

# Statistics for Biology and Health

## *Series Editors:*

M. Gail  
K. Krickeberg  
J. Samet  
A. Tsiatis  
W. Wong

# Statistics for Biology and Health

---

Bacchieri/Cioppa: Fundamentals of Clinical Research  
Borchers/Buckland/Zucchini: Estimating Animal Abundance: Closed Populations  
Burzykowski/Molenberghs/Buyse: The Evaluation of Surrogate Endpoints  
Duchateau/Janssen: The Frailty Model  
Everitt/Rabe-Hesketh: Analyzing Medical Data Using S-PLUS  
Ewens/Grant: Statistical Methods in Bioinformatics: An Introduction, 2nd ed.  
Gentleman/Carey/Huber/Irizarry/Dudoit: Bioinformatics and Computational Biology Solutions Using R and Bioconductor  
Hougaard: Analysis of Multivariate Survival Data  
Keyfitz/Caswell: Applied Mathematical Demography, 3rd ed.  
Klein/Moeschberger: Survival Analysis: Techniques for Censored and Truncated Data, 2nd ed.  
Kleinbaum/Klein: Survival Analysis: A Self-Learning Text, 2nd ed.  
Kleinbaum/Klein: Logistic Regression: A Self-Learning Text, 2nd ed.  
Lange: Mathematical and Statistical Methods for Genetic Analysis, 2nd ed.  
Lazar: The Statistical Analysis of fMRI Data  
Manton/Singer/Suzman: Forecasting the Health of Elderly Populations  
Martinussen/Scheike: Dynamic Regression Models for Survival Data  
Moyé: Multiple Analyses in Clinical Trials: Fundamentals for Investigators  
Nielsen: Statistical Methods in Molecular Evolution  
O'Quigley: Proportional Hazards Regression  
Parmigiani/Garrett/Irizarry/Zeger: The Analysis of Gene Expression Data: Methods and Software  
Proschan/LanWittes: Statistical Monitoring of Clinical Trials: A Unified Approach  
Siegmond/Yakir: The Statistics of Gene Mapping  
Simon/Korn/McShane/Radmacher/Wright/Zhao: Design and Analysis of DNA Microarray Investigations  
Sorensen/Gianola: Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics  
Stallard/Manton/Cohen: Forecasting Product Liability Claims: Epidemiology and Modelling in the Manville Asbestos Case  
Steyerberg: Clinical Prediction Models: A Practical Approach to Model Development, Validation, and Updating  
Sun: The Statistical Analysis of Interval-censored Failure Time Data  
Therneau/Gramsbsch: Modelling Survival Data: Extending the Cox Model  
Ting: Dose Finding in Drug Development  
Vittinghoff/Glidden/Shiboski/McCulloch: Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models  
Wu/Ma/Casella: Statistical Genetics of Quantitative Traits: Linkage, Maps, and QTL  
Zhang/Singer: Recursive Partitioning in the Health Sciences  
Zuur/Ieno/Smith: Analysing Ecological Data

Ewout W. Steyerberg

# Clinical Prediction Models

A Practical Approach to Development,  
Validation, and Updating



Springer

E.W. Steyerberg  
Department of Public Health  
Erasmus MC  
3000 CA Rotterdam  
The Netherlands

*Series Editors*

M. Gail  
National Cancer Institute  
Bethesda, MD 20892  
USA

A. Tsiatis  
Department of Statistics  
North Carolina State University  
Raleigh, NC 27695  
USA

K. Krickeberg  
Le Chatelet  
F-63270 Manglieu  
France

W. Wong  
Department of Statistics  
Stanford University  
Stanford, CA 94305-4065  
USA

J. Sarnet  
Department of Epidemiology  
School of Public Health  
Johns Hopkins University  
615 Wolfe Street  
Baltimore, MD 21205-2103  
USA

ISBN: 978-0-387-77243-1      e-ISBN: 978-0-387-77244-8  
DOI: 10.1007/978-0-387-77244-8

Library of Congress Control Number: 2008929620

© 2009 Springer Science+Business Media, LLC

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

While the advice and information in this book are believed to be true and accurate at the date of going to press, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

springer.com

*For Aleida, Matthijs, Laurens and Suzanne*

# Preface

Prediction models are important in various fields, including medicine, physics, meteorology, and finance. Prediction models will become more relevant in the medical field with the increase in knowledge on potential predictors of outcome, e.g. from genetics. Also, the number of applications will increase, e.g. with targeted early detection of disease, and individualized approaches to diagnostic testing and treatment. The current era of evidence-based medicine asks for an individualized approach to medical decision-making. Evidence-based medicine has a central place for meta-analysis to summarize results from randomized controlled trials; similarly prediction models may summarize the effects of predictors to provide individualized predictions of a diagnostic or prognostic outcome.

## Why Read This Book?

My motivation for working on this book stems primarily from the fact that the development and applications of prediction models are often suboptimal in medical publications. With this book I hope to contribute to better understanding of relevant issues and give practical advice on better modelling strategies than are nowadays widely used.

Issues include:

- (a) Better predictive modelling is sometimes easily possible; e.g. a large data set with high quality data is available, but all continuous predictors are dichotomized, which is known to have several disadvantages.
- (b) Small samples are used:
  - Studies are underpowered, with unreliable answers to difficult questions such as “Which are the most important predictors in this prediction problem?”
  - The problem of small sample size is aggravated by doing a complete case analysis which discards information from nearly complete records. Statistical imputation methods are nowadays available to exploit all available information.

- Predictors are omitted that should reasonably have been included based on subject matter knowledge. Modelers rely too much on the limited data that they have available in their data set, instead of wisely combining information from several sources, such as medical literature and experts in the field.
- Stepwise selection methods are abundant, which are especially risky in small data sets.
- Modelling approaches are used that require higher numbers. Data-hungry techniques, such as tree modelling and neural network modelling, should not be used in small data sets.
- No attempts are made towards validation, or validation is done inefficiently. For example, a split-sample approach is followed, leading to a smaller sample for model development and a smaller sample for model validation. Better methods are nowadays available and should be used far more often.

(c) Claims are exaggerated:

- Often we see statements such as ‘the predictors were identified’; in many instances such findings may not be reproducible and may largely represent noise.
- Models are not internally valid, with overoptimistic expectations of model performance in new patients.
- One modern method with a fancy name is claimed as being superior to a more traditional regression approach, while no convincing evidence exists, and a suboptimal model strategy was followed for the regression model.
- Researchers are insufficiently aware of overfitting, implying that their apparent findings are merely coincidental (“the curse of dimensionality”).

