Biometrics 73, 356–359
March 2017

DOI: 10.1002/biom.12674

BOOK REVIEWS

EDITOR:
DONNA PAULER ANKERST

STEPHEN BURGESS, SIMON G. THOMPSON. **Mendelian Randomization: Methods for Using Genetic Variants in Causal Estimation**. Boca Raton: CRC Press

About 20 years ago there was much talk of whether epidemiology had reached its limits (Taubes and Charles, 1995). The clearest risk factors had been identified and epidemiologists were struggling to distinguish smaller risks from the effects of confounding and the other biases that plague observational studies. Then along came the genetic revolution and new life was breathed into epidemiology, because genetic risk factors are far less likely to be affected by confounding. Very soon it was realized that if a genetic variant is associated with a risk factor and if that risk factor is, in turn, *causally* related to a disease, then the variant will show an association with the disease. So Mendelian randomization was born as an indirect way of testing the causality of risk factor-disease associations in epidemiological studies.

Mendelian randomization (MR) became very popular, very quickly and researchers soon realized that MR has close links with instrumental variable analysis as used in econometrics and causal inference as developed by statisticians. Both of these are complex, highly mathematical areas of research and therein lies the main problem addressed by Burgess and Thompson in their new book (Burgess and Simon, 2015). There is a gulf between the theory and the practice. Users of MR are predominantly biologists, epidemiologists, and other medical researchers who do not all have ready access to the mathematical theory and conversely the mathematicians do not always appreciate the practical constraints of epidemiology. The authors have aimed their book at epidemiologists and medical statisticians but anyone with a basic knowledge of regression will understand most of the contents, because the algebra is kept to a minimum and emphasis is on explaining the ideas that underlie MR.

The first half of the book provides an introduction to MR, links it to causal inference and instrumental variable analysis and then discusses examples. The section half considers some statistical issues associated with MR, in particular the bias induced by weak genetic associations, the way that many variants can be combined to increase power and the role of meta-analysis in MR. The book concludes with a short section on future directions. Here, we find the one limitation of the book, MR research is developing so quickly that topics mentioned under “future directions” are already mainstream and ideally would be covered in much more detail. For instance, the possibility of two-sample MR based on published data is covered in half a page, while this is already established as one of the most important forms of MR.

Burgess and Thompson’s book will serve as an excellent introduction to Mendelian randomization for anyone who wants to understand the underlying statistical issues, but the reader will need other sources if they want to be up to date with the very latest developments.

References

Burgess, S. and Simon, G. T. (2015). *Mendelian Randomization: Methods for Using Genetic Variants in Causal Estimation.* Boca Raton: CRC Press.

Taubes, G. and Charles, C. M. (1995). Epidemiology faces its limits. *Science* **269**(5221), 164.

John Thompson
Dept. of Health Sciences
Univ. of Leicester
Leicester, UK
trj@le.ac.uk

TIE-HUA NG. **Noninferiority Testing in Clinical Trials: Issues and Challenges.** Boca Raton: CRC Press

In drug development, the scenario where it has to be shown that a treatment is not inferior to an active control is a common and critical task. "Noninferiority Testing in Clinical Trials – Issues and Challenges" is written for statisticians as well as non-statisticians and aims to give an overview of the issues and challenges of noninferiority (NI) testing suitable for both groups of researchers involved in drug development. The author states explicitly that the book addresses therapeutic equivalence when referring to NI testing as opposed to bioequivalence, which is mentioned in some chapters, but is not the focus of the book.

In order to achieve these goals, the author starts with an overview of the general topic in the first chapter and introduces the notation. In addition, many different definitions for the equivalence margin and recommendations for its choice reported by the literature are summarized. The second chapter deals with the choice of the NI-margin for the mean difference and covers topics related to NI-trials in general such as the constancy assumption and preservation. In Chapter 2, continuous endpoints and mean differences are covered, whereas in Chapter 3, mean ratios for a continuous endpoint and hazard ratios for a time-to-event endpoint are discussed. Subsequently, in Chapter 4, odds ratios for a binary endpoint are addressed. Chapter 5 discusses two statistical approaches for testing the NI hypothesis. It introduces a fixed-margin method and a synthesis method and addresses concepts of preservation and discounting. Chapter 6 discusses critical issues related to switching between superiority and NI, which is of interest in active equivalence control studies. Chapter 7 covers effect size estimation of standard therapy compared to placebo when there are multiple historical studies as opposed to a single study. Meta-analysis concepts in general, fixed and random effects models, and a discounting-and-then-pooling approach are presented. Chapter 8 discusses the gold standard NI design as well as multiple-arm clinical trials with three treatment groups and continuous endpoints. Chapters 9-11 in summary cover regulatory guidelines, the analysis data set, nonadherence to the trial protocol, and a thrombolytic example. At last, Chapter 12 concludes the book by discussing issues and challenges related to NI testing, design, and analysis of related trials.

