VETERINARY EPIDEMIOLOGIC RESEARCH

‘*A comprehensive text*
for the discipline’

www.upei.ca/ver

VETERINARY EPIDEMIOLOGIC RESEARCH

Ian Dohoo
Professor Emeritus of Epidemiology
Department of Health Management
University of Prince Edward Island
Charlottetown PEI Canada

Wayne Martin
Professor Emeritus of Epidemiology
Department of Population Medicine
University of Guelph
Guelph Ontario Canada

Henrik Stryhn
Associate Professor of Biostatistics
Department of Health Management
University of Prince Edward Island
Charlottetown PEI Canada

VER inc. Published by VER Inc. · Charlottetown · Prince Edward Island · Canada

This book was prepared using the open source office software—OpenOffice (http://www.openoffice.org/)

The book is printed on recycled paper (50% post-consumer waste).

Editor/Proofreader/Compositor/Publication Coordinator: S Margaret McPike
Cover designer: Gregory Mercier
Printer: Friesens Manitoba Canada

© 2014, by VER Inc. All rights reserved. This book is protected by copyright. No part of this book may be reproduced, stored in a retrieval system or transmitted, in any form or by any means—electronic, mechanical, photocopy, recording or otherwise—without the written permission of the publisher. For information write:
VER Inc., Box 491, Charlottetown, Prince Edward Island, Canada, C1A 7L1

Third printing May 2014

Printed in Canada

10 9 8 7 6 5 4 3 2 1

National Library of Canada Cataloguing in Publication
Veterinary epidemiologic research / Ian Dohoo ... [et al].
Includes index.
ISBN 978-0-919013-60-5
1. Veterinary epidemiology – Research – Textbooks
I. Dohoo, Ian Robert

Library catalogue information available from www.upei.ca/ver

Care has been taken to confirm the accuracy of the information presented. Nevertheless, the authors, editor and publisher are not responsible for errors or omissions or for any consequences from application of the information in this book and make no warranty, express or implied, with respect to the content of the publication.

Dedication

This text is dedicated to all of the graduate students who have challenged and inspired us throughout our careers, and to our families who have supported us, especially during the writing of this text.

CONTENTS

Foreword		xv
Preface		xvi
Acknowledgements		xviii
1	**INTRODUCTION AND CAUSAL CONCEPTS**	**1**
1.1	Introduction	2
1.2	A brief history of multiple causation concepts	2
1.3	A brief history of scientific inference	5
1.4	Key components of epidemiologic research	8
1.5	Seeking causes	9
1.6	Models of causation	10
1.7	Counterfactual concepts of causation for a single exposure	17
1.8	Experimental versus observational evidence of causation	20
1.9	Constructing a causal diagram	21
1.10	Causal criteria	23
2	**SAMPLING**	**33**
2.1	Introduction	34
2.2	Non-probability sampling	37
2.3	Probability sampling	37
2.4	Simple random sample	38
2.5	Systematic random sample	38
2.6	Stratified random sample	38
2.7	Cluster sampling	39
2.8	Multistage sampling	40
2.9	Targeted (risk-based) sampling	41
2.10	Analysis of survey data	42
2.11	Sample-size determination	46
2.12	Sampling to detect disease	54
3	**QUESTIONNAIRE DESIGN**	**57**
3.1	Introduction	58
3.2	Designing the question	60
3.3	Open question	61
3.4	Closed question	61
3.5	Wording the question	65
3.6	Structure of questionnaires	65
3.7	Pre-testing questionnaires	66
3.8	Validation	67
3.9	Response rate	67
3.10	Data-coding and editing	68
4	**MEASURES OF DISEASE FREQUENCY**	**73**
4.1	Introduction	74
4.2	Count, proportion, odds and rate	74
4.3	Incidence	75
4.4	Calculating risk	76