(d) Poor generalizability:

- If models are not internally valid, we cannot expect them to generalize.
- Models are developed for each local situation, discarding earlier findings on effects of predictors and earlier models; a framework for continuous improvement and updating of prediction models is required.

In this book; I try to suggest many small improvements in modelling strategies. Combined, these improvements hopefully lead to better prediction models.

## Intended Audience

Readers should have a basic knowledge of biostatistics, especially regression analysis, but no strong background in mathematics is required. The number of formulas is deliberately kept small. Usually a bottom-up approach is followed in teaching regression analysis techniques, starting with model assumptions, estimation methods, and basic interpretation. This book is more top-down: given that we want to predict an outcome, how can we best utilize regression techniques?

Three levels of readers are envisioned:

- (a) The core intended audience is formed by epidemiologists and applied biostatisticians who want to develop or apply a prediction model. Both students and professionals should find practical guidance in this book, especially by the proposed seven steps to develop a valid model (Part II).
- (b) A second group is formed by clinicians, policy makers, and health care professionals who want to judge a study that presents a prediction model. This book should aid them in a critical appraisal, providing explanations of terms and concepts that are common in publications on prediction models. They should try to read chapters of particular interest, or read the main text of the chapters. They can skip the examples and more technical sections (indicated with\*).
- (c) A third group includes more theoretical researchers, such as (bio)statisticians and computer scientists, who want to improve the methods that we use in prediction models. They may find inspiration for further theoretical work and simulation studies in this book. Many of the methods in prediction modelling are not fully developed yet, and common sense underlies some of the proposed approaches in this book.

## Other sources

Many excellent text books exist on regression analysis techniques, but these usually do not have a focus on modelling strategies for prediction. The main exception is Frank Harrell's book "Regression Modelling Strategies".<sup>174</sup> He brings advanced biostatistical concepts to practical application, supported by the Design and Hmisc libraries for S+ software (nowadays: packages for R). Harrell's book may however be too advanced for clinical and epidemiological researchers. This also holds for the Hastie, Tibshirani, and Friedman's quite thorough text book "The Elements of Statistical Learning".<sup>181</sup> These books are very useful for a more in-depth discussion of statistical techniques and strategies. Harrell's book provided the main inspiration for the presented work here. Another good companion book is the Vittinghoff et al. book on "Regression Methods in Biostatistics".<sup>472</sup>

Various sources at the internet can be used that explain terms used in this book. Frank Harrell has a glossary at his web site: [<http://biostat.mc.vanderbilt.edu/twiki/pub/Main/ClinStat/glossary.pdf>]. Other useful sources include [[http://www.aiaccess.net/e\\_gm.htm](http://www.aiaccess.net/e_gm.htm)] and Wikipedia.

## Structure

It has been found that people learn by example, by checklists, and by own discovery. Therefore I provide many examples throughout the text, including the essential computer code and output. I also suggest a checklist for prediction modelling (Part II).



Own discovery is possible with exercises per chapter, with data sets provided at the book's web site: <http://www.clinicalpredictionmodels.org>.

Many statistical techniques and approaches are readily possible with any modern software package. Personally, I work with SPSS for simple, straightforward analyses, but this package is insufficient for more advanced analyses which are essential in prediction modelling. The SAS computer package is more advanced, but may not be so practical for some. A package such as Stata is very suitable. It may be similar in capabilities to S-plus software, which was my preferred program for advanced prediction modelling since a stay at Duke University in 1996. The R software is very similar in nature to S-plus, and has several additional advantages: the software is for free, and innovations in biostatistical methods become readily available for R. Therefore, R is the natural choice as the software accompanying this book. R software is available at <http://www.cran.r-project.org>, with help files and a tutorial.

Some R commands are provided in this book; full programs can be downloaded from a web site (<http://www.clinicalpredictionmodels.org>). This web site also provides a number of data sets that can be downloaded for application of the described techniques. I provide data files in SPSS format that can readily be imported in R and other packages. Further, comments on the text can be submitted electronically.

# Acknowledgements

Many have made small to large contributions to this book. I'm very grateful to all. Frank Harrell has been a source of inspiration for my research in the area of clinical prediction models, together with Hans van Houwelingen, who has developed many of the theoretical innovations that are presented in this book. At my department (Public Health, Erasmus MC), Dik Habbema has encouraged me to further develop my interests in prognosis and prediction research, together with my close coworker René Eijkemans. Hester Lingsma was very supportive in the last phase of finishing this book. Several NIHES MSc students, PhD students, colleagues, and external reviewers made specific comments on various chapters. These include Lex Burdorf, Sarwa Darwish Murad, Sonja Deurloo, René Eijkemans, Frank Harrell, Yolanda van der Graaf, Cecile Janssens, Mike Kattan, Michael Koller, Carl Moons, Jeroen van der Net, Saskia Pluijm, Yvonne Vergouwe, Andrew Vickers, and many others.

I thank John Kimmel from Springer for his encouragement and practical support since we first met in Boston in 2005. I specifically would like to thank investigators who allowed their data sets to be made available for didactic purposes, including Kerry Lee (Duke University) and the GUSTO-I investigators. Finally, I thank my family, in particular my beloved wife Aleida, for their warm and ongoing support, and for allowing me to devote time, often at nights and weekends, to work on this book.

Rotterdam

Ewout Steyerberg

# Contents

<b>Preface</b> .....	vii
<b>Acknowledgements</b> .....	xi
<b>1 Introduction</b> .....	1
1.1 Prognosis and Prediction in Medicine .....	1
1.1.1 Prediction Models and Decision-Making .....	1
1.2 Statistical Modelling for Prediction .....	2
1.2.1 Model Uncertainty .....	3
1.2.2 Sample Size .....	4
1.3 Structure of the Book .....	5
1.3.1 Part I: Prediction Models in Medicine .....	5
1.3.2 Part II: Developing Valid Prediction Models .....	6
1.3.3 Part III: Generalizability of Prediction Models .....	6
1.3.4 Part IV: Applications .....	7
1.3.5 Questions and Exercises .....	7
<b>Part I Prediction Models in Medicine</b>	
<b>2 Applications of Prediction Models</b> .....	9
2.1 Applications: Medical Practice and Research .....	11
2.2 Prediction Models for Public Health .....	12
2.2.1 Targeting of Preventive Interventions .....	12
2.2.2 Example: Incidence of Breast Cancer .....	12
2.3 Prediction Models for Clinical Practice .....	13
2.3.1 Decision Support on Test Ordering .....	13
2.3.2 Example: Predicting Renal Artery Stenosis .....	14
2.3.3 Starting Treatment: the Treatment Threshold .....	15
2.3.4 Example: Probability of Deep Venous Thrombosis .....	16
2.3.5 Intensity of Treatment .....	16
2.3.6 Example: Defining a Poor Prognosis Subgroup in Cancer .....	18

