



EPIDEMIOLOGY

Petra Büttner
Reinhold Müller

OXFORD

EPIDEMIOLOGY

Petra Büttner
Reinhold Müller

OXFORD
UNIVERSITY PRESS
AUSTRALIA & NEW ZEALAND

OXFORD
UNIVERSITY PRESS

Oxford University Press is a department of the University of Oxford. It furthers the University's objective of excellence in research, scholarship, and education by publishing worldwide. Oxford is a registered trademark of Oxford University Press in the UK and in certain other countries.

Published in Australia by
Oxford University Press
253 Normanby Road, South Melbourne, Victoria 3205, Australia

© Petra Büttner and Reinhold Muller 2011

The moral rights of the author have been asserted

First published 2011

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, without the prior permission in writing of Oxford University Press, or as expressly permitted by law, by licence, or under terms agreed with the appropriate reprographics rights organisation. Enquiries concerning reproduction outside the scope of the above should be sent to the Rights Department, Oxford University Press, at the address above.

You must not circulate this work in any other form and you must impose this same condition on any acquirer.

National Library of Australia Cataloguing-in-Publication data

Author: Büttner, Petra.

Title: Epidemiology / Petra Büttner, Reinhold Muller.

ISBN: 9780195573893 (pbk.)

Notes: Includes bibliographical references and index.

Subjects: Epidemiology.

Other Authors/Contributors: Muller, Reinhold, 1958-

Dewey Number: 614.4

Reproduction and communication for educational purposes

The Australian *Copyright Act 1968* (the Act) allows a maximum of one chapter or 10% of the pages of this work, whichever is the greater, to be reproduced and/or communicated by any educational institution for its educational purposes provided that the educational institution (or the body that administers it) has given a remuneration notice to Copyright Agency Limited (CAL) under the Act.



For details of the CAL licence for educational institutions contact:

Copyright Agency Limited
Level 15, 233 Castlereagh Street
Sydney NSW 2000
Telephone: (02) 9394 7600
Facsimile: (02) 9394 7601
Email: info@copyright.com.au

Edited by Carolyn Pike

Text design by eggplant communications

Typeset by diacriTech, Chennai, India

Proofread by Anne Mulvaney

Indexed by Russell Brooks

Printed by Sheck Wah Tong Printing Press Ltd

Links to third party websites are provided by Oxford in good faith and for information only. Oxford disclaims any responsibility for the materials contained in any third party website referenced in this work.

Contents

| | |
|---|-------|
| <i>List of Figures</i> | xii |
| <i>List of Tables</i> | xv |
| <i>Preface</i> | xviii |
| <i>Acknowledgments</i> | xxi |
| <i>Table: Find your level: From Introduction to Beyond the basics</i> | xxii |
| | |
| Chapter 1 What is Epidemiology? | 01 |
| Key learning objectives | 01 |
| Key terms | 01 |
| Introduction | 02 |
| Defining epidemiology | 02 |
| Classification of epidemiology | 07 |
| The epidemiological process | 13 |
| A brief history of epidemiology | 18 |
| Examples of major achievements in public health through epidemiology | 25 |
| Eliminating smallpox | 25 |
| Rheumatic fever and rheumatic heart disease | 25 |
| Iodine deficiency | 26 |
| Smoking and lung cancer | 27 |
| Summary | 28 |
| Further reading | 28 |
| Websites | 29 |
| Bibliography | 29 |
| Answers to critical thinking exercises | 32 |
| | |
| Chapter 2 Disease Concepts in Epidemiology | 35 |
| Key learning objectives | 35 |
| Key terms | 35 |
| Introduction | 36 |



| | |
|---|-----|
| The natural history of disease | 37 |
| Aetiology of disease | 41 |
| An epidemiological approach to causality | 45 |
| Infectious disease concepts | 50 |
| The epidemiological triangle: Agent–host–environment | 52 |
| Agent | 52 |
| Host | 55 |
| Environment | 56 |
| Some basic concepts used in infectious disease epidemiology | 56 |
| Transmission routes | 56 |
| Infectious period and latent period | 57 |
| Investigation of an infectious disease outbreak | 59 |
| The epidemic curve | 61 |
| Modelling infectious diseases | 65 |
| A simple infectious disease model | 69 |
| Summary | 71 |
| Further reading | 72 |
| Websites | 72 |
| Bibliography | 73 |
| Answers to critical thinking exercises | 76 |
| | |
| Chapter 3 Identification of Disease: Diagnostic Tests and Screening | 81 |
| Key learning objectives | 81 |
| Key terms | 81 |
| Introduction | 82 |
| Quality measures of diagnostic tests | 83 |
| Assessing reliability of diagnostic tests | 84 |
| Assessing validity of diagnostic tests | 86 |
| How to calculate predictive values from sensitivity and specificity | 90 |
| Dependence of predictive values on prevalence | 93 |
| Receiver operating characteristics: Optimising cut-off points | 96 |
| Screening for disease | 100 |
| Important issues to be considered when screening | 101 |
| Screening programs | 102 |
| Evaluation of a screening program | 103 |
| Summary | 105 |
| Further reading | 106 |
| Websites | 106 |
| Bibliography | 106 |
| Answers to critical thinking exercises | 107 |

