

Cambridge University Press
978-1-107-01956-0 - Case-Control Studies
Ruth H. Keogh and D. R. Cox
Frontmatter
[More information](#)

Case-Control Studies

The case-control approach is a powerful method for investigating factors that may explain a particular event. It is extensively used in epidemiology to study disease incidence, one of the best-known examples being the investigation by Bradford Hill and Doll of the possible connection between cigarette smoking and lung cancer. More recently, case-control studies have been increasingly used in other fields, including sociology and econometrics.

With a particular focus on statistical analysis, this book is ideal for applied and theoretical statisticians wanting an up-to-date introduction to the field. It covers the fundamentals of case-control study design and analysis as well as more recent developments, including two-stage studies, case-only studies, and methods for case-control sampling in time. The latter have important applications in large prospective cohorts that require case-control sampling designs to make efficient use of resources. More theoretical background is provided in an appendix for those new to the field.

RUTH H. KEOGH is a Lecturer in the Department of Medical Statistics at the London School of Hygiene and Tropical Medicine.

D. R. COX is one of the world's pre-eminent statisticians. His work on the proportional hazards regression model is one of the most-cited and most influential papers in modern statistics. In 2010 he won the Copley Medal of the Royal Society 'for his seminal contributions to the theory and application of statistics'. He is currently an Honorary Fellow at Nuffield College, Oxford.

Cambridge University Press
978-1-107-01956-0 - Case-Control Studies
Ruth H. Keogh and D. R. Cox
Frontmatter
[More information](#)

INSTITUTE OF MATHEMATICAL STATISTICS
MONOGRAPHS

Editorial Board

D. R. Cox (University of Oxford)
A. Agresti (University of Florida)
B. Hambly (University of Oxford)
S. Holmes (Stanford University)
X.-L. Meng (Harvard University)

IMS Monographs are concise research monographs of high quality on any branch of statistics or probability of sufficient interest to warrant publication as books. Some concern relatively traditional topics in need of up-to-date assessment. Others are on emerging themes. In all cases the objective is to provide a balanced view of the field.

Cambridge University Press
978-1-107-01956-0 - Case-Control Studies
Ruth H. Keogh and D. R. Cox
Frontmatter
[More information](#)

Case-Control Studies

RUTH H. KEOGH
*London School of Hygiene
and Tropical Medicine*

D. R. COX
Nuffield College, Oxford



CAMBRIDGE
UNIVERSITY PRESS

Cambridge University Press
 978-1-107-01956-0 - Case-Control Studies
 Ruth H. Keogh and D. R. Cox
 Frontmatter
[More information](#)

CAMBRIDGE
 UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

Published in the United States of America by Cambridge University Press, New York

Cambridge University Press is a 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/9781107019560

© Ruth H. Keogh and D.R. Cox 2014

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 published 2014

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

Keogh, Ruth H., 1979– author.

Case-control studies / Ruth H. Keogh, D.R. Cox.

p. ; cm. – (Institute of Mathematical Statistics monographs ; 4)

Includes bibliographical references and index.

ISBN 978-1-107-01956-0 (hardback)

I. Cox, D.R. (David Roxbee), author. II. Title. III. Series: Institute of Mathematical Statistics monographs ; 4.

[DNLM: 1. Epidemiologic Methods. WA 950]

RA652

614.4 – dc23 2013047249

ISBN 978-1-107-01956-0 Hardback

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.

Contents

<i>Preface</i>	ix
Preamble	1
Notes	7
1 Introduction to case-control studies	8
1.1 Defining a case-control study	8
1.2 Measuring association: the odds ratio	12
1.3 Methods for controlling confounding	13
1.4 Temporal aspects	16
1.5 Further details on the sampling of cases and controls	20
1.6 Bias in case-control studies	23
1.7 Use of case-control studies	26
Notes	29
2 The simplest situation	31
2.1 Preliminaries	31
2.2 Measuring association	34
2.3 Statistical analysis	36
2.4 Model formulation	42
2.5 Design implications for the numbers of cases and controls	44
2.6 Logistic dependence	45
2.7 Explanatory variable with more than two levels	47
2.8 Conditional odds ratios: combining estimates	48
2.9 A special feature of odds ratios	51
2.10 Introducing time	54
2.11 Causality and case-control studies	57
Notes	59
3 Matched case-control studies	62
3.1 Preliminaries	62
3.2 Choice of matched controls	63

