Giuseppe La Torre

Applied Epidemiology and Biostatistics

SEEd

© SE*Ed* srl
Piazza Carlo Emanuele II, 19 – 10123 Torino – Italy
Tel. +39.011.566.02.58 - Fax +39.011.518.68.92
www.edizioniseed.it
info@edizioniseed.it

First edition
September 2010

ISBN 978-88-8968-849-6

Although the information about medication given in this book has been carefully checked, the author and publisher accept no liability for the accuracy of this information. In every individual case the user must check such information by consulting the relevant literature.

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the Italian Copyright Law in its current version, and permission for use must always be obtained from SE*Ed* Medical Publishers srl. Violations are liable to prosecution under the Italian Copyright Law.

Summary

Preface **7**

1. **Measures of Occurrence** 11
   - 1.1. Introduction 11
   - 1.2. Prevalence 12
   - 1.3. Incidence 13
   - 1.4. Practical issues 16
   - 1.5. Practical examples 16
   - References 22

2. **Measures of Association** 23
   - 2.1. Relative risk 23
   - 2.2. Risk difference 25
   - 2.3. Other measures of attributable risk 26
   - 2.4. Practical examples 28
   - References 35

3. **Controlling for Confounding** 37
   - 3.1. What is confounding in epidemiology? 37
   - 3.2. Controlling for confounding factors 38
   - 3.3. How to control for confounding factors 39
   - 3.4. Practical examples 42
   - References 52

4. **Cross-Sectional Studies** 53
   - 4.1. Introduction 53
   - 4.2. Performing a cross-sectional study 55
   - 4.3. A practical example 56
   - References 66

5. Cohort Studies 67
5.1. What is a cohort study? 67
5.2. Why do we need a cohort study? 68
5.3. The eligibility criteria 68
5.4. The structure of a cohort study 69
5.5. Censoring 70
5.6. The statistical analysis in a cohort study 70
5.7. Practical examples 71
References 84

6. Experimental Studies 87
6.1. What is a sample experimental study? 87
6.2. Why do we need an experimental study? 88
6.3. The eligibility criteria 88
6.4. The randomisation process 89
6.5. The blinding 90
6.6. The structure of an experimental study 90
6.7. The statistical analysis in an experimental study 94
6.8. Practical examples 95
References 100

7. Temporal Trend Analysis 101
7.1. Introduction 101
7.2. Basic principles of temporal trend analysis 105
7.3. Practical examples 112
References 116

8. The Surveillance of Sexually Transmitted Infections: the Theory and the Practice 119
8.1. Introduction 119
8.2. Surveillance of sexually transmitted infections in the third millennium 120
8.3. Attributes of a STI surveillance system 122
8.4. Universal versus sentinel surveillance systems 127
8.5. How to perform STI surveillance 128
8.6. Data management and analysis 134
8.7. Practical exercises for analysing a dataset of STIs 135
References 156

9. Systematic Reviews and Meta-Analysis of Clinical Trials 159
9.1. What is a systematic review? What is a meta-analysis? 159
9.2. Why do we need systematic reviews and meta-analyses? 160
9.3. Practical steps of a meta-analysis 165

9.4. A practical example of a meta-analysis of RCTs....174
References....202

10. Meta-Analysis of Observational Studies....207
10.1. Introduction....207
10.2. Practical example....209
10.3. Worked examples....211
References....230

11. Genetic Epidemiology....231
11.1. Key concepts of genetic epidemiology....231
11.2. A practical example: the "candidate gene approach"....234
References....246

12. Analysis of Cost Data Using Bootstrap Technique....249
12.1. Introduction....249
12.2. Basic principles of the bootstrap method....250
12.3. Bootstrap standard normal confidence interval....251
12.4. Percentile method confidence interval....251
12.5. Bias corrected and accelerated (BCa) confidence interval....252
12.6. Application to example....252
References....259

13. Sensitivity, Specificity, and ROC Curves....261
13.1. Study introduction....261
13.2. Sensitivity, specificity, and predictive value....262
13.3. Basic principles of ROC curves....264
13.4. Use of ROC analysis for comparison....273
References....277

14. Measures of Central Tendency and Dispersion....279
14.1. Introduction....279
14.2. Measures of central tendency....279
14.3. Measures of dispersion....282
14.4. Practical exercise....286
References....288

15. Sample Size Calculations....289
15.1. What is a sample size and why do we need a sample?....289
15.2. Steps of a sample size calculation....290
References....302