| | | |
|-----------|---|-----|
| Chapter 4 | Measures of Disease Frequency | 112 |
| | Key learning objectives | 112 |
| | Key terms | 112 |
| | Introduction | 113 |
| | Prevalence | 115 |
| | Characteristics of prevalence | 117 |
| | Incidence | 119 |
| | Cumulative incidence | 120 |
| | Incidence rate | 123 |
| | The concept of 'risk' | 128 |
| | Relationship between prevalence and incidence | 129 |
| | Standardisation of rates | 132 |
| | Crude rates and specific rates | 133 |
| | Direct standardisation of rates | 135 |
| | Indirect standardisation of rates | 136 |
| | Facts about standardisation and standardised rates | 138 |
| | Summary | 140 |
| | Further reading | 141 |
| | Websites | 141 |
| | Bibliography | 141 |
| | Answers to critical thinking exercises | 142 |
| | Appendix 4.1: Age structure of standard populations | 147 |
| Chapter 5 | From Research Topic to Research Hypothesis | 148 |
| | Key learning objectives | 148 |
| | Key terms | 148 |
| | How do you choose a research topic? | 149 |
| | The literature review | 154 |
| | Where to look for literature? | 154 |
| | The hierarchy of evidence | 155 |
| | Some practical tips on conducting a literature review | 156 |
| | Information seeking | 157 |
| | Critical appraisal | 158 |
| | Writing | 160 |
| | The operational research hypothesis | 161 |
| | Summary | 165 |
| | Further reading | 165 |
| | Websites | 166 |
| | Bibliography | 166 |
| | Answers to critical thinking exercises | 168 |



| | | |
|-----------|--|-----|
| Chapter 6 | Quantitative Descriptive Study Designs | 169 |
| | Key learning objectives | 169 |
| | Key terms | 169 |
| | What are quantitative descriptive studies? | 170 |
| | Descriptive epidemiology by person, place, and time | 172 |
| | Characteristics of 'person' | 173 |
| | Characteristics of 'place' | 175 |
| | Characteristics of 'time' | 177 |
| | Routine health statistics: An initial source of descriptive data | 178 |
| | Sources of routine health data | 179 |
| | Use of routine sources of health data | 182 |
| | Limitations of routinely collected health data | 185 |
| | Accuracy | 186 |
| | Completeness | 186 |
| | Timeliness | 187 |
| | Appropriateness | 187 |
| | Vital statistics | 188 |
| | Gaining access to routinely collected data | 191 |
| | Ecological (correlation) studies | 192 |
| | Limitations of ecological studies | 195 |
| | Can ecological studies be relevant for disease aetiology? | 199 |
| | Case report and case series | 200 |
| | Descriptive cross-sectional studies | 201 |
| | General comments on cross-sectional studies | 204 |
| | Summary | 205 |
| | Further reading | 206 |
| | Websites | 206 |
| | Bibliography | 206 |
| | Answers to critical thinking exercises | 211 |
| Chapter 7 | Experimental Designs | 215 |
| | Key learning objectives | 215 |
| | Key terms | 215 |
| | What defines an experimental study? | 216 |
| | Design options for experimental studies | 223 |
| | Historic controls (non-concurrent controls) | 224 |
| | Volunteer control group or researcher-decided control group | 226 |
| | Before-and-after controls | 227 |
| | Randomised controlled trial | 229 |
| | Placebo controls in randomised trials | 234 |

| | |
|---|-----|
| Blinding in randomised controlled trials | 236 |
| Cross-over trial | 238 |
| N-of-1 trial | 239 |
| Zelen's design | 240 |
| Multiple baseline design | 241 |
| Technical aspects of experimental studies | 243 |
| Compliance and non-compliance | 243 |
| Analysis of experimental studies | 244 |
| The study protocol | 247 |
| How to report an RCT | 248 |
| Summary | 249 |
| Further reading | 250 |
| Websites | 250 |
| Bibliography | 250 |
| Answers to critical thinking exercises | 254 |
| | |
| Chapter 8 Observational Designs | 258 |
| Key learning objectives | 258 |
| Key terms | 258 |
| What defines an observational study? | 259 |
| Directionality, timing, and the three basic observational designs | 261 |
| Directionality and timing | 261 |
| The three basic observational study designs | 265 |
| Cohort studies | 267 |
| Case-control studies | 273 |
| Comparative cross-sectional study | 281 |
| Measures of association | 287 |
| The relative risk: closed cohort studies | 287 |
| The exposure odds-ratio: case-control studies | 290 |
| The prevalence odds-ratio: cross-sectional studies | 294 |
| The relative risk for dynamic and fixed cohorts | 297 |
| Difference measures for cohort studies | 298 |
| Comparison between relative risk and difference measures | 301 |
| Difference measures in case-control studies | 302 |
| Other observational study designs | 304 |
| Summary | 306 |
| Further reading | 307 |
| Bibliography | 307 |
| Answers to critical thinking exercises | 310 |



