



Michael Borenstein
Larry V. Hedges
Julian P. T. Higgins
Hannah R. Rothstein

Introduction to Meta-Analysis

SECOND EDITION

WILEY

Introduction to Meta-Analysis

Introduction to Meta-Analysis

Second Edition

Michael Borenstein

Biostat, Inc, New Jersey, USA.

Larry V. Hedges

Northwestern University, Evanston, USA.

Julian P.T. Higgins

University of Bristol, Bristol, UK.

Hannah R. Rothstein

Baruch College, New York, USA.

WILEY

This edition first published 2021
© 2021 John Wiley & Sons Ltd

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, electronic, mechanical, photocopying, recording or otherwise, except as permitted by law. Advice on how to obtain permission to reuse material from this title is available at <http://www.wiley.com/go/permissions>.

The right of Michael Borenstein, Larry V. Hedges, Julian P.T. Higgins and Hannah R. Rothstein to be identified as the author of this work has been asserted in accordance with law.

Registered Offices

John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial Office

9600 Garsington Road, Oxford, OX4 2DQ, UK

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Limit of Liability/Disclaimer of Warranty

In view of ongoing research, equipment modifications, changes in governmental regulations, and the constant flow of information relating to the use of experimental reagents, equipment, and devices, the reader is urged to review and evaluate the information provided in the package insert or instructions for each chemical, piece of equipment, reagent, or device for, among other things, any changes in the instructions or indication of usage and for added warnings and precautions. While the publisher and author have used their best efforts in preparing this work, they make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives, written sales materials or promotional statements for this work. The fact that an organization, website, or product is referred to in this work as a citation and/or potential source of further information does not mean that the publisher and author endorse the information or services the organization, website, or product may provide or recommendations it may make. This work is sold with the understanding that the publisher is not engaged in rendering professional services. The advice and strategies contained herein may not be suitable for your situation. You should consult with a specialist where appropriate. Further, readers should be aware that websites listed in this work may have changed or disappeared between when this work was written and when it is read. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

Library of Congress Cataloging-in-Publication data applied for

HB ISBN: 9781119558354

Cover Design: Wiley
Cover Image: Courtesy of Michael Borenstein

Set in 10.5/13pt TimesLTStd by SPi Global, Chennai, India

10 9 8 7 6 5 4 3 2 1

Contents

List of Tables	xv
List of Figures	xix
Acknowledgements	xxv
Preface	xxvii
Preface to the Second Edition	xxxv
Website	xxxvii

PART 1: INTRODUCTION

1 HOW A META-ANALYSIS WORKS	3
Introduction	3
Individual studies	3
The summary effect	5
Heterogeneity of effect sizes	6
Summary points	7
2 WHY PERFORM A META-ANALYSIS	9
Introduction	9
The streptokinase meta-analysis	10
Statistical significance	11
Clinical importance of the effect	11
Consistency of effects	12
Summary points	13

PART 2: EFFECT SIZE AND PRECISION

3 OVERVIEW	17
Treatment effects and effect sizes	17
Parameters and estimates	18
Outline of effect size computations	19
4 EFFECT SIZES BASED ON MEANS	21
Introduction	21
Raw (unstandardized) mean difference D	21
Standardized mean difference, d and g	25
Response ratios	30
Summary points	31

5	EFFECT SIZES BASED ON BINARY DATA (2×2 TABLES)	33
	Introduction	33
	Risk ratio	33
	Odds ratio	35
	Risk difference	37
	Choosing an effect size index	38
	Summary points	38
6	EFFECT SIZES BASED ON CORRELATIONS	39
	Introduction	39
	Computing r	39
	Other approaches	40
	Summary points	41
7	CONVERTING AMONG EFFECT SIZES	43
	Introduction	43
	Converting from the log odds ratio to d	44
	Converting from d to the log odds ratio	45
	Converting from r to d	45
	Converting from d to r	46
	Summary points	47
8	FACTORS THAT AFFECT PRECISION	49
	Introduction	49
	Factors that affect precision	50
	Sample size	50
	Study design	51
	Summary points	53
9	CONCLUDING REMARKS	55
PART 3: FIXED-EFFECT VERSUS RANDOM-EFFECTS MODELS		
10	OVERVIEW	59
	Introduction	59
	Nomenclature	60
11	FIXED-EFFECT MODEL	61
	Introduction	61
	The true effect size	61
	Impact of sampling error	61
	Performing a fixed-effect meta-analysis	63
	Summary points	64