Consecutive chapters of the book are well structured and built on one another with some that can be read independently of the chronology (Chapters 6, 8, and 9). The author gives interesting examples and discusses all issues critically with respect to existing literature. Furthermore, striking and convincing motivation is given as well as discussions at the ends of most chapters. Many diagrams and graphs clearly illustrate the respective content and statistical issues. The book allows readers from different fields to understand the concept of NI testing as well as issues and challenges related to NI testing.

Sonja Grill
Mathematics Life Science Unit
Technical University of Munich
Munich, Germany
sonja.grill@tum.de

MW-L CHEUNG. **Meta-analysis: A Structural Equation Modeling Approach.** New York: Wiley

This book presents meta-analysis from a structural equation modeling perspective. Meta-analysis is a prominent statistical tool in many research disciplines. It is a statistical method to combine the data of separate independent studies in order to draw overall conclusions based on the pooled results. The popularity of meta-analysis is growing rapidly. Structural equation modeling (SEM) is a technique to evaluate models with intricate relations between observed and latent variables. SEM includes factor analysis and path analysis, and, as Cheung shows in this book, SEM also includes meta-analysis.

As Cheung writes in the preface, he learned the technique of SEM first, after which everything looked like a structural equation model to him. The purpose of the book is therefore to integrate meta-analysis in the SEM-framework (coined *SEM-based meta-analysis*), and to present methods to fit structural equation models on pooled correlation matrices (coined *meta-analytic structural equation modeling* or *MASEM*). SEM-based meta-analysis facilitates the transmission of advances between the fields of SEM and meta-analysis. For example, meta-analysis may benefit from the options to handle missing data that are available in SEM, and SEM may benefit from the amount of data that is available in meta-analytic datasets. The other introduced technique, MASEM, goes one step further than standard (multivariate) meta-analysis, by fitting structural models (such as factor models or path models) to explain the pooled correlations between research variables. As a result, researchers can test all hypothesized relations between variables simultaneously. Some immediate advantages of applying MASEM instead of standard meta-analysis of correlation coefficients are the possibility to evaluate mediation by testing indirect effects, the possibility to model latent variables, and the availability of fit-indices to compare competing theoretical models.

By making the integrated framework of meta-analysis and SEM available and understandable to scientists, and providing detailed illustrations in the R-package metaSEM, this book may be an important starting point for both methodological researchers who wish to develop new meta-analytic methods and for substantive researchers who want to apply the state of the art meta-analytic methods. Given that the book outlines a "structural equation modeling approach" to meta-analysis, good knowledge of SEM is essential to fully understand and appreciate the presented material. The author indicates that only basic knowledge of SEM, similar to the first year of research methods covered in most graduate programs, is required. However, I think some parts of the book are quite more technical than what is understandable after having followed one course in SEM. I would especially expect difficulties if one has learned SEM with a program that does not require specification of matrices, such as AMOS or Mplus. Specifically, I think the first five chapters are still understandable with a basic SEM-knowledge, but from Chapter 6 onward it will get severely more difficult to keep up with the technicalities. Although this may be seen as a disadvantage, I see it as a strength that both the beginner as well as experienced SEM-researchers and meta-analysts will gain a lot of knowledge from reading this book.

After explaining the motivation behind the book in the first chapter, Chapter 2 provides a brief review of structural equation modeling. This chapter immediately shows the ability of the author to explain the core of a subject within only a couple of sentences. Explanations are generally short but complete, with references to relevant other work in case a reader wants to read more about the topic. The introduction in SEM is well aligned with the rest of the book, in that it covers all essential elements of SEM, as well as specific topics that are essential to formulate meta-analytic models as SEMs. These advanced topics are the use of phantom variables, definition variables, and full information maximum likelihood estimation in SEM.

The next four chapters together form the core part on SEM-based meta-analysis. Chapter 3 handles the computation of effect sizes and sampling (co)variances, Chapters 4 and 5 discuss univariate and multivariate meta-analysis, respectively, and Chapter 6 handles modeling dependency using three-level meta-analysis. All chapters have a similar structure, where the topic is introduced in a general way first, followed by the structural equation modeling approach to the specific topic, and a discussion of extensions or related issues. Each chapter ends with extensive illustrations of the techniques in metaSEM (and in some cases the R packages openMx or lavaan). Chapter 7 covers MASEM by discussing conventional approaches and two-stage structural equation modeling. It presents a discussion of related issues and open ends, as well as various examples in metaSEM. Chapter 8 is titled "Advanced topics in SEM-based meta-analysis" and provides a discussion of using restricted (or residual) maximum likelihood versus standard maximum likelihood, as well as options for handling missing values in moderators in SEM-based meta-analysis. The last chapter illustrates how SEM-based meta-analysis can be applied in Mplus, which is the most popular SEM-software at the moment. Although it is clear from the illustrations that the dedicated metaSEM package is much easier to use, the examples in Mplus show that one can in principle apply meta-analysis with any SEM-program, not just with OpenMx.

Throughout the book, Cheung demonstrates skill in explaining (dis)similarities between different concepts. Examples that were real eye-openers to me are the explanation of the connection between the Wald test and likelihood ratio test (Chapter 2), the issue of treating sampling variances as known versus estimated (Chapter 4), the discussion of univariate versus multivariate models from a missing data perspective (Chapter 5), the relationship between multivariate and three-level meta-analyses (Chapter 6), and the discussions of TSSEM in relation to alternative methods (Chapter 7).