| | | |
|------------|---|-----|
| Chapter 9 | Sources of Bias | 318 |
| | Key learning objectives | 318 |
| | Key terms | 318 |
| | The concept of bias | 319 |
| | Selection bias | 328 |
| | Volunteer bias | 329 |
| | Berkson's fallacy | 331 |
| | Detection bias | 335 |
| | Self-selection bias and selective survival | 336 |
| | Selection bias in different study designs | 337 |
| | Information bias | 338 |
| | Misclassification of disease status | 339 |
| | Detection bias | 342 |
| | The Will Rogers phenomenon | 342 |
| | Confounding bias | 344 |
| | Intermediate variable | 346 |
| | Potential confounding | 347 |
| | Confounding and effect modification | 351 |
| | Control of bias | 357 |
| | Control of bias during design phase of study | 357 |
| | Control of bias during statistical analysis | 360 |
| | Direction of bias | 360 |
| | Summary | 365 |
| | Further reading | 366 |
| | Bibliography | 366 |
| | Answers to critical thinking exercises | 368 |
| Chapter 10 | Sampling Strategy and Sample Size Calculation | 375 |
| | Key learning objectives | 375 |
| | Key terms | 375 |
| | Introduction | 376 |
| | Sampling strategy | 377 |
| | Simple random sampling | 378 |
| | Stratified random sampling | 379 |
| | Systematic sampling | 381 |
| | Cluster sampling | 382 |
| | Non-probability sampling | 385 |
| | Sample size calculation | 387 |
| | Sample size calculation for prevalence estimated by a cross-sectional study | 389 |

| | |
|---|-----|
| Summary | 392 |
| Further reading | 392 |
| Websites | 393 |
| Bibliography | 393 |
| Answers to critical thinking exercises | 394 |
| | |
| Chapter 11 Quantitative Methods of Data Collection | 399 |
| Key learning objectives | 399 |
| Key terms | 399 |
| Observations | 403 |
| Interviews | 406 |
| Types of interviews | 406 |
| Face-to-face interviews versus distance interviews | 407 |
| Recording interview information | 408 |
| Questionnaires | 408 |
| Questionnaire design: the process | 409 |
| The structure of a questionnaire | 409 |
| The format of responses | 414 |
| Visual analogue scale (VAS) and numerical rating scales (NRS) | 416 |
| Adjectival scale | 417 |
| Likert scale | 417 |
| Semantic differential scale | 417 |
| Vignettes | 419 |
| Validated instruments | 419 |
| What can go wrong with data collection? | 421 |
| The 'don't know' problem | 421 |
| The effect of wording on responses | 422 |
| Recall loss and telescoping error | 422 |
| Summary | 423 |
| Further reading | 424 |
| Websites | 424 |
| Bibliography | 424 |
| Answers to critical thinking exercises | 426 |
| | |
| Chapter 12 Statistics with Confidence | 432 |
| Key learning objectives | 432 |
| Key terms | 432 |
| Why statistics? | 433 |
| Types of quantitative data | 434 |

| | |
|--|-----|
| Descriptive statistics | 436 |
| Inferential statistics | 441 |
| Confidence interval | 442 |
| Confidence interval for the relative risk | 447 |
| Confidence interval for the odds-ratio | 448 |
| Statistical hypothesis testing | 451 |
| Errors in statistical testing and the problem of multiple testing | 453 |
| Selecting an appropriate bivariate statistical test | 455 |
| Survival analysis | 459 |
| The pitfalls of agreement and equivalence | 459 |
| Multivariable statistical analysis | 462 |
| Data preparations for multivariable analysis: correlation matrix | 462 |
| Data preparations for multivariable analysis: coding | 463 |
| Selecting an appropriate multivariable model | 464 |
| Presenting results from multivariable modelling | 466 |
| The use of computer programs for data analysis | 469 |
| Summary | 469 |
| Further reading | 470 |
| Websites | 471 |
| Bibliography | 471 |
| Answers to critical thinking exercises | 473 |
| Appendix 12.1: Mathematical signs, symbols, and abbreviations | 480 |
| | |
| Chapter 13 Ethical Considerations | 482 |
| Key learning objectives | 482 |
| Key terms | 482 |
| What is ethics? | 483 |
| An historical perspective of ethics in health research | 484 |
| Principles of biomedical ethics | 489 |
| Respect for autonomy | 490 |
| Beneficence and non-maleficance | 492 |
| Justice | 492 |
| Veracity and fidelity | 494 |
| Privacy and confidentiality | 494 |
| Further ethical issues in health research | 496 |
| Responsibility | 496 |
| Quality research | 497 |
| Debriefing | 497 |
| Ethical concerns in clinical research | 498 |
| Health research conducted in Aboriginal and Torres Strait Islander peoples | 499 |
| Coercion and deception | 501 |
| Scientific misconduct | 502 |

| | |
|--|-----|
| Ethical theory and public health | 503 |
| Utilitarianism | 503 |
| Liberalism | 504 |
| Communitarianism | 505 |
| Which point of view is correct? | 506 |
| Summary | 506 |
| Further reading | 507 |
| Websites | 507 |
| Bibliography | 507 |
| Answers to critical thinking exercises | 510 |
| | |
| Chapter 14 How to Read and Write Scientific Publications | 514 |
| Key learning objectives | 514 |
| Key terms | 514 |
| Introduction | 515 |
| The structure of a scientific publication | 515 |
| The introduction | 516 |
| The methodology | 517 |
| The results | 518 |
| The discussion | 519 |
| The abstract | 519 |
| The references | 520 |
| The title page | 520 |
| The acknowledgments | 520 |
| Guidelines for writing scientific articles | 521 |
| Technical aspects of publishing a scientific article | 524 |
| Formatting the manuscript | 524 |
| The peer-review process | 524 |
| The impact factor | 525 |
| How to read a scientific publication | 526 |
| Critical appraisal of published articles | 529 |
| Summary | 544 |
| Further reading | 545 |
| Websites | 545 |
| Bibliography | 545 |
| | |
| Glossary | 547 |
| Bibliography | 564 |
| Index | 592 |