4.5	Calculating incidence rates	77
4.6	Relationship between risk and rate	79
4.7	Prevalence	80
4.8	Mortality statistics	81
4.9	Other measures of disease frequency	81
4.10	Standard errors and confidence intervals	83
4.11	Standardisation of risks and rates	85
5	**SCREENING AND DIAGNOSTIC TESTS**	**91**
5.1	Introduction	92
5.2	Attributes of the test *per se*	92
5.3	The ability of a test to detect disease or health	100
5.4	Predictive values	103
5.5	Interpreting test results that are measured on a continuous scale	105
5.6	Using multiple tests	111
5.7	Evaluation of diagnostic tests	114
5.8	Evaluation when there is no gold standard	116
5.9	Other considerations in test evaluation	122
5.10	Sample size requirements	123
5.11	Herd-level testing	123
5.12	Use of pooled samples	127
6	**MEASURES OF ASSOCIATION**	**135**
6.1	Introduction	136
6.2	Measures of association	137
6.3	Measures of effect	139
6.4	Study design and measures of association	143
6.5	Hypothesis testing and confidence intervals	143
6.6	Multivariable estimation of measures of association	148
7	**INTRODUCTION TO OBSERVATIONAL STUDIES**	**151**
7.1	Introduction	152
7.2	A unified approach to study design	154
7.3	Descriptive studies	156
7.4	Observational studies	157
7.5	Cross-sectional studies	158
7.6	Repeated cross-sectional versus cohort studies	162
8	**COHORT STUDIES**	**167**
8.1	Introduction	168
8.2	Study group	169
8.3	The exposure	171
8.4	Ensuring exposed and non-exposed groups are comparable	174
8.5	Follow-up period	175
8.6	Measuring the outcome	175
8.7	Analysis	176
8.8	Reporting of cohort studies	177

9 CASE-CONTROL STUDIES 181
9.1 Introduction 182
9.2 The study base 182
9.3 The case series 183
9.4 Principles of control selection 184
9.5 Selecting controls in risk-based designs 185
9.6 Selecting controls in rate-based designs 187
9.7 Other sources of controls 190
9.8 The number of controls per case 193
9.9 The number of control groups 193
9.10 Exposure and covariate assessment 193
9.11 Keeping the cases and controls comparable 193
9.12 Analysis of case-control data 194
9.13 Reporting guidelines for case-control studies 195

10 HYBRID STUDY DESIGNS 199
10.1 Introduction 200
10.2 Case-crossover studies 200
10.3 Case-case studies 203
10.4 Case-series studies 204
10.5 Case-cohort studies 206
10.6 Case-only studies 208
10.7 Two-stage sampling designs 209

11 CONTROLLED STUDIES 213
11.1 Introduction 214
11.2 Stating the objectives 215
11.3 The study group 216
11.4 Allocation of study subjects 221
11.5 Specifying the intervention 225
11.6 Masking (blinding) 225
11.7 Follow-up/compliance 226
11.8 Measuring the outcome 227
11.9 Analysis 227
11.10 Clinical trial designs for prophylaxis of communicable organisms 230
11.11 Ethical considerations 233
11.12 Reporting of clinical trials 235

12 VALIDITY IN OBSERVATIONAL STUDIES 243
12.1 Introduction 244
12.2 Selection bias 244
12.3 Examples of selection bias 249
12.4 Reducing selection bias 254
12.5 Information bias 255
12.6 Bias from misclassification 257
12.7 Validation studies to correct misclassification 263
12.8 Measurement error 264
12.9 Errors in surrogate measures of exposure 265

12.10 The impact of information bias on sample size 266

13 CONFOUNDING: DETECTION AND CONTROL 271
13.1 Introduction 272
13.2 Control of confounding prior to data analysis 275
13.3 Matching on confounders 276
13.4 Matching using propensity scores 281
13.5 Detection of confounding 283
13.6 Analytic control of confounding 288
13.7 Other approaches to control confounding and estimate causal effects 295
13.8 Multivariable modelling to control confounding 301
13.9 Instrumental variables to control confounding 302
13.10 External adjustment and sensitivity analysis for unmeasured confounders 304
13.11 Understanding causal relationships 306
13.12 Summary of effects of extraneous variables 315