- 2.3.7 Cost-Effectiveness of Treatment. . . . . 18
- 2.3.8 Delaying Treatment. . . . . 19
- 2.3.9 Example: Spontaneous Pregnancy Chances . . . . . 19
- 2.3.10 Surgical Decision-Making . . . . . 21
- 2.3.11 Example: Replacement of Risky Heart Valves . . . . . 21
- 2.4 Prediction Models for Medical Research . . . . . 23
  - 2.4.1 Inclusion and Stratification in an RCT . . . . . 23
  - 2.4.2 Example: Selection for TBI Trials . . . . . 24
  - 2.4.3 Covariate Adjustment in an RCT . . . . . 25
  - 2.4.4 Gain in Power by Covariate Adjustment . . . . . 26
  - \*2.4.5 Example: Analysis of the GUSTO-III Trial . . . . . 27
  - 2.4.6 Prediction Models and Observational Studies . . . . . 27
  - 2.4.7 Propensity Scores . . . . . 28
  - 2.4.8 Example: Statin Treatment Effects . . . . . 28
  - 2.4.9 Provider Profiling . . . . . 29
  - 2.4.10 Example: Ranking Cardiac Outcome . . . . . 29
- 2.5 Concluding Remarks. . . . . 30
- 3 Study Design for Prediction Models. . . . . 33**
  - 3.1 Study Design. . . . . 33
  - 3.2 Cohort Studies for Prognosis . . . . . 33
    - 3.2.1 Retrospective Designs. . . . . 35
    - 3.2.2 Example: Predicting Early Mortality in Oesophageal Cancer . . . . . 35
    - 3.2.3 Prospective Designs . . . . . 35
    - 3.2.4 Example: Predicting Long-Term Mortality in Oesophageal Cancer . . . . . 36
    - 3.2.5 Registry Data . . . . . 36
    - 3.2.6 Example: Surgical Mortality in Oesophageal Cancer. . . . . 37
    - 3.2.7 Nested Case–Control Studies . . . . . 37
    - 3.2.8 Example: Perioperative Mortality in Major Vascular Surgery . . . . . 38
  - 3.3 Studies for Diagnosis . . . . . 38
    - 3.3.1 Cross-Sectional Study Design and Multivariable Modelling 38
    - 3.3.2 Example: Diagnosing Renal Artery Stenosis . . . . . 38
    - 3.3.3 Case–Control Studies . . . . . 39
    - 3.3.4 Example: Diagnosing Acute Appendicitis . . . . . 39
  - 3.4 Predictors and Outcome . . . . . 39
    - 3.4.1 Strength of Predictors . . . . . 39
    - 3.4.2 Categories of Predictors . . . . . 40
    - 3.4.3 Costs of Predictors . . . . . 40
    - 3.4.4 Determinants of Prognosis . . . . . 41
    - 3.4.5 Prognosis in Oncology . . . . . 41

- 3.5 Reliability of Predictors . . . . . 42
  - 3.5.1 Observer Variability . . . . . 42
  - 3.5.2 Example: Histology in Barrett’s Oesophagus . . . . . 42
  - 3.5.3 Biological Variability . . . . . 43
  - 3.5.4 Regression Dilution Bias . . . . . 43
  - 3.5.5 Example: Simulation Study on Reliability of a Binary Predictor . . . . . 43
  - 3.5.6 Choice of Predictors . . . . . 44
- 3.6 Outcome . . . . . 44
  - 3.6.1 Types of Outcome . . . . . 44
  - 3.6.2 Survival Endpoints . . . . . 45
  - 3.6.3 Example: Relative Survival in Cancer Registries . . . . . 45
  - 3.6.4 Composite End Points . . . . . 46
  - 3.6.5 Example: Mortality and Composite End Points in Cardiology . . . . . 46
  - 3.6.6 Choice of Prognostic Outcome . . . . . 46
  - 3.6.7 Diagnostic End Points . . . . . 47
  - 3.6.8 Example: PET Scans in Oesophageal Cancer . . . . . 47
- 3.7 Phases of Biomarker Development . . . . . 47
- 3.8 Statistical Power . . . . . 48
  - 3.8.1 Statistical Power to Identify Predictor Effects . . . . . 49
  - 3.8.2 Examples of Statistical Power Calculations . . . . . 49
  - 3.8.3 Statistical Power for Reliable Predictions . . . . . 50
- 3.9 Concluding Remarks . . . . . 51
  
- 4 Statistical Models for Prediction . . . . . 53**
  - 4.1 Continuous Outcomes . . . . . 53
    - 4.1.1 Examples of Linear Regression . . . . . 54
    - 4.1.2 Economic Outcomes . . . . . 54
    - 4.1.3 Example: Prediction of Costs . . . . . 54
    - 4.1.4 Transforming the Outcome . . . . . 54
    - 4.1.5 Performance: Explained Variation . . . . . 55
    - 4.1.6 More Flexible Approaches . . . . . 55
  - 4.2 Binary Outcomes . . . . . 57
    - 4.2.1  $R^2$  in Logistic Regression Analysis . . . . . 58
    - 4.2.2 Calculation of  $R^2$  on the Log Likelihood Scale . . . . . 58
    - 4.2.3 Models Related to Logistic Regression . . . . . 60
    - 4.2.4 Bayes Rule . . . . . 61
    - 4.2.5 Example: Calculations with Likelihood Ratios . . . . . 62
    - 4.2.6 Prediction with Naïve Bayes . . . . . 63
    - 4.2.7 Examples of Naïve Bayes . . . . . 65
    - 4.2.8 Calibration and Naïve Bayes . . . . . 65
    - 4.2.9 Logistic Regression and Bayes . . . . . 65
    - 4.2.10 More Flexible Approaches to Binary Outcomes . . . . . 65