List of Figures

| | | |
|-------------|--|----|
| Figure 1.1 | A disease may be endemic within a population, with a relatively constant number of cases, but an outbreak of the disease, causing a spike in the number of cases, is an epidemic | 4 |
| Figure 1.2 | The 10 leading causes of death in the United States, (a) 1900 and (b) 2006 | 5 |
| Figure 1.3 | Classification of epidemiological studies | 7 |
| Figure 1.4 | The scientific method | 15 |
| Figure 1.5 | An idealised concept of the epidemiological process | 16 |
| Figure 1.6 | Hippocrates—engraving by Peter Paul Rubens, 1638 | 19 |
| Figure 1.7 | John Graunt (1620–74) | 20 |
| Figure 1.8 | Dr John Snow (1813–58), British physician | 22 |
| Figure 2.1 | Timeline of the natural history of disease | 38 |
| Figure 2.2 | Example of visualisation of sufficient cause according to Kenneth Rothman. Factors A, B, C, D, and E create three different combinations for a sufficient cause each. Factor A is assumed to be a necessary cause. Factors B, C, D, and E are contributory factors | 43 |
| Figure 2.3 | Macro, intermediate, and micro factors | 44 |
| Figure 2.4 | Types of possible dose–response relationships: (a) ideal linear form, (b) threshold, and (c) saturation | 47 |
| Figure 2.5 | The epidemiological triangle | 52 |
| Figure 2.6 | Methods of developing immunity to biological agents | 55 |
| Figure 2.7 | The life cycle of schistosomiasis | 57 |
| Figure 2.8 | Infectious period and latent period of an infectious disease | 58 |
| Figure 2.9 | Original map made by John Snow in 1854, with locations of cholera cases highlighted in black | 60 |
| Figure 2.10 | Epidemic curve of a point source outbreak | 61 |
| Figure 2.11 | Epidemic curve of an outbreak of Legionnaires' disease in 1976 in Pennsylvania (Garrett 1994), an example of a typical point source outbreak | 62 |
| Figure 2.12 | An epidemic curve of a continuous source outbreak | 63 |
| Figure 2.13 | An epidemic curve of a person-to-person outbreak | 63 |
| Figure 2.14 | Epidemic curve of a nosocomial outbreak of hepatitis B virus among patients with diabetes in a Californian hospital in 1989–90 | 64 |

| | | |
|-------------|--|-----|
| Figure 2.15 | Schematic spread of an infection in a population: the darker blue index case (upper left corner) introduced the infectious disease into a susceptible population (pale blue dots) and the spread of the infection through the population can be traced by the medium blue dots | 65 |
| Figure 2.16 | Hypothetical epidemic in a population of 1000 people | 70 |
| Figure 3.1 | Reliability of the neuropsychological diagnostic test for Alzheimer's disease (hypothetical example) | 84 |
| Figure 3.2 | Hypothetical results of ELISA testing for HIV seropositivity | 97 |
| Figure 3.3 | Receiver operating characteristics curve (in pale blue) from example given in Box 3.7 | 99 |
| Figure 4.1 | Closed, fixed, and dynamic cohorts over a study period | 123 |
| Figure 4.2 | The relationship between incidence and prevalence | 130 |
| Figure 4.3 | Incidence rates of syphilis in the United States (1981–90): (a) crude rates; (b) gender-specific rates; (c) gender- and ethnicity-specific rates | 133 |
| Figure 4.4 | Age-specific incidence rates in two populations | 139 |
| Figure 6.1 | Age-specific incidence rates per 100 000 population for cancer of the trachea, bronchus, and lung stratified by gender, Australia 2005 | 174 |
| Figure 6.2 | A global view of active trachoma, 2005 | 175 |
| Figure 6.3 | Age-standardised incidence rate for cutaneous melanoma (per 100 000 inhabitants) by state and territory within Australia, 1999–2003 | 176 |
| Figure 6.4 | Age-standardised (world standard population) incidence rate for lung cancer per 100 000 inhabitants in Australia, 1982–2005 | 178 |
| Figure 6.5 | Census day in the Netherlands, 1925 | 180 |
| Figure 6.6 | Mortality rates per 100 000 population for diphtheria and measles in the United States between 1900 and 1970 | 184 |
| Figure 6.7 | Age-standardised mortality rates per 100 000 inhabitants for Indigenous and non-Indigenous residents of Western Australia between 1991 and 2005 | 185 |
| Figure 6.8 | Life expectancy at birth for different countries for 2005–2010 | 191 |
| Figure 6.9 | Correlation between per capita dietary fat intake and age-adjusted mortality from breast cancer in different countries | 193 |
| Figure 6.10 | Solar UV radiation index and age-standardised incidence of cutaneous melanoma per 100 000 inhabitants | 194 |
| Figure 6.11 | Hypothetical example of an ecological study investigating the relationship between per capita income and suicide rate in three populations* | 198 |
| Figure 7.1 | Before-and-after control design in its simplest form | 227 |
| Figure 7.2 | Table of simple random numbers | 230 |
| Figure 7.3 | Process of simple randomisation (R) | 232 |
| Figure 7.4 | Process of stratified randomisation (R) | 233 |
| Figure 7.5 | The process of filtering participants | 234 |