14 LINEAR REGRESSION 323
14.1 Introduction 324
14.2 Regression analysis 324
14.3 Hypothesis testing and effect estimation 326
14.4 Nature of the X-variables 333
14.5 Detecting highly correlated (collinear) variables 338
14.6 Detecting and modelling interaction 340
14.7 Causal interpretation of a multivariable linear model 341
14.8 Evaluating the least squares model 344
14.9 Evaluating the major assumptions 349
14.10 Assessment of individual observations 356
14.11 Time-series data 360

15 MODEL-BUILDING STRATEGIES 365
15.1 Introduction 366
15.2 Steps in building a model 367
15.3 Building a causal model 367
15.4 Reducing the number of predictors 368
15.5 The problem of missing values 374
15.6 Effects of continuous predictors 375
15.7 Identifying interaction terms of interest 381
15.8 Building the model 383
15.9 Evaluate the reliability of the model 388
15.10 Presenting the results 390

16 LOGISTIC REGRESSION 395
16.1 Introduction 396
16.2 The logistic model 396
16.3 Odds and odds ratios 397
16.4 Fitting a logistic regression model 398
16.5 Assumptions in logistic regression 399
16.6 Likelihood ratio statistics 400

16.7 Wald tests 401
16.8 Interpretation of coefficients 402
16.9 Assessing interaction and confounding 405
16.10 Model-building 408
16.11 Generalised linear models 408
16.12 Evaluating logistic regression models 410
16.13 Sample size considerations 421
16.14 Exact logistic regression 421
16.15 Conditional logistic regression for matched studies 422

17 MODELLING ORDINAL AND MULTINOMIAL DATA 427
17.1 Introduction 428
17.2 Overview of models 429
17.3 Multinomial logistic regression 431
17.4 Modelling ordinal data 436
17.5 Proportional odds model (constrained cumulative logit model) 437
17.6 Adjacent-category model 441
17.7 Continuation-ratio model 443

18 MODELLING COUNT AND RATE DATA 445
18.1 Introduction 446
18.2 The Poisson distribution 447
18.3 Poisson regression model 448
18.4 Interpretation of coefficients 449
18.5 Evaluating Poisson regression models 451
18.6 Negative binomial regression 454
18.7 Problems with zero counts 461

19 MODELLING SURVIVAL DATA 467
19.1 Introduction 468
19.2 Non-parametric analyses 473
19.3 Actuarial life tables 473
19.4 Kaplan-Meier estimate of survivor function 475
19.5 Nelson-Aalen estimate of cumulative hazard 478
19.6 Statistical inference in non-parametric analyses 479
19.7 Survivor, failure and hazard functions 480
19.8 Semi-parametric analyses 485
19.9 Parametric models 503
19.10 Accelerated failure time models 507
19.11 Frailty models and clustering 510
19.12 Multiple outcome event data 517
19.13 Discrete-time survival analysis 518
19.14 Sample sizes for survival analyses 522

20 INTRODUCTION TO CLUSTERED DATA 529
20.1 Introduction 530
20.2 Clustering arising from the data structure 530
20.3 Effects of clustering 536

20.4 Simulation studies on the impact of clustering 540
20.5 Introduction to methods for dealing with clustering 542

21 MIXED MODELS FOR CONTINUOUS DATA 553
21.1 Introduction 554
21.2 Linear mixed model 555
21.3 Random slopes 560
21.4 Contextual effects 564
21.5 Statistical analysis of linear mixed models 565

22 MIXED MODELS FOR DISCRETE DATA 579
22.1 Introduction 580
22.2 Logistic regression with random effects 580
22.3 Poisson regression with random effects 584
22.4 Generalised linear mixed model 587
22.5 Statistical analysis of GLMMs 593
22.6 Summary remarks on analysis of discrete clustered data 603