- 4.2.11 Classification and Regression Trees . . . . . 67
- 4.2.12 Example: Mortality in Acute MI Patients. . . . . 67
- 4.2.13 Advantages and Disadvantages of Tree Models. . . . . 67
- 4.2.14 Trees as Special Cases of Logistic Regression Modelling . . . . . 69
- 4.2.14 Other Methods for Binary Outcomes . . . . . 70
- 4.2.15 Summary on Binary Outcomes. . . . . 71
- 4.3 Categorical Outcomes. . . . . 71
  - 4.3.1 Polytomous Logistic Regression. . . . . 72
  - 4.3.2 Example: Histology of Residual Masses . . . . . 72
  - 4.3.3 Alternative Models . . . . . 73
  - 4.3.4 Comparison of Modelling Approaches. . . . . 74
- 4.4 Ordinal Outcomes. . . . . 74
  - 4.4.1 Proportional Odds Logistic Regression . . . . . 75
  - 4.4.2 Alternative: Continuation Ratio Model. . . . . 77
- 4.5 Survival Outcomes . . . . . 77
  - 4.5.1 Cox Proportional Hazards Regression . . . . . 77
  - 4.5.2 Predicting with Cox . . . . . 78
  - 4.5.3 Proportionality Assumption . . . . . 78
  - 4.5.4 Kaplan–Meier Analysis . . . . . 79
  - 4.5.5 Example: NFI After Treatment of Leprosy. . . . . 79
  - 4.5.6 Parametric Survival. . . . . 80
  - 4.5.7 Example: Replacement of Risky Heart Valves . . . . . 80
  - 4.5.8 Summary on Survival Outcomes . . . . . 81
- 4.6 Concluding Remarks. . . . . 81
- 5 Overfitting and Optimism in Prediction Models. . . . . 83**
  - 5.1 Overfitting and Optimism . . . . . 83
    - 5.1.1 Example: Surgical Mortality in Oesophagectomy . . . . . 84
    - 5.1.2 Variability within One Centre. . . . . 84
    - 5.1.3 Variability between Centres: Noise vs. True Heterogeneity . . . . . 85
    - 5.1.4 Predicting Mortality by Centre: Shrinkage. . . . . 87
  - 5.2 Overfitting in Regression Models. . . . . 87
    - 5.2.1 Model Uncertainty: Testimation. . . . . 87
    - 5.2.2 Other Biases . . . . . 89
    - 5.2.3 Overfitting by Parameter Uncertainty. . . . . 90
    - 5.2.4 Optimism in Model Performance . . . . . 90
    - 5.2.5 Optimism-Corrected Performance . . . . . 92
  - 5.3 Bootstrap Resampling . . . . . 92
    - 5.3.1 Applications of the Bootstrap . . . . . 93
    - 5.3.2 Bootstrapping for Regression Coefficients . . . . . 93
    - 5.3.3 Bootstrapping for Optimism Correction. . . . . 94
    - 5.3.4 Calculation of Optimism-Corrected Performance . . . . . 95

- 5.3.5 Example: Stepwise Selection in 429 Patients . . . . . 96
- 5.4 Cost of Data Analysis . . . . . 97
  - 5.4.1 Example: Cost of Data Analysis in a Tree Model . . . . . 98
  - 5.4.2 Practical Implications . . . . . 98
- 5.5 Concluding Remarks. . . . . 99
- 6 Choosing Between Alternative Statistical Models. . . . . 101**
  - 6.1 Prediction with Statistical Models . . . . . 101
    - 6.1.1 Testing of Model Assumptions and Prediction. . . . . 102
    - 6.1.2 Choosing a Type of Model . . . . . 102
  - 6.2 Modelling Age–Outcome Relationships. . . . . 103
    - 6.2.1 Age and Mortality After Acute MI. . . . . 103
    - 6.2.2 Age and Operative Mortality . . . . . 103
    - 6.2.3 Age–Outcome Relationships in Other Diseases. . . . . 106
  - 6.3 Head-to-Head Comparisons . . . . . 107
    - 6.3.1 StatLog Results. . . . . 107
    - 6.3.2 GUSTO-I Modelling Comparisons. . . . . 108
    - 6.3.3 GUSTO-I Results . . . . . 109
  - 6.4 Concluding Remarks. . . . . 110

**Part II Developing Valid Prediction Models**

- 7 Dealing with Missing Values . . . . . 113**
  - 7.1 Missing Values in Predictors. . . . . 115
    - 7.1.1 Inefficiency of Complete Case Analysis. . . . . 116
    - 7.1.2 Interpretation of Analyses with Missing Data . . . . . 117
    - 7.1.3 Missing Data Mechanisms . . . . . 117
    - 7.1.4 Summary Points . . . . . 118
  - 7.2 Regression Coefficients Under MCAR, MAR, and MNAR. . . . . 118
    - 7.2.1 R Code . . . . . 120
  - 7.3 Missing Values in Regression Analysis . . . . . 121
    - 7.3.1 Imputation Principle . . . . . 121
    - 7.3.2 Simple and More Advanced Single Imputation Methods . . . . . 122
    - 7.3.3 Multiple Imputation . . . . . 123
  - 7.4 Defining the Imputation Model. . . . . 124
    - 7.4.1 Transformations of Variables . . . . . 125
    - 7.4.2 Imputation Models for SI . . . . . 125
    - 7.4.3 Summary Points . . . . . 126
  - 7.5 Simulations of Imputation Under MCAR, MAR, and MNAR. . . . . 126
    - 7.5.1 Multiple Predictors . . . . . 127
  - 7.6 Imputation of Missing Outcomes . . . . . 128
  - 7.7 Guidance to Missing Values in Prediction Research . . . . . 129

