

Cambridge University Press
978-1-107-03003-9 - Applied Longitudinal Data Analysis for Epidemiology: A Practical Guide: Second Edition
Jos W. R. Twisk
Frontmatter
[More information](#)

Applied Longitudinal Data Analysis for Epidemiology
A Practical Guide

Cambridge University Press
978-1-107-03003-9 - Applied Longitudinal Data Analysis for Epidemiology: A Practical Guide: Second Edition
Jos W. R. Twisk
Frontmatter
[More information](#)

Cambridge University Press

978-1-107-03003-9 - Applied Longitudinal Data Analysis for Epidemiology: A Practical Guide: Second Edition

Jos W. R. Twisk

Frontmatter

[More information](#)

Applied Longitudinal Data Analysis for Epidemiology

A Practical Guide

Second Edition

Jos W. R. Twisk

Department of Epidemiology and Biostatistics Medical Center and

Department of Health Sciences, Vrije Universiteit Amsterdam, the Netherlands



CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

Cambridge University Press is part of the University of Cambridge.

It furthers the University’s mission by disseminating knowledge in the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781107699922

First edition © Jos W. R. Twisk 2003
Second edition © Jos W. R. Twisk 2013

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First edition first published 2003
Second edition first published 2013
3rd printing 2015

Printed in the United Kingdom by Clays, St Ives plc.

A catalogue record for this publication is available from the British Library

Library of Congress Cataloguing in Publication data
Twisk, Jos W. R., 1962–
Applied longitudinal data analysis for epidemiology : a practical guide / Jos W. R. Twisk, Department of Epidemiology and Biostatistics, Medical Centre and the Department of Health Sciences of the Vrije Universiteit, Amsterdam. – Second edition.
pages cm
Includes bibliographical references and index.
ISBN 978-1-107-03003-9 (hardback) – ISBN 978-1-107-69992-2 (paperback)
1. Epidemiology – Research – Statistical methods. 2. Epidemiology – Longitudinal studies.
3. Epidemiology – Statistical methods. I. Title.
RA652.2.M3T95 2013
614.4 – dc23 2012050470

ISBN 978-1-107-03003-9 Hardback
ISBN 978-1-107-69992-2 Paperback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

Every effort has been made in preparing this book to provide accurate and up-to-date information which is in accord with accepted standards and practice at the time of publication. Although case histories are drawn from actual cases, every effort has been made to disguise the identities of the individuals involved. Nevertheless, the authors, editors and publishers can make no warranties that the information contained herein is totally free from error, not least because clinical standards are constantly changing through research and regulation. The authors, editors and publishers therefore disclaim all liability for direct or consequential damages resulting from the use of material contained in this book. Readers are strongly advised to pay careful attention to information provided by the manufacturer of any drugs or equipment that they plan to use.

An eye for an eye
A tooth for a tooth
And anyway I told the truth
And I'm not afraid to die
NICK CAVE

To Marjon, Mike, and Nick

Cambridge University Press
978-1-107-03003-9 - Applied Longitudinal Data Analysis for Epidemiology: A Practical Guide: Second Edition
Jos W. R. Twisk
Frontmatter
[More information](#)

Contents

	<i>Preface</i>	<i>page</i> xiii
	<i>Acknowledgements</i>	xiv
1	Introduction	1
	1.1 Introduction	1
	1.2 General approach	2
	1.3 Prior knowledge	2
	1.4 Example	2
	1.5 Software	4
	1.6 Data structure	4
	1.7 Statistical notation	4
	1.8 What’s new in the second edition?	5
2	Study design	6
	2.1 Introduction	6
	2.2 Observational longitudinal studies	7
	2.2.1 Period and cohort effects	8
	2.2.2 Other confounding effects	11
	2.2.3 Example	12
	2.3 Experimental (longitudinal) studies	13
3	Continuous outcome variables	16
	3.1 Two measurements	16
	3.1.1 Example	17
	3.2 Non-parametric equivalent of the paired <i>t</i> -test	18
	3.2.1 Example	19
	3.3 More than two measurements	20
	3.3.1 The “univariate” approach: a numerical example	23
	3.3.2 The shape of the relationship between an outcome variable and time	26

