 provides two examples: one for logistic regression and the other for Poisson regression. Through these two examples, all comprehensive analyses are well illustrated. The whole analysis process provides a good guideline of CDA.

The final chapter covers various additional topics: binary responses and testing error, exact inference, categorical data analysis in complex survey designs, and "Choose all that apply" data. In addition, mixed models and estimating equations for correlated data, and Bayesian methods for categorical data are also covered. This chapter is very vast and ranges around 120 pages. I think some sections such as 6.3, 6.5, and 6.6 can be handled in separate chapters. In particular, mixed models and generalized estimating equations are important topics in CDA analysis which needs to be handled more extensively.

In summary, I think this book is well organized and nicely written. I really enjoyed reading it, though I did not have time to run all of the R codes by myself. I want to use this book as a textbook in a graduate course for CDA. This book has many advantages. Compared to other standard textbooks, its complete coverage of examples from many different research areas and the R codes would let the students (other readers) become experts in CDA in fields. Use of the same examples throughout different chapters consistently provides excellent process of data analysis. Furthermore, an extensive set of exercises at the end of each chapter (over 65 pages on all) that differ in scope and subject manner would be good supporting materials for enhancing practical experiences of real data analysis.

Acknowledgement

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIP) (No. 2012R1A3A2026438).

Reference

Agresti, A (2012). Analysis of Categorical Data Analysis, 3rd edition. NY: John and Wiley.

Taesung Park
Department of Statistics
Seoul National University
Seoul, Korea
tspark@stats.snu.ac.kr