12	RANDOM-EFFECTS MODEL	65
	Introduction	65
	The true effect sizes	65
	Impact of sampling error	66
	Performing a random-effects meta-analysis	68
	Summary points	70
13	FIXED-EFFECT VERSUS RANDOM-EFFECTS MODELS	71
	Introduction	71
	Definition of a summary effect	71
	Estimating the summary effect	72
	Extreme effect size in a large study or a small study	73
	Confidence interval	73
	The null hypothesis	76
	Which model should we use?	76
	Model should not be based on the test for heterogeneity	78
	Concluding remarks	79
	Summary points	79
14	WORKED EXAMPLES (PART 1)	81
	Introduction	81
	Worked example for continuous data (Part 1)	81
	Worked example for binary data (Part 1)	85
	Worked example for correlational data (Part 1)	90
	Summary points	94
PART 4: HETEROGENEITY		
15	OVERVIEW	97
	Introduction	97
	Nomenclature	98
	Worked examples	98
16	IDENTIFYING AND QUANTIFYING HETEROGENEITY	99
	Introduction	99
	Isolating the variation in true effects	99
	Computing Q	101
	Estimating τ^2	106
	The I^2 statistic	109
	Comparing the measures of heterogeneity	111
	Confidence intervals for τ^2	114
	Confidence intervals (or uncertainty intervals) for I^2	115
	Summary points	116

17	PREDICTION INTERVALS	119
	Introduction	119
	Prediction intervals in primary studies	119
	Prediction intervals in meta-analysis	121
	Confidence intervals and prediction intervals	123
	Comparing the confidence interval with the prediction interval	123
	Summary points	125
18	WORKED EXAMPLES (PART 2)	127
	Introduction	127
	Worked example for continuous data (Part 2)	127
	Worked example for binary data (Part 2)	131
	Worked example for correlational data (Part 2)	134
	Summary points	138
19	AN INTUITIVE LOOK AT HETEROGENEITY	139
	Introduction	139
	Motivating example	140
	The Q -value and the p -value do not tell us how much the effect size varies	141
	The confidence interval does not tell us how much the effect size varies	142
	The I^2 statistic does not tell us how much the effect size varies	142
	What I^2 tells us	142
	The I^2 index vs. the prediction interval	145
	The prediction interval	145
	Prediction interval is clear, concise, and relevant	147
	Computing the prediction interval	147
	How to use I^2	149
	How to explain heterogeneity	149
	How much does the effect size vary across studies?	150
	Caveats	150
	Conclusion	150
	Further reading	151
	Summary points	151
	The meaning of I^2 in Figure 19.2	151
20	CLASSIFYING HETEROGENEITY AS LOW, MODERATE, OR HIGH	155
	Introduction	155
	Interest should generally focus on an index of absolute heterogeneity	155
	The classifications lead themselves to mistakes of interpretation	158
	Classifications focus attention in the wrong direction	158
	Summary points	158

PART 5: EXPLAINING HETEROGENEITY

21	SUBGROUP ANALYSES	161
	Introduction	161
	Fixed-effect model within subgroups	163
	Computational models	172
	Random effects with separate estimates of τ^2	174
	Random effects with pooled estimate of τ^2	181
	The proportion of variance explained	189
	Mixed-effects model	192
	Obtaining an overall effect in the presence of subgroups	193
	Summary points	195
22	META-REGRESSION	197
	Introduction	197
	Fixed-effect model	198
	Fixed or random effects for unexplained heterogeneity	203
	Random-effects model	206
	Summary points	212
23	NOTES ON SUBGROUP ANALYSES AND META-REGRESSION	213
	Introduction	213
	Computational model	213
	Multiple comparisons	216
	Software	216
	Analyses of subgroups and regression analyses are observational	217
	Statistical power for subgroup analyses and meta-regression	218
	Summary points	219
PART 6: PUTTING IT ALL IN CONTEXT		
24	LOOKING AT THE WHOLE PICTURE	223
	Introduction	223
	Methylphenidate for adults with ADHD	226
	Impact of GLP-1 mimetics on blood pressure	228
	Augmenting clozapine with a second antipsychotic	228
	Conclusions	231
	Caveats	231
	Summary points	232
25	LIMITATIONS OF THE RANDOM-EFFECTS MODEL	233
	Introduction	233
	Assumptions of the random-effects model	234