- 7.7.1 Patterns of Missingness. . . . . 129
- 7.7.2 Simple Approaches. . . . . 130
- 7.7.3 Maximum Fraction of Missing Values Before Omitting  
a Predictor . . . . . 131
- 7.7.4 Single or Multiple Imputation for Predictor Effects? . . . . 131
- 7.7.5 Single or Multiple Imputation for Predictions? . . . . . 132
- 7.7.6 Reporting of Missing Values in Prediction Research . . . . 133
- 7.8 Concluding Remarks. . . . . 134
  - 7.8.1 Summary Statements . . . . . 135
  - 7.8.2 Currently Available Software and Challenges . . . . . 136
- 8 Case Study on Dealing with Missing Values . . . . . 139**
  - 8.1 Introduction. . . . . 139
    - 8.1.1 Aim . . . . . 139
    - 8.1.2 Patient Selection . . . . . 140
    - 8.1.3 Selection of Potential Predictors. . . . . 140
    - 8.1.4 Coding and Time Dependency of Predictors . . . . . 141
  - 8.2 Missing Values in the IMPACT Study . . . . . 142
    - 8.2.1 Missing Values in Outcome . . . . . 142
    - 8.2.2 Quantification of Missingness of Predictors. . . . . 143
    - 8.2.3 Patterns of Missingness. . . . . 144
  - 8.3 Imputation of Missing Predictor Values . . . . . 147
    - 8.3.1 Correlations Between Predictors . . . . . 147
    - 8.3.2 Imputation Model . . . . . 147
    - 8.3.3 Distributions of Imputed Values . . . . . 149
  - 8.4 Estimating Adjusted Effects . . . . . 149
    - 8.4.1 Adjusted Analysis for Complete Predictors:  
Age and Motor Score . . . . . 151
    - 8.4.2 Adjusted Analysis for Incomplete Predictors: Pupils . . . . 154
  - 8.5 Multivariable Analyses . . . . . 155
  - 8.6 Concluding Remarks. . . . . 155
- 9 Coding of Categorical and Continuous Predictors . . . . . 159**
  - 9.1 Categorical Predictors. . . . . 159
    - 9.1.1 Examples of Categorical Coding . . . . . 160
  - 9.2 Continuous Predictors. . . . . 161
    - 9.2.1 Examples of Continuous Predictors . . . . . 161
    - 9.2.2 Categorization of Continuous Predictors . . . . . 162
  - 9.3 Non-Linear Functions for Continuous Predictors. . . . . 163
    - 9.3.1 Polynomials . . . . . 164
    - 9.3.2 Fractional Polynomials . . . . . 164
    - 9.3.3 Splines. . . . . 165
    - 9.3.4 Example: Functional Forms with RCS or FP . . . . . 166
    - 9.3.5 Extrapolation and Robustness. . . . . 166



- 9.4 Outliers and Truncation . . . . . 167
  - 9.4.1 Example: Glucose Values and Outcome of TBI . . . . . 168
- 9.5 Interpretation of Effects of Continuous Predictors . . . . . 170
  - 9.5.1 Example: Predictor Effects in TBI . . . . . 171
- 9.6 Concluding Remarks. . . . . 172
  - 9.6.1 Software . . . . . 172
- 10 Restrictions on Candidate Predictors . . . . . 175**
  - 10.1 Selection Before Studying the Predictor–Outcome Relationship . . . . . 175
    - 10.1.1 Selection Based on Subject Knowledge . . . . . 175
    - 10.1.2 Example: Too Many Candidate Predictors . . . . . 176
    - 10.1.3 Meta-Analysis for Candidate Predictors . . . . . 176
    - 10.1.4 Example: Predictors in Testicular Cancer. . . . . 176
    - 10.1.5 Selection Based on Distributions . . . . . 177
  - 10.2 Combining Similar Variables . . . . . 177
    - 10.2.1 Example: Coding of Comorbidity. . . . . 178
    - 10.2.2 Assessing the Equal Weights Assumption . . . . . 178
    - 10.2.3 Logical Weighting. . . . . 179
    - 10.2.4 Statistical Combination. . . . . 180
  - 10.3 Averaging Effects . . . . . 180
    - 10.3.1 Example: Chlamydia Trachomatis Infection Risks . . . . . 180
    - 10.3.2 Example: Acute Surgery Risk Relevant for Elective Patients? . . . . . 180
  - 10.4 Case study: Family History for Prediction of a Genetic Mutation . . . . . 181
    - 10.4.1 Clinical Background and Patient Data . . . . . 181
    - 10.4.2 Similarity of Effects . . . . . 182
    - 10.4.3 CRC and Adenoma in a Proband . . . . . 184
    - 10.4.4 Age of CRC in Family History . . . . . 185
    - 10.4.5 Full Prediction Model for Mutations . . . . . 186
  - 10.5 Concluding Remarks. . . . . 187
- 11 Selection of Main Effects. . . . . 191**
  - 11.1 Predictor Selection . . . . . 191
    - 11.1.1 Reduction Before Modelling. . . . . 191
    - 11.1.2 Reduction While Modelling . . . . . 192
    - 11.1.3 Collinearity . . . . . 192
    - 11.1.4 Parsimony . . . . . 193
    - 11.1.5 Should Non-Significant Variables Be Removed? . . . . . 193
    - 11.1.6 Summary Points . . . . . 194
  - 11.2 Stepwise Selection . . . . . 194
    - 11.2.1 Stepwise Selection Variants . . . . . 194
    - 11.2.2 Stopping Rules in Stepwise Selection. . . . . 195

- 11.3 Advantages of Stepwise Methods . . . . . 196
- 11.4 Disadvantages of Stepwise Methods. . . . . 197
  - 11.4.1 Instability of selection. . . . . 197
  - 11.4.2 Biased Estimation of Coefficients. . . . . 199
  - 11.4.3 Bias of Stepwise Selection and Events Per Variable. . . . . 199
  - 11.4.4 Misspecification of Variability. . . . . 201
  - 11.4.5 Exaggeration of P-Values . . . . . 204
  - 11.4.6 Predictions of Worse Quality Than from a Full Model . . . . . 204
- 11.5 Influence of Noise Variables . . . . . 205
- 11.6 Univariate Analyses and Model Specification . . . . . 206
  - 11.6.1 Pros and Cons of Univariate Pre-Selection. . . . . 207
  - 11.6.2 Testing of Predictors within Domains. . . . . 207
- 11.7 Modern Selection Methods . . . . . 207
  - 11.7.1 Bootstrapping for Selection . . . . . 208
  - 11.7.2 Bagging and Boosting. . . . . 208
  - 11.7.3 Bayesian Model Averaging (BMA) . . . . . 208
  - 11.7.4 Practical Advantages of BMA . . . . . 209
  - 11.7.5 Shrinkage of Regression Coefficients to Zero . . . . . 210
- 11.8 Concluding Remarks. . . . . 210