| | | |
|-------------|---|-----|
| Figure 7.6 | The process of a cross-over trial | 238 |
| Figure 7.7 | Timeline for intervention and multiple baseline study in three communities | 242 |
| Figure 8.1 | Prospective versus retrospective timing | 264 |
| Figure 8.2 | The three basic observational study designs | 265 |
| Figure 8.3 | Relationship between type of job in the rubber industry and occurrence of respiratory symptoms | 284 |
| Figure 9.1 | The hierarchy of populations | 319 |
| Figure 9.2 | Schematic representation of random and systematic errors | 322 |
| Figure 9.3 | An example selection process: from the target population to the sample | 329 |
| Figure 9.4 | Relationship between skin cancer (= outcome), hypertension (= study factor), and accidental bone fractures (= controls) in the target population | 333 |
| Figure 9.5 | Relationship between skin cancer (= outcome), hypertension (= study factor), and accidental bone fractures (= controls) restricted to 'diseased' people in the hospital | 334 |
| Figure 9.6 | Schematic representation of potential confounding | 345 |
| Figure 9.7 | Intermediate variable: three scenarios | 346 |
| Figure 9.8 | Checking for potential confounders | 348 |
| Figure 9.9 | Direction of bias using relative risk as the effect measure | 361 |
| Figure 11.1 | A framework for questionnaires | 402 |
| Figure 11.2 | The visual analogue scale | 416 |
| Figure 11.3 | The numerical rating scale | 417 |
| Figure 11.4 | Sir Francis Galton (1822–1911) | 420 |
| Figure 12.1 | The use of statistics in epidemiology: descriptive and inferential statistics | 434 |
| Figure 12.2 | Mean and median of two series of observations | 440 |
| Figure 12.3 | Population and sample means | 443 |
| Figure 12.4 | Classification of bivariate statistical tests | 455 |
| Figure 13.1 | Immanuel Kant (1724–1804), German philosopher | 504 |

List of Tables

| | | |
|------------|--|-----|
| Table 1.1 | Causes of death according to John Graunt | 20 |
| Table 1.2 | Water supply and death from Cholera (adapted from Snow 1855) | 23 |
| Table 1.3 | Age-standardised lung cancer death rates per 100 000 population stratified by smoking habits and exposure to asbestos | 27 |
| Table 2.1 | Incubation periods for selected infectious diseases | 39 |
| Table 2.2 | Biological agents of infectious diseases | 53 |
| Table 2.3 | Average times for incubation, latent, and infectious periods of common infectious diseases | 58 |
| Table 3.1 | Reliability of CT scans for spinal fractures—hypothetical results | 85 |
| Table 3.2 | Tabulating sensitivity and specificity | 87 |
| Table 3.3 | Hypothetical results of diagnostic tests for malaria: reference laboratory light microscope versus CyScope fluorescence microscope | 88 |
| Table 3.4 | Hypothetical results of the fasting blood glucose test for diabetes testing 1000 people | 92 |
| Table 3.5 | Diagnostic test with sensitivity 80%, specificity 90%, and prevalence 30%; assume a total number of 1000 | 94 |
| Table 3.6 | Diagnostic test with sensitivity 80%, specificity 90%, and prevalence 10%; assume a total number of 1000 | 94 |
| Table 3.7 | Diagnostic test with sensitivity 80%, specificity 90%, and prevalence 1%; assume a total number of 1000 | 95 |
| Table 3.8 | Hypothetical results of ELISA testing for HIV seropositivity | 97 |
| Table 3.9 | Sensitivity and specificity using a cut-off value of 1 | 98 |
| Table 3.10 | Sensitivity and specificity for different cut-off values of ELISA test | 98 |
| Table 3.11 | Sensitivity and specificity using a cut-off value of 5 | 98 |
| Table 4.1 | Hypothetical frequencies of chlamydia in two communities | 114 |
| Table 4.2 | Chlamydia prevalences in perceived high-risk groups | 118 |
| Table 4.3 | Results of an hypothetical follow-up study on HIV/AIDS in intravenous drug users | 125 |
| Table 4.4 | Crude and age-specific death rates for two communities: Mayhemberg and Lazyacres | 134 |
| Table 4.5 | Direct standardisation of rates | 135 |