A textbook case	234
When studies are pulled from the literature	235
A useful fiction	237
Transparency	238
A narrowly defined universe	238
Two important caveats	239
In context	239
Extreme cases	240
Summary points	241
26 KNAPP–HARTUNG ADJUSTMENT	243
Introduction	243
Adjustment is rarely employed in simple analyses	243
Adjusting the standard error	244
The Knapp–Hartung adjustment for other effect size indices	246
<i>t</i> distribution vs. <i>Z</i> distribution	247
Limitations of the Knapp–Hartung adjustment	248
Summary points	249
PART 7: COMPLEX DATA STRUCTURES	
27 OVERVIEW	253
28 INDEPENDENT SUBGROUPS WITHIN A STUDY	255
Introduction	255
Combining across subgroups	255
Comparing subgroups	260
Summary points	260
29 MULTIPLE OUTCOMES OR TIME-POINTS WITHIN A STUDY	263
Introduction	263
Combining across outcomes or time-points	264
Comparing outcomes or time-points within a study	270
Summary points	275
30 MULTIPLE COMPARISONS WITHIN A STUDY	277
Introduction	277
Combining across multiple comparisons within a study	277
Differences between treatments	278
Summary points	279
31 NOTES ON COMPLEX DATA STRUCTURES	281
Introduction	281
Summary effect	281
Differences in effect	282

PART 8: OTHER ISSUES

32	OVERVIEW	287
33	VOTE COUNTING – A NEW NAME FOR AN OLD PROBLEM	289
	Introduction	289
	Why vote counting is wrong	290
	Vote counting is a pervasive problem	291
	Summary points	293
34	POWER ANALYSIS FOR META-ANALYSIS	295
	Introduction	295
	A conceptual approach	295
	In context	299
	When to use power analysis	300
	Planning for precision rather than for power	301
	Power analysis in primary studies	301
	Power analysis for meta-analysis	304
	Power analysis for a test of homogeneity	309
	Summary points	312
35	PUBLICATION BIAS	313
	Introduction	313
	The problem of missing studies	314
	Methods for addressing bias	316
	Illustrative example	317
	The model	317
	Getting a sense of the data	318
	Is there evidence of any bias?	320
	How much of an impact might the bias have?	320
	Summary of the findings for the illustrative example	324
	Conflating bias with the small-study effect	325
	Using logic to disentangle bias from small-study effects	326
	These methods do not give us the ‘correct’ effect size	327
	Some important caveats	327
	Procedures do not apply to studies of prevalence	328
	The model for publication bias is simplistic	328
	Concluding remarks	329
	Putting it all together	330
	Summary points	330

PART 9: ISSUES RELATED TO EFFECT SIZE

36	OVERVIEW	335
-----------	-----------------	------------

37	EFFECT SIZES RATHER THAN p-VALUES	337
	Introduction	337
	Relationship between p -values and effect sizes	337
	The distinction is important	339
	The p -value is often misinterpreted	340
	Narrative reviews vs. meta-analyses	341
	Summary points	342
38	SIMPSON'S PARADOX	343
	Introduction	343
	Circumcision and risk of HIV infection	343
	An example of the paradox	345
	Summary points	348
39	GENERALITY OF THE BASIC INVERSE-VARIANCE METHOD	349
	Introduction	349
	Other effect sizes	350
	Other methods for estimating effect sizes	353
	Individual participant data meta-analyses	354
	Bayesian approaches	355
	Summary points	357
PART 10: FURTHER METHODS		
40	OVERVIEW	361
41	META-ANALYSIS METHODS BASED ON DIRECTION AND p-VALUES	363
	Introduction	363
	Vote counting	363
	The sign test	363
	Combining p -values	364
	Summary points	368
42	FURTHER METHODS FOR DICHOTOMOUS DATA	369
	Introduction	369
	Mantel–Haenszel method	369
	One-step (Peto) formula for odds ratio	373
	Summary points	376
43	PSYCHOMETRIC META-ANALYSIS	377
	Introduction	377
	The attenuating effects of artifacts	378
	Meta-analysis methods	380
	Example of psychometric meta-analysis	381
	Comparison of artifact correction with meta-regression	384