viii	Contents	
	3.3.3 A numerical example	27
	3.3.4 Example	29
	3.4 The “univariate” or the “multivariate” approach?	34
	3.5 Comparing groups	35
	3.5.1 The “univariate” approach: a numerical example	37
	3.5.2 Example	38
	3.6 Comments	42
	3.7 Post-hoc procedures	44
	3.7.1 Example	44
	3.8 Different contrasts	45
	3.8.1 Example	46
	3.9 Non-parametric equivalent of MANOVA for repeated measurements	48
	3.9.1 Example	50
4	Continuous outcome variables – relationships with other variables	51
	4.1 Introduction	51
	4.2 “Traditional” methods	51
	4.3 Example	53
	4.4 Longitudinal methods	55
	4.5 Generalized estimating equations	57
	4.5.1 Introduction	57
	4.5.2 Working correlation structures	57
	4.5.3 Interpretation of the regression coefficients derived from GEE analysis	60
	4.5.4 Example	61
	4.5.4.1 Introduction	61
	4.5.4.2 Results of a GEE analysis	63
	4.5.4.3 Different correlation structures	66
	4.6 Mixed model analysis	69
	4.6.1 Introduction	69
	4.6.2 Mixed models for longitudinal studies	70
	4.6.3 Example	73
	4.6.4 Comments	80
	4.7 Comparison between GEE analysis and mixed model analysis	81
	4.7.1 The “adjustment for covariance” approach	83
	4.7.2 Extensions of mixed model analysis	84
	4.7.3 Comments	84
5	The modeling of time	86
	5.1 The development over time	86
	5.2 Comparing groups	95
	5.3 The adjustment for time	99

ix	Contents	
6	Other possibilities for modeling longitudinal data	103
	6.1 Introduction	103
	6.2 Alternative models	103
	6.2.1 Time-lag model	103
	6.2.2 Model of changes	105
	6.2.3 Autoregressive model	107
	6.2.4 Overview	108
	6.2.5 Example	108
	6.2.5.1 Introduction	108
	6.2.5.2 Data structure for alternative models	109
	6.2.5.3 GEE analysis	109
	6.2.5.4 Mixed model analysis	113
	6.3 Comments	116
	6.4 Another example	117
7	Dichotomous outcome variables	119
	7.1 Simple methods	119
	7.1.1 Two measurements	119
	7.1.2 More than two measurements	121
	7.1.3 Comparing groups	121
	7.1.4 Example	122
	7.1.4.1 Introduction	122
	7.1.4.2 Development over time	122
	7.1.4.3 Comparing groups	124
	7.2 Relationships with other variables	125
	7.2.1 “Traditional” methods	125
	7.2.2 Example	126
	7.2.3 Sophisticated methods	126
	7.2.4 Example	128
	7.2.4.1 Generalized estimating equations	128
	7.2.4.2 Mixed model analysis	133
	7.2.5 Comparison between GEE analysis and mixed model analysis	136
	7.2.6 Alternative models	138
	7.2.7 Comments	139
8	Categorical and “count” outcome variables	141
	8.1 Categorical outcome variables	141
	8.1.1 Two measurements	141
	8.1.2 More than two measurements	142
	8.1.3 Comparing groups	143
	8.1.4 Example	143