| | | |
|------------|---|-----|
| Table 4.6 | Indirect standardisation of rates | 137 |
| Table 5.1 | Hierarchy of evidence | 155 |
| Table 5.2 | Example of two entries from a table created for a literature review on nutritional women's health in refugee situations | 160 |
| Table 6.1 | Definitions of some vital statistics | 188 |
| Table 6.2 | Life expectancy at birth in Australia 1996–2001 and 2005–2007 | 191 |
| Table 6.3 | Cancer mortality rates and fluoridated water supply in 46 US cities | 196 |
| Table 7.1 | Unbalanced random allocation list | 231 |
| Table 7.2 | Example of a balanced randomisation | 232 |
| Table 8.1 | Distribution of lung cancer cases and controls according to the average number of cigarettes smoked regularly per day before the onset of the current illness | 275 |
| Table 8.2 | A 2×2 cross-tabulation of the study factor and the outcome in a cohort study | 287 |
| Table 8.3 | Retrospective cohort study: exposure to kindergarten (study factor) and uptake of smoking in adulthood (outcome)* | 289 |
| Table 8.4 | Retrospective cohort study: exposure to not having attended kindergarten (study factor) and uptake of smoking in adulthood (outcome): switching the baseline category | 289 |
| Table 8.5 | A 2×2 cross-tabulation of the study factor and the outcome in a case–control study | 290 |
| Table 8.6 | Case–control study: skin cancer and previous dietary fat intake | 292 |
| Table 8.7 | Hypothetical case–control studies with ratio of cases to controls = 1:1 and 1:5 | 294 |
| Table 8.8 | A 2×2 cross-tabulation of study factor and outcome in a cross-sectional study | 295 |
| Table 8.9 | Cross-sectional study of parenting style and mental health outcome for children | 296 |
| Table 8.10 | A 2×2 cross-tabulation of results from a fixed or a dynamic cohort | 297 |
| Table 8.11 | Frequently used difference measures for closed, fixed, and dynamic cohorts | 298 |
| Table 8.12 | Mortality rates per 100 000 person-years for lung cancer and coronary artery disease stratified for smokers and non-smokers and the resulting relative risk and attributable rate | 301 |
| Table 9.1 | Results of the Salk poliomyelitis vaccine field trial conducted in 1954 | 330 |
| Table 9.2 | Association between hypertension (study factor) and skin cancer (outcome) in the target population of 50 000 people | 332 |
| Table 9.3 | Result of case–control study in the entire 'diseased' population from the target population | 333 |
| Table 9.4 | Result of case–control study in hospital population | 335 |
| Table 9.5 | True disease classification in the target population | 340 |
| Table 9.6 | Result of diagnostic test in exposed and unexposed groups from Table 9.5 | 340 |

| | | |
|------------|---|-----|
| Table 9.7 | Disease status in cohort study | 341 |
| Table 9.8 | Six-month survival probabilities of patients with lung cancer stratified by TNM stage | 343 |
| Table 9.9 | Percentage of patients with lung cancer stratified by TNM stages | 344 |
| Table 9.10 | Case–control study of alcohol consumption during pregnancy and birthweight of child | 350 |
| Table 9.11 | Case–control study of alcohol consumption during pregnancy and birthweight of child stratified by smoking status of mother | 350 |
| Table 9.12 | Confounding and effect modification | 352 |
| Table 9.13 | Overall results of a hypothetical cross-sectional study on ergonomically designed chairs and chronic low back pain in office workers | 353 |
| Table 9.14 | Results of a hypothetical cross-sectional study on ergonomically designed chairs and chronic low back pain in office workers stratified by age | 353 |
| Table 9.15 | Cross-sectional study: sexual orientation and sexual violence among Australian male prisoners | 362 |
| Table 10.1 | Relationship between prevalence p and $p(1 - p)$ | 390 |
| Table 10.2 | Relationship between precision, d , and sample size, assuming $p = 0.5$ and $z = 1.96$ | 391 |
| Table 11.1 | Percentage of respondents answering ‘yes’ by age | 422 |
| Table 12.1 | Overview descriptive statistics | 440 |
| Table 12.2 | 2×2 cross-tabulation of study factor and outcome in a closed and in a dynamic cohort study: Relative risk, rate ratio and their confidence intervals (Rothman, K. J., and Greenland, S., 1998) | 447 |
| Table 12.3 | Types of errors in statistical hypothesis testing | 453 |
| Table 12.4 | Overall alpha error and number of statistical tests conducted | 454 |
| Table 12.5 | Examples of study questions and suitable statistical tests | 456 |
| Table 12.6 | Dummy and stepwise coding | 463 |
| Table 12.7 | Overview of multivariable models most frequently used in the health sciences | 464 |
| Table 12.8 | Descriptive and multivariable log-binomial regression results of a cohort study of predictors of low birthweight for gestational age | 467 |
| Table 14.1 | The standard structure of a scientific publication in the health sciences | 516 |
| Table 14.2 | CONSORT checklist for randomised trials | 521 |

Preface

Epidemiology is a fairly young discipline which is highlighted by the fact that the first ever epidemiology textbook, written by Brian MacMahon and Thomas Pugh, was published just about 50 years ago in 1960. The core business of 'classical epidemiology' was, and still is, the identification of occurrences and distributions of diseases in populations and the detection and evaluation of causes of diseases. However, epidemiology has been a very dynamic discipline with new focus areas being added constantly—molecular, social, spatial, and cognitive epidemiology are just a few examples of more recent specialisations.