Sources of information about artifact values	384
How heterogeneity is assessed	385
Reporting in psychometric meta-analysis	386
Concluding remarks	386
Summary points	387

PART 11: META-ANALYSIS IN CONTEXT

44 OVERVIEW	391
45 WHEN DOES IT MAKE SENSE TO PERFORM A META-ANALYSIS?	393
Introduction	393
Are the studies similar enough to combine?	394
Can I combine studies with different designs?	395
How many studies are enough to carry out a meta-analysis?	399
Summary points	400
46 REPORTING THE RESULTS OF A META-ANALYSIS	401
Introduction	401
The computational model	402
Forest plots	402
Sensitivity analysis	404
Summary points	405
47 CUMULATIVE META-ANALYSIS	407
Introduction	407
Why perform a cumulative meta-analysis?	409
Summary points	412
48 CRITICISMS OF META-ANALYSIS	413
Introduction	413
One number cannot summarize a research field	414
The file drawer problem invalidates meta-analysis	414
Mixing apples and oranges	415
Garbage in, garbage out	416
Important studies are ignored	417
Meta-analysis can disagree with randomized trials	417
Meta-analyses are performed poorly	420
Is a narrative review better?	420
Concluding remarks	422
Summary points	422
49 COMPREHENSIVE META-ANALYSIS SOFTWARE	425
Introduction	425
Features in CMA	426

Teaching elements	427
Documentation	427
Availability	427
Acknowledgments	427
Motivating example	428
Data entry	428
Basic analysis	429
What is the <i>average</i> effect size?	430
How much does the effect size vary?	430
Plot showing distribution of effects	431
High-resolution plot	432
Subgroup analysis	433
Meta-regression	435
Publication bias	438
Explaining results	439
50 HOW TO EXPLAIN THE RESULTS OF AN ANALYSIS	443
Introduction	443
The overview	444
The mean effect size	444
Variation in effect size	444
Notations	444
Impact of resistance exercise on pain	445
Correlation between letter knowledge and word recognition	450
Statins for prevention of cardiovascular events	455
Bupropion for smoking cessation	460
Mortality following mitral-valve procedures in elderly patients	465
PART 12: RESOURCES	
51 SOFTWARE FOR META-ANALYSIS	471
Comprehensive meta-analysis	471
Metafor	471
Stata	472
Revman	472
52 WEB SITES, SOCIETIES, JOURNALS, AND BOOKS	473
Web sites	473
Professional societies	476
Journals	476
Special issues dedicated to meta-analysis	477
Books on systematic review methods and meta-analysis	477
REFERENCES	479
INDEX	491

List of Tables

Table 3.1	Roadmap of formulas in subsequent chapters	19
Table 5.1	Nomenclature for 2×2 table of outcome by treatment	34
Table 5.2	Fictional data for a 2×2 table	34
Table 8.1	Impact of sample size on variance	51
Table 8.2	Impact of study design on variance	52
Table 14.1	Dataset 1 – Part A (basic data)	82
Table 14.2	Dataset 1 – Part B (fixed-effect computations)	83
Table 14.3	Dataset 1 – Part C (random-effects computations)	85
Table 14.4	Dataset 2 – Part A (basic data)	86
Table 14.5	Dataset 2 – Part B (fixed-effect computations)	87
Table 14.6	Dataset 2 – Part C (random-effects computations)	89
Table 14.7	Dataset 3 – Part A (basic data)	90
Table 14.8	Dataset 3 – Part B (fixed-effect computations)	91
Table 14.9	Dataset 3 – Part C (random-effects computations)	93
Table 16.1	Factors affecting measures of dispersion	111
Table 18.1	Dataset 1 – Part D (intermediate computations)	128
Table 18.2	Dataset 1 – Part E (variance computations)	128
Table 18.3	Dataset 2 – Part D (intermediate computations)	131
Table 18.4	Dataset 2 – Part E (variance computations)	131
Table 18.5	Dataset 3 – Part D (intermediate computations)	135
Table 18.6	Dataset 3 – Part E (variance computations)	135
Table 19.1	Relationship between observed effects and true effects in Figure 19.2, Panel A	152
Table 21.1	Fixed effect model – computations	164
Table 21.2	Fixed-effect model – summary statistics	167
Table 21.3	Fixed-effect model – ANOVA table	169
Table 21.4	Fixed-effect model – subgroups as studies	170
Table 21.5	Random-effects model (separate estimates of τ^2) – computations	176
Table 21.6	Random-effects model (separate estimates of τ^2) – summary statistics	177
Table 21.7	Random-effects model (separate estimates of t^2) – ANOVA table	180
Table 21.8	Random-effects model (separate estimates of τ^2) – subgroups as studies	181