This book is a unique contribution to the literature, as it is the only work that discusses meta-analysis from a structural equation modeling perspective. Books that may be close to the present work are the introductions to meta-analysis from Noel Card or Michael Borenstein and colleagues. These authors, however, take a more conceptual approach and cover a broader range of topics in research synthesis, such as searching and selecting the literature. Moreover, these authors do not discuss the connection with structural equation modeling. I myself have written a book on meta-analytic structural equation modeling, which is the topic of Chapter 7. However, my book does not cover SEM-based meta-analysis. The book by Cheung is therefore the only one that fully covers the relation between and the combination of SEM and meta-analysis. I would recommend anyone with a (methodological) interest in meta-analysis to read this book. For me (experienced with SEM, and interested in meta-analysis), this book was exactly what I needed. The work will probably be within reach on my desk for a long time from now.

Suzanne Jak
Methods and Statistics
Research Institute of Child Development and Education
University of Amsterdam
Amsterdam, The Netherlands
S.Jak@uva.nl

DANIEL ZELTERMAN, **Applied Multivariate Statistics with R.** Heidelberg: Springer

Those of us who teach Multivariate Statistics at the upper undergraduate and graduate levels know the struggles of finding a single book that strikes the right balance between theory and implementation. I myself have oscillated between the fundamental theoretical book by Anderson, the more applied classic by Johnson and Wichern, the informatics tour-de-force by Izenman, and the recent theoretical presentation by Fujikoshi, Ulyanov, and Shimizu. None of these books used the ubiquitous academic freeware R package, and their emphasis on visualization and analysis of real data where confined to brief examples. For those of us stranded in the multivariate desert, Applied Multivariate Statistics with R is a refreshing oasis.

A handy feature of the book is that it is organized pedagogically, which makes it an all-in-one resource for educators. It starts with an Introduction that lures the reader into wanting to read further by interesting examples. One of the initial examples had me at hamburger. Let us be honest, most examples used in textbooks on this subject are boring. But what non-vegetarian university student would not be interested in the quintiple of fat, saturated fat, sodium, carbs, and protein in a Chili's Bacon Burger versus a Sonic Double-Cheeseburger? The book transitions to a brief introduction to the basics of R, with painless coverage of programming and then to the power of R graphics for multivariate data, with all illustrations in color. After seeing the Chernoff faces for the hamburger vectors, I had to read on (and go order a hamburger). What is nice is that the R review is comprehensive enough to not require tutelage elsewhere. As one floats in an anesthesia of pretty graphics, a painless dose of basic linear algebra is administered. This is kept minimal with user-friendly analogies, such as apples and oranges, so that before one knows it they are waking up from a dream and diving into the heart of the subject. Data, pretty graphs, and R code break up the formulas during the rest of the book, making it one of the most readable multivariate statistics texts out there. And to add icing on the cake, each chapter ends with a nice doable set of exercises, something not often found.

The meat of topic begins with univariate distributions in Chapter 5. I like that extensive treatment is given to powerlaw transformations to achieve Normality and to graphical and multiple goodness-of-fit tests for Normality, that simulations and real examples with R code are provided to show what confidence levels and asymptotic distributions for means and variances really mean. For the bivariate Normal distribution in Chapter 6, it was nice to see the various methods for testing that a correlation equals zero, with additional nonparametric tests in the exercises, and the demonstration of how to maximize the likelihood using general maximizing routines in addition to the pre-canned Normal ones. Chapters 7, 8, and 9 contain the traditional course material on the multivariate Normal distribution, factor methods, and multivariable regression, with the continued added spice of enticing data, including US housing prices before and after the financial crash, the nutritional value of popular chocolate candies and health surveys and outcomes.

Chapters 10 and 11 on discrimination/classification and clustering foray into useful machine learning terrain. The new R alternatives for displaying cluster dendograms, including the unrooted tree, triangular branches, and leaves spread in a circle were inspirational to see. This multivariate book was the first that I had seen that included a chapter on time series (Chapter 12). Time series presents another form of multivariate data, that of a long auto-correlated series, or a collection of correlated series, as opposed to independent replications of vectors. Again the book provides provocative real-world examples, such as examination of trends in New York City birth rates and climate change. Inclusion of time series represents the book's effort to provide a comprehensive toolkit. Chapter 13 takes the mission even further into the hot new topics that until now have by and far not appeared, such as how to perform ranking based on multivariate attributes, methods for extreme order statistics, and big and wide data.

In summary, this book is so clearly explained with R code throughout that it could be used as a self-learning text for an applied multivariate course and should be assigned as a self-learning adjunct assignment for a graduate level theoretical multivariate course. The real-word examples are page turners and ubiquitous use of color and fancy graphs easily explained make this usually dry topic an exciting one. The author is to be congratulated on a much needed book filling an important gap in the field.

Donna Pauler Ankerst
Department of Mathematics
Technical University Munich
Munich, Germany
ankerst@tum.de