16. Representation of Data....303
16.1. Introduction....303

16.2. Representation of qualitative variables 304
16.3. Representation of quantitative variables 314
References 324

17. Running Multiple Regression With Quantitative and Qualitative Variables With R 325
17.1. Introduction 325
17.2. The regression model with quantitative and qualitative variables 326
17.3. Practical example: multiple regression with 2 qualitative variables 331
References 347

18. Methods for Assessing Normality of Quantitative Variables 349
18.1. Introduction 349
18.2. Definition of normality 349
18.3. Parametric and nonparametric statistics 351
18.4. How to verify normality of data 351
18.5. Practical examples 353
References 359

19. Quality of Life Evaluation 361
19.1. Quality of life in the general population 361
19.2. Quality of life in the clinical setting 372
References 380
Appendix. Algorithm to create the SF-36 scales 382

20. Disability Adjusted Life Years (DALY) Summary Measure of Population Health 389
20.1. Introduction 389
20.2. Disability adjusted life year (DALY): the concept and its uses 390
20.3. Method for DALY estimation used in the serbian burden of disease study 390
20.4. Practical example: calculation of DALY for colorectal cancer, Serbia, 2000 395
Acknowledgments 397
References 398

Preface

When this book was conceived, as a discussion among members of the section of Public Health Epidemiology of the European Public Health Associations (EUPHA), the main idea was to describe not only theory, but above all how to use the available software for epidemiologic and statistical data analysis.

In the era of Evidence Based Medicine, health professionals are required to fully understand design, analysis and interpretation of the results of research. Furthermore, they should be able to assess the needs of their communities and respond accordingly. To achieve these goals, one needs to be familiar with the basic concepts of epidemiology and biostatistics.

But epidemiology is more than "the study of." Its application and practice are essential to address public health issues.

So, the purpose of the book is to give the reader either the theory concerning specific aspects of technical disciplines as epidemiology and biostatistics, and in the mean time to give the opportunity of replicating under guidance the analysis done by each chapter's author and already published in a given research article. The idea is to use the available software for epidemiologic and statistical data analysis, that each reader can download freely from the Internet.

Concerning the way in which the purpose of the book is to be achieved, it is important to underline that each chapter will present one or more specific examples on how to perform an epidemiological or statistical data analysis. The single chapter will give the reader the possibility of conducting an epidemiological or a statistical analysis, using a step by step approach. In other words, the reader will be able to do the analysis following the detailed description of the commands to use and the figures that represent a picture of the software command and/or output.

Why do we believe the book is needed?

The answer is mainly of technical reason. Up to now, many books concerning epidemiology and biostatistics are available, but no one could give practical examples using different freely available software. This book will use software such as Epi Info, Episheet, Simcalc, StatCalc, RevMan, that are downloadable from the web, and could cover most arguments concerning the two disciplines. In selected cases, we will make examples using commercial statistical software, such as Stata and SPSS.

The reader will be interested in this book because he/she will find a resolution of an epidemiological/biostatistical problem with practical example and a guide to use the software in a very detailed and efficient way.

Have you ever been interested in performing an epidemiological data analysis, but you thought to be not able to?

Have you ever been in trouble in making a statistical analysis, because you considered statistics a matter of statisticians only?

Applied Epidemiology and Biostatistics is the answer to you.

Questions as following will find an answer:

- How to perform a multiple logistic regression using your own data?
- How to calculate the 95% confidence intervals of that odds ratio?
- How to perform a meta-analysis of papers of your interest?
- How to make graphs for your report?
- How to make a ROC curve or control for a possible confounding?
- How to calculate the sample size needed for the clinical trial?

This is a manual designed for using software, that are freely downloadable from the web, and could cover most arguments concerning the two disciplines, epidemiology and biostatistics. In selected cases, examples will use commercial statistical packages.

Who is the best reader of this book?

Considering that epidemiology can be seen as the study of factors affecting the health and illness of a certain population, and the Biostatistics is one of the main pillar of the research, in our intention this manual will have as principle possible targets the following:

- Public Health practitioners (professionals, researchers).
- Clinicians (researchers).

- Health Managers (professionals, researchers).
- Teachers of Epidemiology.
- Teachers of Biostatistics.

Finally, I would like to thank all the contributors to this manual. Without their support and suggestions, it would have been impossible to achieve this goal.

Now, do you want to start?

Let's make Epidemiology and Biostatistics together!

Giuseppe La Torre

Instructions for Downloading

To download the software and databases described in this book, you need to:

- access the website **http://download.edizioniseed.it**
- access the "**Download area**" and
- type the code **-------**