Table 21.9	Statistics for computing a pooled estimate of τ^2	183
Table 21.10	Random-effects model (pooled estimate of τ^2) – computations	183
Table 21.11	Random-effects model (pooled estimate of τ^2) – summary statistics	185
Table 21.12	Random-effects model (pooled estimate of τ^2) – ANOVA table	187
Table 21.13	Random-effects model (pooled estimate of τ^2) – subgroups as studies	188
Table 22.1	The BCG dataset	200
Table 22.2	Fixed-effect model – Regression results for BCG	200
Table 22.3	Fixed-effect model – ANOVA table for BCG regression	200
Table 22.4	Random-effects model – regression results for BCG	207
Table 22.5	Random-effects model – test of the model	207
Table 22.6	Random-effects model – comparison of model (latitude) versus the null model	211
Table 26.1	Knapp–Hartung computations for ADHD analysis	244
Table 26.2	Original vs. Knapp–Hartung	246
Table 26.3	Impact of using t distribution on the confidence interval width	248
Table 28.1	Independent subgroups – five fictional studies	256
Table 28.2	Independent subgroups – summary effect	257
Table 28.3	Independent subgroups – synthetic effect for study 1	257
Table 28.4	Independent subgroups – summary effect across studies	258
Table 29.1	Multiple outcomes – five fictional studies	264
Table 29.2	Creating a synthetic variable as the mean of two outcomes	265
Table 29.3	Multiple outcomes – summary effect	267
Table 29.4	Multiple outcomes – impact of correlation on variance of summary effect	269
Table 29.5	Creating a synthetic variable as the difference between two outcomes	271
Table 29.6	Multiple outcomes – difference between outcomes	272
Table 29.7	Multiple outcomes – Impact of correlation on the variance of difference	274
Table 38.1	HIV as function of circumcision (by subgroup)	344
Table 38.2	HIV as function of circumcision – by study	345
Table 38.3	HIV as a function of circumcision – full population	346
Table 38.4	HIV as a function of circumcision – by risk group	346
Table 38.5	HIV as a function of circumcision/risk group – full population	347
Table 39.1	Simple example of a genetic association study	352
Table 41.1	Streptokinase data – calculations for meta-analyses of p -values	367
Table 42.1	Nomenclature for 2×2 table of events by treatment	370
Table 42.2	Mantel–Haenszel – odds ratio	371

Table 42.3	Mantel–Haenszel – variance of summary effect	372
Table 42.4	One-step – odds ratio and variance	375
Table 43.1	Fictional data for psychometric meta-analysis	382
Table 43.2	Observed (attenuated) correlations	382
Table 43.3	Unattenuated correlations	383

List of Figures

Figure 1.1	High-dose versus standard-dose of statins (adapted from Cannon <i>et al.</i> , 2006)	4
Figure 2.1	Impact of streptokinase on mortality (adapted from Lau <i>et al.</i> , 1992)	10
Figure 4.1	Response ratios are analyzed in log units	30
Figure 5.1	Risk ratios are analyzed in log units	34
Figure 5.2	Odds ratios are analyzed in log units	36
Figure 6.1	Correlations are analyzed in Fisher's z units	40
Figure 7.1	Converting among effect sizes	44
Figure 8.1	Impact of sample size on variance	51
Figure 8.2	Impact of study design on variance	52
Figure 10.1	Symbols for true and observed effects	60
Figure 11.1	Fixed-effect model – true effects	62
Figure 11.2	Fixed-effect model – true effects and sampling error	62
Figure 11.3	Fixed-effect model – distribution of sampling error	63
Figure 12.1	Random-effects model – distribution of the true effects	66
Figure 12.2	Random-effects model – true effects	66
Figure 12.3	Random-effects model – true and observed effect in one study	67
Figure 12.4	Random-effects model – between-study and within-study variance	68
Figure 13.1	Fixed-effect model – forest plot showing relative weights	72
Figure 13.2	Random-effects model – forest plot showing relative weights	72
Figure 13.3	Very large studies under fixed-effect model	74
Figure 13.4	Very large studies under random-effects model	74
Figure 14.1	Forest plot of Dataset 1 – fixed-effect weights	84
Figure 14.2	Forest plot of Dataset 1 – random-effects weights	85
Figure 14.3	Forest plot of Dataset 2 – fixed-effect weights	88
Figure 14.4	Forest plot of Dataset 2 – random-effects weights	90
Figure 14.5	Forest plot of Dataset 3 – fixed-effect weights	92
Figure 14.6	Forest plot of Dataset 3 – random-effects weights	94
Figure 16.1	Dispersion across studies relative to error within studies	100
Figure 16.2	Q in relation to df as measure of dispersion	102
Figure 16.3	Flowchart showing how T^2 and I^2 are derived from Q and df	104
Figure 16.4	Impact of Q and number of studies on the p -value	105
Figure 16.5	Impact of excess dispersion and absolute dispersion on T^2	107
Figure 16.6	Impact of excess and absolute dispersion on T	108