**12 Assumptions in Regression Models: Additivity**

- and Linearity . . . . . 213**
- 12.1 Additivity and Interaction Terms . . . . . 213
  - 12.1.1 Potential Interaction Terms to Consider . . . . . 214
  - 12.1.2 Interactions with Treatment . . . . . 214
  - 12.1.3 Other Potential Interactions . . . . . 215
  - 12.1.4 Example: Time and Survival After Valve Replacement . . . . . 216
- 12.2 Selection, Estimation and Performance with Interaction Terms. . . . . 216
  - 12.2.1 Example: Age Interactions in GUSTO-I. . . . . 217
  - 12.2.2 Estimation of Interaction Terms . . . . . 217
  - 12.2.3 Better Prediction with Interaction Terms? . . . . . 219
  - 12.2.4 Summary Points . . . . . 220
- 12.3 Non-linearity in Multivariable Analysis . . . . . 220
  - 12.3.1 Multivariable Restricted Cubic Splines (RCS). . . . . 220
  - 12.3.2 Multivariable Fractional Polynomials (FP). . . . . 221
  - 12.3.3 Multivariable Splines in GAM . . . . . 222
- 12.4 Example: Non-Linearity in Testicular Cancer Case Study . . . . . 222
  - 12.4.1 Details of Multivariable FP and GAM Analyses . . . . . 224
  - 12.4.2 GAM in Univariate and Multivariable Analysis. . . . . 224
  - 12.4.3 Predictive Performance . . . . . 226
  - 12.4.4 R code for Non-Linear Modelling . . . . . 227
- 12.5 Concluding Remarks. . . . . 227
  - 12.5.1 Recommendations. . . . . 228

**13 Modern Estimation Methods** . . . . . 231

13.1 Predictions from Regression and Other Models . . . . . 231

13.2 Shrinkage . . . . . 232

13.2.1 Uniform Shrinkage . . . . . 233

13.2.2 Uniform Shrinkage in GUSTO-1 . . . . . 233

13.3 Penalized Estimation. . . . . 234

13.3.1 Penalized Maximum Likelihood Estimation. . . . . 234

13.3.2 Penalized ML in Sample4. . . . . 235

13.3.3 Shrinkage, Penalization, and Model Selection . . . . . 238

13.4 Lasso. . . . . 238

13.4.1 Estimation of Lasso Model . . . . . 238

13.4.2 Lasso in GUSTO-I . . . . . 239

13.4.3 Predictions after Shrinkage. . . . . 239

13.4.4 Model Performance after Shrinkage. . . . . 240

13.5 Concluding Remarks. . . . . 240

**14 Estimation with External Information** . . . . . 243

14.1 Combining Literature and Individual Patient Data . . . . . 243

14.1.1 Adaptation Method 1 . . . . . 244

14.1.2 Adaptation Method 2 . . . . . 244

14.1.3 Estimation . . . . . 245

14.1.4 Simulation Results . . . . . 245

14.1.5 Performance of Adapted Model . . . . . 247

14.1.6 Improving Calibration. . . . . 247

14.2 Example: Mortality of Aneurysm Surgery . . . . . 248

14.2.1 Meta-Analysis. . . . . 248

14.2.2 Individual Patient Data Analysis. . . . . 249

14.2.3 Adaptation Results . . . . . 250

14.3 Alternative Approaches. . . . . 251

14.3.1 Overall Calibration . . . . . 251

14.3.2 Bayesian Methods: Using Data Priors to Regression Modelling . . . . . 251

14.3.3 Example: Predicting Neonatal Death . . . . . 252

14.3.4 Example: Mortality of Aneurysm Surgery . . . . . 252

14.4 Concluding Remarks. . . . . 253

**15 Evaluation of Performance**. . . . . 255

15.1 Overall Performance Measures. . . . . 255

15.1.1 Explained Variation:  $R^2$ . . . . . 255

15.1.2 Brier Score . . . . . 257

15.1.3 Example: Performance of Testicular Cancer Prediction Model . . . . . 257

15.1.4 Overall Performance Measures in Survival . . . . . 258

- 15.1.5 Decomposition in Discrimination and Calibration . . . . . 259
- 15.1.6 Summary Points . . . . . 259
- 15.2 Discriminative Ability. . . . . 260
  - 15.2.1 Sensitivity and Specificity of Prediction Models . . . . . 260
  - 15.2.2 Example: Sensitivity and Specificity of Testicular Cancer Prediction Model . . . . . 260
  - 15.2.3 ROC Curve . . . . . 260
  - 15.2.4  $R^2$  vs.  $c$  . . . . . 262
  - 15.2.5 Box Plots and Discrimination Slope . . . . . 264
  - 15.2.6 Lorenz Curve. . . . . 264
  - 15.2.7 Discrimination in Survival Data . . . . . 267
  - 15.2.8 Example: Discrimination of Testicular Cancer Prediction Model. . . . . 268
  - 15.2.9 Verification Bias and Discriminative Ability . . . . . 269
  - 15.2.10 R Code . . . . . 269
- 15.3 Calibration. . . . . 270
  - 15.3.1 Calibration Plot . . . . . 270
  - 15.3.2 Calibration in Survival . . . . . 271
  - 15.3.3 Calibration-in-the-Large . . . . . 271
  - 15.3.4 Calibration Slope . . . . . 272
  - 15.3.5 Estimation of Calibration-in-the-Large and Calibration Slope . . . . . 272
  - 15.3.6 Other Calibration Measures . . . . . 273
  - 15.3.7 Calibration Tests . . . . . 274
  - 15.3.8 Goodness-of-Fit Tests . . . . . 274
  - 15.3.9 Calibration of Survival Predictions. . . . . 276
  - 15.3.10 Example: Calibration in Testicular Cancer Prediction Model. . . . . 276
  - 15.3.11 Calibration and Discrimination. . . . . 278
  - 15.3.12 R Code . . . . . 278
- 15.4 Concluding Remarks. . . . . 278
  - 15.4.1 Bibliographic Notes . . . . . 279
- 16 Clinical Usefulness . . . . . 281**
  - 16.1 Clinical Usefulness . . . . . 281
    - 16.1.1 Intuitive Approach to the Cutoff . . . . . 282
    - 16.1.2 Decision-Analytic Approach to the Cutoff . . . . . 282
    - 16.1.3 Error Rate and Accuracy . . . . . 283
    - 16.1.4 Accuracy Measures for Clinical Usefulness . . . . . 284
    - 16.1.5 Decision Curves . . . . . 284
    - 16.1.6 Examples of NB in Decision Curves . . . . . 285
    - 16.1.7 Example: Clinical Usefulness of Prediction Model for Testicular Cancer. . . . . 286