x	Contents	
	8.1.5 Relationship with other variables	146
	8.1.5.1 “Traditional” methods	146
	8.1.5.2 Example	147
	8.1.5.3 Sophisticated methods	148
	8.1.5.4 Example	148
	8.2 “Count” outcome variables	153
	8.2.1 Example	153
	8.2.1.1 Introduction	153
	8.2.1.2 GEE analysis	154
	8.2.1.3 Mixed model analysis	158
	8.2.2 Comparison between GEE analysis and mixed model analysis	160
	8.3 Comments	162
9	Analysis of experimental studies	163
	9.1 Introduction	163
	9.2 Continuous outcome variables	165
	9.2.1 Experimental models with only one follow-up measurement	165
	9.2.1.1 Example	170
	9.2.2 Experimental studies with more than one follow-up measurement	179
	9.2.2.1 Simple analysis	180
	9.2.2.2 Summary statistics	182
	9.2.2.3 MANOVA for repeated measurements	183
	9.2.2.4 MANOVA for repeated measurements adjusted for the baseline value	184
	9.2.2.5 Sophisticated analysis	187
	9.2.3 Conclusion	200
	9.3 Dichotomous outcome variables	201
	9.3.1 Introduction	201
	9.3.2 Simple analysis	202
	9.3.3 Sophisticated analysis	203
	9.3.4 Other approaches	208
	9.4 Comments	210
10	Missing data in longitudinal studies	212
	10.1 Introduction	212
	10.2 Ignorable or informative missing data?	214
	10.3 Example	215
	10.3.1 Generating datasets with missing data	215
	10.3.2 Analysis of determinants for missing data	216
	10.4 Analysis performed on datasets with missing data	218
	10.4.1 Example	219

© in this web service Cambridge University Press www.cambridge.org

xii	Contents	
	12.4.4 SPSS	260
	12.4.5 MLwiN	264
	12.4.6 Overview	266
	12.5 Mixed model analysis with dichotomous outcome variables	267
	12.5.1 Introduction	267
	12.5.2 Stata	268
	12.5.3 SAS	268
	12.5.4 R	271
	12.5.5 SPSS	274
	12.5.6 MLwiN	278
	12.5.7 Overview	280
	12.6 Categorical and “count” outcome variables	281
	12.7 The “adjustment for covariance approach”	282
	12.7.1 Example	283
13	One step further	292
	13.1 Introduction	292
	13.2 Outcome variables with upper or lower censoring	292
	13.2.1 Introduction	292
	13.2.2 Example	294
	13.2.3 Remarks	300
	13.3 Classification of subjects with different developmental trajectories	301
	<i>References</i>	305
	<i>Index</i>	316

Preface

The most important feature of this book is the word “applied” in the title. This implies that the emphasis of this book lies more on the application of statistical techniques for longitudinal data analysis and not so much on the mathematical background. In most other books on the topic of longitudinal data analysis, the mathematical background is the major issue, which may not be surprising since (nearly) all the books on this topic have been written by statisticians. Although statisticians fully understand the difficult mathematical material underlying longitudinal data analysis, they often have difficulty in explaining this complex material in a way that is understandable for the researchers who have to use the technique or interpret the results. Therefore, this book is not written by a statistician, but by an epidemiologist. In fact, an epidemiologist is not primarily interested in the basic (difficult) mathematical background of the statistical methods, but in finding the answer to a specific research question; the epidemiologist wants to know how to apply a statistical technique and how to interpret the results. Owing to their different basic interests and different level of thinking, communication problems between statisticians and epidemiologists are quite common. This, in addition to the growing interest in longitudinal studies, initiated the writing of this book: a book on longitudinal data analysis, which is especially suitable for the “non-statistical” researcher (e.g. the epidemiologist). The aim of this book is to provide a practical guide on how to handle epidemiological data from longitudinal studies. The purpose of this book is to build a bridge over the communication gap that exists between statisticians and epidemiologists when addressing the complicated topic of longitudinal data analysis.

Acknowledgements

I am very grateful to all my colleagues and students who came to me with (mostly) practical questions on longitudinal data analysis. This book is based on all those questions. Furthermore, I would like to thank Dick Bezemer, Maarten Boers, Bernard Uitdehaag, Wieke de Vente, Michiel de Boer, and Martijn Heymans who critically read preliminary drafts of some chapters and provided very helpful comments.