Figure 16.7	Impact of excess dispersion on I^2	110
Figure 16.8	Factors affecting T^2 but not I^2	112
Figure 16.9	Factors affecting I^2 but not T^2	112
Figure 17.1	Prediction interval based on population parameters μ and τ^2	121
Figure 17.2	Prediction interval based on sample estimates M^* and T^2	122
Figure 17.3	Simultaneous display of confidence interval and prediction interval	123
Figure 17.4	Impact of number of studies on confidence interval and prediction interval	124
Figure 18.1	Forest plot of Dataset 1 – random-effects weights with prediction interval	130
Figure 18.2	Forest plot of Dataset 2 – random-effects weights with prediction interval	134
Figure 18.3	Forest plot of Dataset 3 – random-effects weights with prediction interval	137
Figure 19.1	Alcohol use and mortality. Risk ratio < 1 favors drinkers. Three possible distributions of true effects	141
Figure 19.2	Alcohol use and mortality. Risk ratio < 1 favors drinkers. Three possible distributions of true effects (inner) and observed effects (outer)	144
Figure 19.3	Alcohol use and mortality (Forest plot). Risk ratio < 1 favors drinkers.	146
Figure 19.4	Alcohol use and mortality (true effects). Risk ratio < 1 favors drinkers.	148
Figure 20.1	True effects for two meta-analyses	156
Figure 20.2	True effects (inner) and observed effects (outer) for two meta-analyses	157
Figure 21.1	Fixed-effect model – studies and subgroup effects	164
Figure 21.2	Fixed-effect – subgroup effects	167
Figure 21.3	Fixed-effect model – treating subgroups as studies	170
Figure 21.4	Flowchart for selecting a computational model	174
Figure 21.5	Random-effects model (separate estimates of τ^2) – studies and subgroup effects	175
Figure 21.6	Random-effects model (separate estimates of τ^2) – subgroup effects	178
Figure 21.7	Random-effects model (separate estimates of τ^2) – treating subgroups as studies	180
Figure 21.8	Random-effects model (pooled estimate of τ^2) – studies and subgroup effects	182
Figure 21.9	Random-effects model (pooled estimate of τ^2) – subgroup effects	185
Figure 21.10	Random-effects model (pooled estimate of τ^2) – treating subgroups as studies	188
Figure 21.11	A primary study showing subjects within groups	190