vi	<i>Contents</i>	
3.3	Analysis of matched data	65
3.4	Discussion	70
3.5	Alternative analyses of matched data	71
3.6	Overmatching	73
3.7	Further theoretical points	75
3.8	Generalizations: several controls per case	76
3.9	Logistic dependence	78
3.10	Introducing time	80
	Notes	81
4	A general formulation	83
4.1	Preliminaries	83
4.2	Logistic regression for case-control studies	84
4.3	Extensions	89
4.4	More on non-collapsibility	93
4.5	Non-logistic models	97
4.6	Statistical efficiency	99
4.7	Estimating other quantities	102
4.8	Use of the distribution of exposure	104
4.9	Beyond case-control studies: a more general view	105
	Notes	108
5	Case-control studies with more than two outcomes	111
5.1	Preliminaries	111
5.2	Polychotomous case-control studies	111
5.3	Polychotomous studies with matching	117
5.4	More than one control group	119
5.5	Ordered outcome groups	120
5.6	Case-only studies	122
	Notes	130
6	Special sampling designs	132
6.1	Preliminaries	132
6.2	Two-stage sampling design	133
6.3	Likelihood analyses for two-stage studies	143
6.4	Discussion	148
6.5	Special sampling in matched case-control studies	150
6.6	Case-control studies using family members	152
6.7	Augmenting case-control data using family members	155
	Notes	157
7	Nested case-control studies	160
7.1	Preliminaries	160
7.2	The nested case-control design	162

<i>Contents</i>		vii
7.3	Analysis of nested case-control studies	165
7.4	Further comments on analysis	167
7.5	Some extensions	169
7.6	More elaborate sampling procedures	170
7.7	Further comments on the sampling of controls	175
7.8	Some alternative models	176
7.9	Extended use of controls	180
7.10	Estimating absolute risk	182
7.11	Using full-cohort information	184
	Notes	186
8	Case-subcohort studies	191
8.1	Preliminaries	191
8.2	Essential outline of the case-subcohort design	193
8.3	Analysis disregarding time	194
8.4	Case-subcohort study using event times	198
8.5	Size of the subcohort	202
8.6	Stratified case-subcohort studies	204
8.7	Some further approaches	207
8.8	Comparison between nested case-control and case-subcohort studies	208
	Notes	209
9	Misclassification and measurement error	212
9.1	Preliminaries	212
9.2	Misclassification in an exposure variable	214
9.3	Misclassification of cases and controls	218
9.4	Misclassification in matched case-control studies	220
9.5	Error in continuous exposures	221
9.6	Correcting for error in a continuous exposure	228
	Notes	236
10	Synthesis of studies	240
10.1	Preliminaries	240
10.2	Some large-sample methods	241
10.3	Graphical representation	243
10.4	Study heterogeneity	244
10.5	An exact analysis	248
10.6	Combining studies of different type	250
	Notes	250
Appendix	A theoretical diversion	252
A.1	Preliminaries	252
A.2	Multidimensional parameter	253

A.3	The maximum-likelihood estimate	254
A.4	Significance tests	256
A.5	More detailed theory	256
A.6	Pseudo-likelihood	256
A.7	Weighted estimation	258
A.8	Bayesian analysis	258
	Notes	261
	<i>References</i>	262
	<i>Index</i>	281

Preface

The retrospective case-control approach provides a powerful method for studying rare events and their dependence on explanatory features. The method is extensively used in epidemiology to study disease incidence, one of the best known and early examples being the investigation by Bradford Hill and Doll of the possible impact of smoking and pollution on lung cancer. More recently the approach has been ever more widely used, by no means only in an epidemiological setting. There have also been various extensions of the method.

A definitive account in an epidemiological context was given by Breslow and Day in 1980 and their book remains a key source with many important insights. Our book is addressed to a somewhat more statistical readership and aims to cover recent developments. There is an emphasis on the analysis of data arising in case-control studies, but we also focus in a number of places on design issues. We have tried to make the book reasonably self-contained; some familiarity with simple statistical methods and theory is assumed, however. Many methods described in the book rely on the use of maximum likelihood estimation, and the extension of this to pseudo-likelihoods is required in the later chapters. We have therefore included an appendix outlining some theoretical details.

There is an enormous statistical literature on case-control studies. Some of the most important fundamental work appeared in the late 1970s, while the later 1980s and the 1990s saw the establishment of methods for case-control sampling in time. The latter have important applications in large prospective cohorts which collect large amounts of information, for example biological samples, but which require case-control sampling designs to make efficient use of resources. There continue to appear in the literature many innovations in case-control study design and, in particular, methodology covering a wide range of areas.

We hope that the book will be useful both to applied and to theoretical statisticians wanting an introduction to the field. Parts of it might be useful

as a basis for a short course for postgraduate students, although we have not written with that use specifically in mind.

The EPIC-Norfolk study provided some examples; we are grateful for permission to use this data and thank the staff and participants of the study.

We are grateful for very helpful comments on an initial draft from the following colleagues and friends: Amy Berrington de Gonzalez, Vern Farewell, Chris Keogh and Ian White. We are very grateful also to Diana Gillooly at Cambridge University Press for her encouragement and helpful advice and to Susan Parkinson for her meticulous and constructive copy-editing.

Ruth Keogh and David Cox