- 16.1.8 Decision Curves for Testicular Cancer Example . . . . . 287
- 16.1.9 Verification Bias and Clinical Usefulness. . . . . 288
- 16.1.10 R Code . . . . . 289
- 16.2 Discrimination, Calibration, and Clinical Usefulness. . . . . 289
  - 16.2.1 Aim of the Prediction Model and Performance Measures . . . . . 290
  - 16.2.2 Summary Points . . . . . 291
- 16.3 From Prediction Models to Decision Rules . . . . . 291
  - 16.3.1 Performance of Decision Rules . . . . . 292
  - 16.3.2 Treatment Benefit in Prognostic Subgroups . . . . . 294
  - 16.3.3 Evaluation of Classification Systems . . . . . 294
- 16.4 Concluding Remarks. . . . . 295
  - 16.4.1 Bibliographic notes . . . . . 296
- 17 Validation of Prediction Models. . . . . 299**
  - 17.1 Internal vs. External Validation, and Validity . . . . . 299
  - 17.2 Internal Validation Techniques . . . . . 300
    - 17.2.1 Apparent Validation . . . . . 300
    - 17.2.2 Split-Sample Validation . . . . . 301
    - 17.2.3 Cross-Validation . . . . . 302
    - 17.2.4 Bootstrap Validation . . . . . 303
  - 17.3 External Validation Studies. . . . . 304
    - 17.3.1 Temporal Validation . . . . . 305
    - 17.3.2 Example: Development and Validation of a Model for Lynch Syndrome . . . . . 306
    - 17.3.3 Geographic Validation . . . . . 307
    - 17.3.4 Fully Independent Validation . . . . . 308
    - 17.3.5 Reasons for Poor Validation . . . . . 309
  - 17.4 Concluding Remarks. . . . . 310
- 18 Presentation Formats . . . . . 313**
  - 18.1 Prediction vs. Decision Rules . . . . . 313
  - 18.2 Clinical Prediction Models . . . . . 315
    - 18.2.1 Regression Formula . . . . . 315
    - 18.2.2 Confidence Intervals for Predictions. . . . . 316
    - 18.2.3 Nomograms. . . . . 317
    - 18.2.4 Score Chart . . . . . 319
    - 18.2.5 Tables with Predictions . . . . . 320
    - 18.2.6 Specific Formats . . . . . 321
  - 18.3 Case Study: Clinical Prediction Model for Testicular Cancer Model . . . . . 321
    - 18.3.1 Regression Formula from Logistic Model . . . . . 321
    - 18.3.2 Nomogram . . . . . 324

18.3.3	Score Chart . . . . .	324
18.3.4	Coding with Categorization . . . . .	327
18.3.5	Summary Points . . . . .	327
18.4	Clinical Decision Rules. . . . .	328
18.4.1	Regression Tree. . . . .	328
18.4.2	Score Chart Rule. . . . .	328
18.4.3	Survival Groups . . . . .	329
18.4.4	Meta-Model. . . . .	329
18.5	Concluding Remarks. . . . .	330

### Part III Generalizability of Prediction Models

<b>19</b>	<b>Patterns of External Validity . . . . .</b>	<b>333</b>
19.1	Determinants of External Validity . . . . .	335
19.1.1	Case-Mix. . . . .	335
19.1.2	Differences in Case-Mix . . . . .	336
19.1.3	Differences in Regression Coefficients. . . . .	336
19.2	Impact on Calibration, Discrimination, and Clinical Usefulness. . . . .	337
19.2.1	Simulation Set-Up. . . . .	338
19.2.2	Performance Measures . . . . .	339
19.3	Distribution of Predictors . . . . .	340
19.3.1	More- or Less-Severe Case-Mix According to X. . . . .	340
19.3.2	Example: Interpretation of Testicular Cancer Validation . . . . .	341
19.3.3	More or Less Heterogeneous Case-Mix According to X . . . . .	341
19.3.4	More- or Less-Severe Case-Mix According to Z . . . . .	342
19.3.5	More or Less Heterogeneous Case-Mix According to Z . . . . .	344
19.4	Distribution of Observed Outcomes $Y$ . . . . .	344
19.5	Coefficients $\beta$ . . . . .	345
19.5.1	Coefficient of Linear Predictor $< 1$ . . . . .	345
19.5.2	Coefficients Different. . . . .	346
19.5.3	R Code . . . . .	346
19.5.4	Influence of Different Coefficients . . . . .	347
19.5.5	Other Scenarios of Invalidity . . . . .	348
19.5.6	Summary of Patterns of Invalidity . . . . .	348
19.6	Reference Values for Performance . . . . .	349
19.6.1	Calculation of Reference Values. . . . .	349
19.6.2	R Code . . . . .	350
19.6.3	Performance with Refitting. . . . .	350
19.6.4	Examples: Testicular Cancer and TBI . . . . .	351

19.7	Estimation of Performance . . . . .	352
19.7.1	Uncertainty in Validation of Performance . . . . .	352
19.7.2	Estimating Standard Errors in Validation Studies. . . . .	354
19.7.3	Summary Points . . . . .	354
19.8	Design of External Validation Studies . . . . .	355
19.8.1	Power of External Validation Studies . . . . .	355
19.8.2	Required Sample Sizes for Validation Studies . . . . .	356
19.8.3	Summary Points . . . . .	357
19.9	Concluding Remarks. . . . .	358
<b>20</b>	<b>Updating for a New Setting . . . . .</b>	<b>361</b>
20.1	Updating the Intercept. . . . .	361
20.1.1	Simple Updating Methods . . . . .	362
20.1.2	Bayesian Updating . . . . .	362
20.2	Approaches to More-Extensive Updating. . . . .	363
20.2.1	A comparison of Eight Updating Methods . . . . .	364
20.3	Case Study: Validation and Updating in GUSTO-I . . . . .	366
20.3.1	Validity of TIMI-II Model for GUSTO-I . . . . .	366
20.3.2	Updating the TIMI-II Model for GUSTO-I . . . . .	368
20.3.3	Performance of Updated Models . . . . .	369
20.3.4	R Code for Updating Methods . . . . .	370
20.4	Shrinkage and Updating . . . . .	371
20.4.1	Example: Shrinkage towards Re-calibrated Values in GUSTO-I. . . . .	371
20.4.2	R code for Shrinkage and Penalization in Updating. . . . .	372
20.5	Sample Size and Updating Strategy . . . . .	373
20.5.1	Simulations of Sample Size, Shrinkage, and Updating Strategy . . . . .	374
20.6	Validation and Updating of Tree Models . . . . .	376
20.6.1	Example: Tree Modelling in Testicular Cancer . . . . .	377
20.7	Validation and Updating of Survival Models . . . . .	378
20.7.1	Case Study: Validation of a Simple Index for Non-Hodgkin’s Lymphoma . . . . .	379
20.7.2	Updating the Prognostic Index . . . . .	380
20.7.3	Re-calibration for Groups by Time Points . . . . .	380
20.7.4	Re-calibration with a Cox Regression Model. . . . .	381
20.7.5	Parametric Re-calibration . . . . .	382
20.7.6	Summary Points . . . . .	384
20.8	Continuous Updating . . . . .	384
20.8.1	A Continuous Updating Strategy . . . . .	385
20.8.2	Example: Continuous Updating in GUSTO-I. . . . .	386
20.9	Concluding Remarks. . . . .	388