Figure 21.12	Random-effects model – variance within and between subgroups	191
Figure 21.13	Proportion of variance explained by subgroup membership	191
Figure 22.1	Fixed-effect model – forest plot for the BCG data	199
Figure 22.2	Fixed-effect model – regression of log risk ratio on latitude	202
Figure 22.3	Fixed-effect model – population effects as function of covariate	204
Figure 22.4	Random-effects model – population effects as a function of covariate	204
Figure 22.5	Random-effects model – forest plot for the BCG data	206
Figure 22.6	Random-effects model – regression of log risk ratio on latitude	208
Figure 22.7	Between-studies variance (T^2) with no covariate	210
Figure 22.8	Between-studies variance (T^2) with covariate	210
Figure 22.9	Proportion of variance explained by latitude	212
Figure 24.1	Three fictional examples where the mean effect is 0.00	224
Figure 24.2	Three fictional examples where the mean effect is 0.40	225
Figure 24.3	Three fictional examples where the mean effect is 0.80	226
Figure 24.4	Methylphenidate for adults with ADHD (Forest plot). Effect size > 0 favors treatment	227
Figure 24.5	Methylphenidate for adults with ADHD (True effects). Effect size > 0 favors treatment	228
Figure 24.6	GLP-1 mimetics and diastolic BP (Forest plot). Mean difference < 0 favors treatment	229
Figure 24.7	GLP-1 mimetics and diastolic BP (True effects). Mean difference < 0 favors treatment	229
Figure 24.8	Augmenting clozapine (Forest plot). Std mean difference < 0 favors augmentation	230
Figure 24.9	Augmenting clozapine (True effects). Std mean difference < 0 favors augmentation	230
Figure 25.1	Random effects. Confidence interval 60 points wide	234
Figure 25.2	Methylphenidate for adults with ADHD. Effect size > 0 favors treatment	236
Figure 28.1	Creating a synthetic variable from independent subgroups	257
Figure 33.1	The p -value for each study is > 0.20 but the p -value for the summary effect is < 0.02	290
Figure 34.1	Power for a primary study as a function of n and δ	304
Figure 34.2	Power for a meta-analysis as a function of number studies and δ	307
Figure 34.3	Power for a meta-analysis as a function of number studies and heterogeneity	309
Figure 35.1	Passive smoking and lung cancer – forest plot	318
Figure 35.2	Passive smoking and lung cancer – funnel plot	319
Figure 35.3	Observed studies only	321

Figure 35.4	Observed studies and studies imputed by Trim and Fill	322
Figure 35.5	Passive smoking and lung cancer – cumulative forest plot	323
Figure 37.1	Estimating the effect size versus testing the null hypothesis	338
Figure 37.2	The p -value is a poor surrogate for effect size	339
Figure 37.3	Studies where p -values differ but effect sizes is the same	340
Figure 37.4	Studies where p -values are the same but effect sizes differ	341
Figure 37.5	Studies where the more significant p -value corresponds to weaker effect size	341
Figure 38.1	Circumcision and HIV. Odds Ratio > 1 indicates circumcision is associated with lower risk of HIV.	344
Figure 38.2	HIV as function of circumcision – in three sets of studies	347
Figure 41.1	Effect size in four fictional studies	366
Figure 46.1	Forest plot using lines to represent the effect size	403
Figure 46.2	Forest plot using boxes to represent the effect size and relative weight	404
Figure 47.1	Impact of streptokinase on mortality – forest plot	408
Figure 47.2	Impact of streptokinase on mortality – cumulative forest plot	409
Figure 48.1	Forest plot of five fictional studies and a new trail (consistent effects)	418
Figure 48.2	Forest plot of five fictional studies and a new trial (heterogeneous effects)	419
Figure 49.1	Data-entry screen in CMA.	428
Figure 49.2	Basic analysis screen in CMA	429
Figure 49.3	Average effect size (top), Variation in effect size (bottom)	430
Figure 49.4	Plotting distribution of true effects. ADHD	432
Figure 49.5	High-resolution plot in CMA	433
Figure 49.6	Impact of treatment as a function of subgroup: Forest plot	434
Figure 49.7	Impact of treatment as a function of subgroup: Statistics	434
Figure 49.8	Results for regression, random effects	436
Figure 49.9	Regression of effect size on Dose, with SUD held constant	437
Figure 49.10	Funnel plot of observed effects	438
Figure 49.11	Funnel plot of observed and imputed effects	439
Figure 49.12	Regression of effect size (d) on Dose and SUD. Plot created in Excel (TM)	441
Figure 50.1	Impact of resistance exercise on pain. Data-entry screen	447
Figure 50.2	Impact of resistance exercise on pain. $g > 0$ indicates exercise reduced pain	448
Figure 50.3	Impact of resistance exercise on pain. Heterogeneity statistics	449
Figure 50.4	Impact of resistance exercise on pain. Distribution of true effects	449
Figure 50.5	Predicting reading scores. Data-entry screen	452
Figure 50.6	Predicting reading scores	453
Figure 50.7	Predicting reading scores. Heterogeneity statistics	454
Figure 50.8	Predicting reading scores. Distribution of true correlations	454