23 REPEATED MEASURES DATA 607
23.1 Introduction to repeated measures data 608
23.2 Univariate and multivariate approaches to repeated measures data 611
23.3 Linear mixed models with correlation structure 616
23.4 Mixed models for discrete repeated measures data 624
23.5 Generalised estimating equations 627

24 INTRODUCTION TO BAYESIAN ANALYSIS 637
24.1 Introduction 638
24.2 Bayesian analysis 638
24.3 Markov chain Monte Carlo (MCMC) estimation 642
24.4 Statistical analysis based on MCMC estimation 646
24.5 Extensions of Bayesian and MCMC Modelling 651

25 ANALYSIS OF SPATIAL DATA: INTRODUCTION AND VISUALISATION 663
25.1 Introduction 664
25.2 Spatial data 664
25.3 Spatial data analysis 667
25.4 Additional topics 673

26 ANALYSIS OF SPATIAL DATA 679
26.1 Introduction 680
26.2 Issues specific to statistical analysis of spatial data 680
26.3 Exploratory spatial analysis 682
26.4 Global spatial clustering 690
26.5 Localised spatial cluster detection 697
26.6 Space-time association 700
26.7 Modelling 704

27 CONCEPTS OF INFECTIOUS DISEASE EPIDEMIOLOGY 715
27.1 Introduction 716
27.2 Infection vs disease 718
27.3 Transmission 719
27.4 Mathematical modelling of infectious disease transmission 721
27.5 Estimating R_0 and other infectious disease parameters 725

28 SYSTEMATIC REVIEWS AND META-ANALYSIS 739
28.1 Introduction 740
28.2 Narrative reviews 740
28.3 Systematic Reviews 741
28.4 Meta-analysis – Introduction 745
28.5 Fixed- and random-effects models 746
28.6 Presentation of results 749
28.7 Heterogeneity 750
28.8 Publication bias 758
28.9 Influential studies 760
28.10 Outcome scales and data issues 760
28.11 Meta-analysis of observational studies 764
28.12 Meta-analysis of diagnostic tests 766
28.13 Use of meta-analysis 766

29 ECOLOGICAL AND GROUP-LEVEL STUDIES 773
29.1 Introduction 774
29.2 Rationale for group level studies 774
29.3 Types of ecologic variable 775
29.4 Issues related to modelling approaches in ecologic studies 776
29.5 Issues related to inferences 778
29.6 Sources of ecologic bias 778
29.7 Non-ecologic group-level studies 782

30 A STRUCTURED APPROACH TO DATA ANALYSIS 789
30.1 Introduction 790
30.2 Data-collection sheets 790
30.3 Data coding 791
30.4 Data entry 791
30.5 Keeping track of files 792
30.6 Keeping track of variables 792
30.7 Program mode versus interactive processing 793
30.8 Data-editing 794
30.9 Data verification 795
30.10 Data processing—outcome variable(s) 795
30.11 Data processing—predictor variables 796
30.12 Data processing—multilevel data 796
30.13 Unconditional associations 796
30.14 Keeping track of your analyses 797

31 DESCRIPTION OF DATASETS 799

FOREWORD

When I was asked to write the Foreword for the first edition of Veterinary Epidemiologic Research six years ago, I saw publication of the book as a major step in the continuing maturation and growth of the field of veterinary epidemiology. Finally we would have a single source that we could go to for information on how to use and interpret epidemiologic methods – confident that investigational designs and analytical techniques were comprehensively and accurately described, with appropriate guidance on their strengths and limitations.

Six years on, that promise has been amply fulfilled, and the book has found a place in the centre of the bookshelf of epidemiologists throughout the world. It can either be pulled out quickly to check a point, or an entire chapter can be read carefully and methodically to widen the horizons and increase the skills of both the budding epidemiologist and the experienced practitioner of the field. We have all gained benefit from the use of the book.

The first edition used the wide investigational experience of the authors and brought together in one place most of the methods we use in our work. This Second Edition adds chapters covering a number of areas that have grown considerably in importance over the last decade, and justified detailed coverage in this new edition. I am particularly pleased to see that spatial epidemiology, survival analysis and Bayesian analysis are now covered in detail, and that a number of other areas have been strengthened.