<b>21</b>	<b>Updating for Multiple Settings</b> . . . . .	391
21.1	Differences Between Settings . . . . .	391
21.1.1	Testing for Calibration-in-the Large . . . . .	391
21.1.2	Illustration of Heterogeneity in GUSTO-I . . . . .	392
21.1.3	Updating for Better Calibration-in-the Large . . . . .	393
21.1.4	Empirical Bayes Estimates . . . . .	394
21.1.5	Illustration of Updating in GUSTO-I . . . . .	394
21.1.6	Testing and Updating of Predictor Effects . . . . .	396
21.1.7	Heterogeneity of Predictor Effects in GUSTO-I . . . . .	396
21.1.8	R Code for Random Effect Analyses . . . . .	397
21.2	Provider Profiling . . . . .	398
21.2.1	Indicators for Differences Between Centres . . . . .	398
21.2.2	Ranking of Centres . . . . .	399
21.2.3	Example: Provider Profiling in Stroke . . . . .	401
21.2.4	Testing of Differences Between Centres . . . . .	401
21.2.5	Estimation of Differences Between Centres . . . . .	402
21.2.6	Uncertainty in Differences . . . . .	403
21.2.7	Ranking of Centres . . . . .	404
21.2.8	Essential R Code for Provider Profiling . . . . .	405
21.2.9	Guidelines for Provider Profiling . . . . .	406
21.3	Concluding Remarks . . . . .	406
21.3.1	Bibliographic Notes . . . . .	407

## Part IV Applications

<b>22</b>	<b>Prediction of a Binary Outcome: 30-Day Mortality After Acute Myocardial Infarction</b> . . . . .	411
22.1	GUSTO-I Study . . . . .	411
22.1.1	Acute Myocardial Infarction . . . . .	411
22.1.2	Treatment Results from GUSTO-I . . . . .	412
22.1.3	Prognostic Modelling in GUSTO-I . . . . .	412
22.2	General Considerations of Model Development . . . . .	415
22.2.1	Research Question and Intended Application . . . . .	415
22.2.2	Outcome and Predictors . . . . .	416
22.2.3	Study Design and Analysis . . . . .	416
22.3	Seven Modelling Steps in GUSTO-I . . . . .	417
22.3.1	Data Inspection . . . . .	417
22.3.2	Coding of Predictors . . . . .	418
22.3.3	Model Specification . . . . .	418
22.3.4	Model Estimation . . . . .	418
22.3.5	Model Performance . . . . .	419
22.3.6	Model Validation . . . . .	419
22.3.7	Presentation . . . . .	420



- 22.4 Validity . . . . . 421
  - 22.4.1 Internal Validity: Overfitting. . . . . 421
  - 22.4.2 External Validity: Generalizability . . . . . 421
  - 22.4.3 Summary Points . . . . . 421
- 22.5 Translation into Clinical Practice . . . . . 422
  - 22.5.1 Score Chart for Choosing Thrombolytic Therapy . . . . . 422
  - 22.5.2 Predictions for Choosing Thrombolytic Therapy . . . . . 423
  - 22.5.3 Covariate Adjustment in GUSTO-I . . . . . 424
- 22.6 Concluding Remarks. . . . . 425
  
- 23 Case Study on Survival Analysis: Prediction of Secondary Cardiovascular Events . . . . . 427**
  - 23.1 Prognosis in the SMART Study . . . . . 427
    - 23.1.1 Patients in SMART . . . . . 428
  - 23.2 General Considerations in SMART . . . . . 429
    - 23.2.1 Research Question and Intended Application. . . . . 429
    - 23.2.2 Outcome and Predictors . . . . . 429
    - 23.2.3 Study Design and Analysis . . . . . 432
  - 23.3 Data Inspection Steps in the SMART Cohort. . . . . 432
  - 23.4 Coding of Predictors . . . . . 435
    - 23.4.1 Extreme Values . . . . . 435
    - 23.4.2 Transforming Continuous Predictors . . . . . 436
    - 23.4.2 Combining Predictors with Similar Effects . . . . . 437
  - 23.5 Model Specification . . . . . 438
    - 23.5.1 Selection . . . . . 440
  - 23.6 Model Estimation, Performance, Validation, and Presentation . . 440
    - 23.6.1 Model Estimation . . . . . 440
    - 23.6.2 Model Performance. . . . . 442
    - 23.6.3 Model Validation: Stability. . . . . 442
    - 23.6.4 Model Validation: Optimism. . . . . 444
    - 23.6.5 Model Presentation . . . . . 444
  - 23.7 Concluding Remarks. . . . . 444
  
- 24 Lessons from Case Studies . . . . . 447**
  - 24.1 Sample Size. . . . . 447
    - 24.1.1 Example: Sample Size and Number of Predictors . . . . . 447
    - 24.1.2 Number of Predictors . . . . . 448
    - 24.1.3 Potential Solutions . . . . . 449
  - 24.2 Validation . . . . . 450
    - 24.2.1 Examples of Internal and External Validation . . . . . 450
  - 24.3 Subject Matter Knowledge . . . . . 451
  - 24.4 Data Sets . . . . . 452
    - 24.4.1 GUSTO-I Prediction Models . . . . . 453
    - 24.4.2 Modern Learning Methods in GUSTO-I . . . . . 453

24.4.3	Modelling Strategies in Small Data Sets from GUSTO-I . . . . .	453
24.4.4	SMART Case Study . . . . .	453
24.4.5	Testicular Cancer Case Study . . . . .	455
24.4.6	Abdominal Aortic Aneurysm Case Study. . . . .	455
24.4.7	Traumatic Brain Injury Data Set. . . . .	459
24.5	Concluding Remarks. . . . .	459
<b>References.</b> . . . . .		463
<b>Index</b> . . . . .		487