In the early 2000s some geneticists predicted the end of epidemiology as we know it, suggesting that genetics would soon be proven as the ultimate cause for most if not all diseases. This rather simplistic view may still be hampering serious genetics research but in the meantime genetic epidemiology, assessing the interplay between environmental and genetic factors and their role for health events, is thriving.

WHY ANOTHER EPIDEMIOLOGY BOOK?

Epidemiology belongs to the health sciences and medical scientists have traditionally dominated the discipline, bringing a disease centred approach to epidemiology. Our backgrounds, however, are in general research methodology, including biostatistics, and we hope to be able to contribute with our book a somewhat different point of view to further promote and progress modern epidemiology towards its full potential: a methodologically well grounded and versatile tool-kit to conduct evidence-based research in all health sciences.

This 'methods based' approach finds its expression already in the overall structure of this introductory level book that follows the logic of the epidemiological research 'cycle', which is probably better described as an upward spiral in the sense that each research study contributes novel evidence to the knowledge base of the respective discipline. We hope that this structure and our methods based approach as well as the plethora of examples from a broad spectrum of health related disciplines render this textbook to be a practical guide for conducting quantitative research in the broad field of the health sciences.

We surely advocate the integration of qualitative and quantitative research designs. However, in this book we focus only on quantitative research methods since qualitative

methods have become a very large discipline in its own right, including such varied concepts and techniques that only a separate book could do it justice.

This introductory textbook to epidemiology is largely based on our lecture notes for both postgraduate students of epidemiology for public health, and under-graduate students of public health and other health sciences (including medicine, nursing, physiotherapy, sports and exercise sciences, and occupational therapy). As with our lecture notes, we have opted for a two-level approach differentiating between 'introductory' knowledge that all students of epidemiology should be familiar with and 'beyond the basics' information for the interested or more advanced reader. The levels of the different sections are indicated in an overview table starting on page xxvi.

The first four chapters of the book introduce the reader to epidemiology, its historical roots and some basic concepts. The remaining chapters 5 to 14 follow the logic of the epidemiological research spiral.

Chapter 5 clarifies the concepts of a literature review, the first step of every epidemiological research study. This is followed by the formulation of an operational research hypothesis, a complete and quantitative precise statement of the question the research will aim to confirm or reject. Since the study design and the development of the research hypothesis are closely interlinked, different types of studies, their applications, advantages and disadvantages are subsequently discussed in some detail in Chapters 6, 7, and 8.

Chapter 9 discusses 'bias', systematic error in the conduct of epidemiological studies, and how bias can be avoided or at least controlled to achieve valid conclusions. Chapter 10 introduces basic sampling strategies and sample size considerations necessary to achieve an 'optimal' sample size for a planned study.

The logically following steps in the research spiral are then to develop measurement tools, such as questionnaires, collecting the data, entering the collected information into a data base (Chapter 11), and analysing the data statistically in order to reject or confirm the stated research hypothesis (Chapter 12). The chapter on statistical analyses is kept very concise and in non-technical language where possible; it is not intended to be a text for general statistical analysis but an introduction to the main concepts and techniques relevant in statistical hypothesis testing.

Chapter 13 deals with the important issues of ethics in human research. Finally, to achieve an 'upward' momentum in the research spiral, the results have to be published (Chapter 14) to be accessible to other scientists and to contribute novel evidence to the area-specific knowledge base.

Each chapter is accompanied by numerous critical thinking exercises (some conceptual, but more often practical also necessitating some calculations) with answers given in the back of the chapter.

The glossary at the end of the book provides an abridged 'dictionary' for epidemiology. However, the definitions provided there are working definitions as introduced and used in the context of this book; they may not be universally acceptable to all epidemiologists.



Throughout the book we apply examples of recent epidemiological research conducted in Australia and New Zealand. These examples are drawn from a wide range of disciplines demonstrating that epidemiological methodologies are essential research tools across the health sciences. We aimed at presenting as much original research as possible, though sometimes the actual figures had to be adjusted for ease of calculations or other pedagogical purposes. We hope that the examples showcase ongoing research in Australia and New Zealand and provide some inspiration to students of epidemiology.

Students sometimes may perceive epidemiology as a daunting subject. We thus aimed at introducing the concepts structured and in non-technical terms whenever possible and provided ample examples and exercises to stimulate reflection on theoretical concepts as well as on practical issues. However, every textbook has its limits—to really appreciate epidemiology you will have to jump into the deep end and conduct epidemiological research yourself. Our aim was to inspire readers to do exactly that!

Acknowledgments

Naturally this textbook would not have been written without the support of many people. First and foremost we would like to thank our friend and mentor at the Free University in Berlin, Germany, the late Professor Irene Guggenmoos-Holzmann (1948–1997). We would like to thank also in particular all our epidemiology students who helped us improve our lecture notes over the last decades. We gratefully acknowledge the contributions of Robert MacLennan, Professor Emeritus of Epidemiology, Queensland Institute of Medical Research, to an earlier version of our lecture notes. We also thank numerous friends and colleagues who supported and informed this book project over the last two years.

We would like to express our gratitude to Dr Madeleine Nowak and Dr Clive Wilkinson who helped us tirelessly in reviewing a first version of the manuscript and to all our colleagues who were happy enough to showcase their research in our book and to confide to us some of the intricacies of their epidemiological research.