Epidemiologists are practical people, and they benefit from the opportunity to learn through practical examples. This book is very strong in providing real-world examples of the various methods and how to use them, which makes it easy to learn the 'how' of epidemiology as well as the 'why'.

For the first edition, I wrote "This is a comprehensive text for the discipline of veterinary epidemiology, written by authors who have the standing to provide wise and insightful guidance on epidemiological research methods for the novice and the expert alike". This second edition is even more comprehensive, and has also been refined through the experience of using it for teaching and for practical application in epidemiologic investigations and research.

I commend it to all who are involved in epidemiology, as the book you absolutely must have on your shelf.

Professor Roger S Morris CNZM, MVSc, PhD, FACVSc, FAmerCE, FRSNZ
Massey University EpiCentre, Palmerston North, New Zealand

PREFACE

The 2nd edition of *Veterinary Epidemiologic Research* arrives 6 years after the publication of the 1st edition (2003). One major motivation for the preparation of this 2nd edition was a desire to expand the coverage to include several new topic areas which are becoming increasingly important for veterinary epidemiologists. In addition, our use of the 1st edition in graduate level courses identified a number of areas in which we thought more detail, or clarification, would be beneficial. Finally, we believe that it would be helpful to readers if referencing of source material was more thorough throughout the text, so we have endeavoured to do that. As with the 1st edition, the text grew as we wrote it. We have tried to achieve a balance between being comprehensive and making the material accessible to researchers and graduate students.

Specific encouragement to prepare a 2nd edition came from the many users of the text who have provided us with numerous positive comments about the value of the book. The use of the text around the world certainly exceeded our expectation and we are grateful to you, the readers, who have made it so successful. We have also appreciated the many constructive suggestions from users of the 1st edition, and have tried to incorporate many of these into this edition.

Before reviewing the content of the text we thought we should address the three most common questions we received about the 1st edition.

- Why are the 2X2 tables oriented the way they are (disease in rows, exposure in columns)? The answer to this was actually included in the 1st edition (page 671—Glossary and Terminology). We feel that the text Modern Epidemiology (Rothman *et al*, 2008) is a key reference text in the field of epidemiology and have chosen to be consistent with their format.
- Why does the title use the word 'epidemiologic' instead of 'epidemiological'? According to "Scientific Style and Format—The CSE manual for Authors, Editors and Publishers" (Council of Science Editors—Style Manual Committee, 2006), either is acceptable. Once again, we deferred to a text which we felt was seminal in the development of epidemiologic methods "Epidemiologic Research: Principles and Quantitative Methods" (Kleinbaum *et al*, 1982).
- Why does the title on the spine of the book run the opposite direction from all other books on my bookshelf? This was just an oversight on our part. We have conformed with tradition for this edition.

This text focuses on both design and analytic issues. The general structure is the same as the 1st edition. Chapters 1 through 6 focus on basic epidemiologic principles. The biggest change from the 1st edition is the substantial expansion of the material covered in Chapter 5 (Screening and Diagnostic Tests) and this reflects the substantial growth in methods for evaluating tests that has taken place in the last 6 years. All other chapters in this section have undergone modest expansion.

As in the 1st edition, Chapters 7-11 focus on study design issues for observational studies and controlled trails. The most substantial changes in this section are: much more extensive referencing in all chapters and a substantial expansion of the material in Chapter 11 (Controlled Studies). In addition, there has been much discussion over the past decade about the need for epidemiologists to thoroughly report their research findings (and by doing so this will help ensure high quality study designs in the future) and we have cited the summary recommendations in these chapters.

Once again, Chapters 14-19 cover a range of multivariable models. In all chapters, new information has been incorporated, but the largest change is in Chapter 19 (Modelling Survival Data). The analysis of time-to-event data is expanding rapidly in veterinary epidemiology and we wanted to provide a more complete coverage of this topic.