Table: Find your level: From Introduction to Beyond the basics

| Chapter | Introductory level | Page no | Beyond the basics | Page no |
|---|---|---------|--|---------|
| Chapter 1 WHAT IS EPIDEMIOLOGY? | Introduction | 2 | A brief history of epidemiology | 18 |
| | Defining epidemiology | 2 | | |
| | Classification of epidemiology | 7 | Examples of major achievements in public health through epidemiology | 25 |
| | The epidemiological process | 13 | | |
| Chapter 2 DISEASE CONCEPTS IN EPIDEMIOLOGY | Introduction | 36 | Infectious disease concepts | 50 |
| | The natural history of disease | 37 | | |
| | Aetiology of disease | 41 | | |
| | An epidemiological approach to causality | 45 | | |
| Chapter 3 IDENTIFICATION OF DISEASE: DIAGNOSTIC TESTS AND SCREENING | Introduction | 82 | Dependence of predictive values on prevalence | 93 |
| | Quality measures of diagnostic tests | 83 | Receiver operating characteristics: | |
| | Assessing reliability of diagnostic tests | 84 | Optimising cut-off points | 96 |
| | Assessing validity of diagnostic tests | 86 | Screening for disease | 100 |
| | How to calculate predictive values from sensitivity and specificity | 90 | | |
| Chapter 4 MEASURES OF DISEASE FREQUENCY | Introduction | 113 | Standardisation of rates | 132 |
| | Prevalence | 115 | Direct standardisation of rates | 135 |
| | Incidence | 119 | Indirect standardisation of rates | 136 |
| | Relationship between prevalence and incidence | 129 | Appendix 4.1: Age structure of standard populations | 147 |
| Chapter 5 FROM RESEARCH TOPIC TO RESEARCH HYPOTHESIS | How do you choose a research topic? | 149 | | |
| | The literature review | 154 | | |
| | The operational research hypothesis | 161 | | |

| Chapter | Introductory level | Page no | Beyond the basics | Page no |
|--|---|---------|--|---------|
| Chapter 6 QUANTITATIVE DESCRIPTIVE STUDY DESIGNS | What are quantitative descriptive studies? | 170 | | |
| | Descriptive epidemiology by person, place, and time | 172 | | |
| | Routine health statistics: An initial source of descriptive data | 178 | | |
| | Ecological studies | 192 | | |
| | Case report and case series | 200 | | |
| | Descriptive cross-sectional studies | 201 | | |
| Chapter 7 EXPERIMENTAL DESIGNS | What defines an experimental study? | 216 | Zelen's design | 240 |
| | Design options for experimental studies | 223 | Multiple baseline design | 241 |
| | | | Technical aspects of experimental studies: | 243 |
| | | | Compliance and non-compliance | 243 |
| | | | Analysis of experimental studies | 244 |
| | | | The study protocol | 247 |
| | | | How to report an RCT | 248 |
| Chapter 8 OBSERVATIONAL DESIGNS | What defines an observational study? | 259 | The relative risk for dynamic and fixed cohorts | 297 |
| | Directionality, timing, and the three basic observational designs | 261 | Difference measures for cohort studies | 298 |
| | The three basic observational study designs | 265 | Comparison between relative risk and difference measures | 301 |
| | Measures of association | 287 | Difference measures in case-control studies | 302 |
| | | | Other observational study designs | 304 |
| Chapter 9 SOURCES OF BIAS | The concept of bias | 319 | Direction of bias | 360 |
| | Selection bias | 328 | | |
| | Information bias | 338 | | |
| | Confounding bias | 344 | | |
| | Control of bias | 357 | | |
| Chapter 10 SAMPLING STRATEGY AND SAMPLE SIZE CALCULATION | Sampling strategy | 377 | | |
| | Sample size calculation | 387 | | |

| Chapter | Introductory level | Page no | Beyond the basics | Page no |
|---|--|---------|---|---------|
| Chapter 11 QUANTITATIVE METHODS OF DATA COLLECTION | Observations | 403 | | |
| | Interviews | 406 | | |
| | Questionnaires | 408 | | |
| | What can go wrong with data collection? | 421 | | |
| Chapter 12 STATISTICS WITH CONFIDENCE | Why statistics? | 433 | Confidence interval for the relative risk | 447 |
| | Types of quantitative data | 434 | | |
| | Descriptive statistics | 436 | Confidence interval for the odds-ratio | 448 |
| | Inferential statistics | 441 | | |
| | Confidence interval | 442 | Multivariable statistical analysis | 462 |
| Chapter 13 ETHICAL CONSIDERATIONS | Statistical hypothesis testing | 451 | | |
| | What is ethics? | 483 | Ethical theory and public health | 503 |
| | An historical perspective of ethics in health research | 484 | | |
| | Principles of biomedical ethics | 489 | | |
| Chapter 14 HOW TO READ AND WRITE SCIENTIFIC PUBLICATIONS | Further ethical issues in health research | 496 | | |
| | The structure of a scientific publication | 515 | | |
| | Guidelines for writing scientific articles | 521 | | |
| | Technical aspects of publishing a scientific article | 524 | | |
| | How to read a scientific publication | 526 | | |