As in the 1st edition, Chapters 20-23 deal with the issue of clustered data, but the material has been updated and reorganised considerably. Chapter 20 includes a more substantial coverage of the impact of clustering and incorporates some of the material previously in Chapter 23. Chapter 23 now provides a much more thorough description of methods for analysing repeated measures data.

Chapters 24-30 include both new material and updates to existing chapters. New topics covered include: Introduction to Bayesian Analysis (Chapter 24—contributed by William Browne and Henrik Stryhn), two chapters on presenting and analysing spatial data (Chapter 25 and 26—contributed by Javier Sanchez and Dirk Pfeiffer), and Concepts of Infectious Disease Epidemiology (Chapter 27—contributed by Ian Dohoo and Graham Medley). There has also been a substantial expansion in the coverage of meta-analysis (Chapter 28) compared with the 1st edition.

In order to make some room for the new material, some changes and section/topic deletions were required. The size (height and width) of the text has been slightly increased and the sections in which we provided 'program files' and a complete bibliography for the text have been removed. Both of these items are available at www.upei.ca/ver.

As we did in the 1st edition, we have made extensive use of examples. All of the datasets used in these examples are described in the text (Chapter 31) and are available through the book's website. Virtually all of the examples have been analysed using the statistical program Stata™—a program which provides a unique combination of statistical and epidemiological tools and which we use extensively in our teaching. Version 10 of Stata was used throughout although Version 11 was released shortly before the book went to print. Updates to some of the program files which take advantage of features in Version 11 will be added to the website as they become available.

We hope that you find this second edition of *Veterinary Epidemiologic Research* useful in your studies and your research.

REFERENCES

Council of Science Editors - Style Manual Committee. Scientific style and format: The CSE manual for authors, editors and publishers. 7th ed. Reston (VA): The Council; 2006.

Kleinbaum D, Kupper L, Morgenstern H. Epidemiologic Research: Principles and Quantitative Methods. London: Lifetime Learning Publications; 1982.

Rothman K, Greenland S, Lash T. Modern Epidemiology, 3rd Ed. Philadelphia: Lippincott Williams & Wilkins; 2008.

ACKNOWLEDGEMENTS

As noted in the Preface, many individuals have provided constructive feedback on the 1st edition of this text, and we are grateful for that feedback. However, one individual—Gary Anderson of Melbourne University—stands out in this regard. Over the past 5 years, Gary provided a wealth of valuable feedback, on VER 1, which we very much appreciated. As is usually the case, if you do a good job at something, you get given more work (no good work should go unpunished!)—and this is true for Gary. Thus, as we planned, and then wrote, the 2nd edition we asked Gary to review many of the analytical chapters and he has kindly done so in a very thorough and constructive manner. We are deeply indebted to him for this valuable contribution.

We believe the value of this book has been greatly enhanced by the provision of a substantial number of 'real-life' datasets, both for use in the examples and the sample problems. Rather than include a list of the contributors (it is becoming a very long list), we would like to collectively acknowledge their contribution and point out that they are all identified as contributors in Chapter 31 in which the datasets are described.

Putting this book together has been both a learning experience and a lot of fun. We chose to prepare this book using open source software (OpenOffice—www.openoffice.org) and in the process have learned a lot about how to use a general purpose office package for preparing a text. We are deeply indebted to Margaret McPike who has done all of the editing, proofreading, and formatting of this text. As with the 1st edition, we chose to publish this book ourselves, which entailed taking complete responsibility for these activities. Margaret has dedicated herself to this task. All of the credit for layout of the book, and the clarity of the format, goes to Margaret.

We would like to thank Gregory Mercier, who did the graphic design work for the cover of the 1st edition of this book cover and updated it for the 2nd edition. We would also like to thank Bill Rising of Stata Corp. who reviewed a number of the analytical methods chapters and provided some very constructive feedback, particularly in terms of the program files (Stata -do- files) which are available on the book's website.