Figure 50.9	Statins for prevention of cardiovascular events. Data-entry screen	457
Figure 50.10	Statins for prevention of cardiovascular events. Odds ratio < 1 shows reduction in events	458
Figure 50.11	Statins for prevention of cardiovascular events. Heterogeneity statistics	459
Figure 50.12	Statins for prevention of cardiovascular events. Distribution of true effects	459
Figure 50.13	Bupropion for smoking cessation. Data-entry screen	462
Figure 50.14	Bupropion for smoking cessation. Risk ratio > 1 shows reduction in smoking	463
Figure 50.15	Bupropion for smoking cessation. Heterogeneity statistics	464
Figure 50.16	Bupropion for smoking cessation. Distribution of true effects	464
Figure 50.17	Mortality following mitral-valve surgery in elderly patients. Data-entry screen	466
Figure 50.18	Mortality following mitral-valve surgery in elderly patients	467
Figure 50.19	Mortality following mitral-valve surgery in elderly patients. Heterogeneity statistics	468
Figure 50.20	Mortality following mitral-valve surgery in elderly patients. Distribution of true risks	468

Acknowledgements

This book was funded by the following grants from the National Institutes of Health: *Combining data types in meta-analysis* (AG021360), *Publication bias in meta-analysis* (AG20052), *Software for meta-regression* (AG024771), from the National Institute on Aging, under the direction of Dr. Sidney Stahl; and *Forest plots for meta-analysis* (DA019280), from the National Institute on Drug Abuse, under the direction of Dr. Thomas Hilton.

These grants allowed us to convene a series of workshops on meta-analysis, and parts of this volume reflect ideas developed as part of these workshops. We would like to acknowledge and thank Doug Altman, Betsy Becker, Jesse Berlin, Michael Brannick, Harris Cooper, Kay Dickersin, Sue Duval, Roger Harbord, Despina Contopoulos-Ioannidis, John Ioannidis, Spyros Konstantopoulos, Mark Lipsey, Mike McDaniel, Ingram Olkin, Fred Oswald, Terri Pigott, Simcha Pollack, David Rindskopf, Stephen Senn, Will Shadish, Jonathan Sterne, Alex Sutton, Thomas Trikalinos, Jeff Valentine, Jack Vevea, Vish Viswesvaran, and David Wilson.

Steven Tarlow helped to edit this book and to ensure the accuracy of all formulas and examples. We would like to acknowledge and thank our editors at Wiley, including Kathryn Sharples, Ashley Alliano, Alison Oliver, Sarah Keegan, Kimberly Monroe-Hill and Viktoria Hartl-Vida. We would especially like to thank Adalfin Jayasingh who served as the production editor for this volume. His attention to detail and his patience in working through revisions are very much appreciated.

Preface

In his best-selling book *Baby and Child Care*, Dr. Benjamin Spock wrote ‘I think it is preferable to accustom a baby to sleeping on his stomach from the beginning if he is willing’. This statement was included in most editions of the book, and in most of the 50 million copies sold from the 1950s into the 1990s. The advice was not unusual, in that many pediatricians made similar recommendations at the time.

During this same period, from the 1950s into the 1990s, more than 100,000 babies died of sudden infant death syndrome (SIDS), also called *crib death* in the United States and *cot death* in the United Kingdom, where a seemingly healthy baby goes to sleep and never wakes up.

In the early 1990s, researchers became aware that the risk of SIDS decreased by at least 50% when babies were put to sleep on their backs rather than face down. Governments in various countries launched educational initiatives such as the *Back to sleep* campaigns in the United Kingdom and the United States, which led to an immediate and dramatic drop in the number of SIDS deaths.

While the loss of more than 100,000 children would be unspeakably sad in any event, the real tragedy lies in the fact that many of these deaths could have been prevented. Gilbert *et al.* (2005) write

Advice to put infants to sleep on the front for nearly half a century was contrary to evidence available from 1970 that this was likely to be harmful. Systematic review of preventable risk factors for SIDS from 1970 would have led to earlier recognition of the risks of sleeping on the front and might have prevented over 10,000 infant deaths in the UK and at least 50,000 in Europe, the USA and Australasia.

AN ETHICAL IMPERATIVE

This example is one of several cited by Sir Iain Chalmers in a talk entitled *The scandalous failure of scientists to cumulate scientifically* (Chalmers, 2006). The theme of this talk was that we live in a world where the utility of almost any intervention will be tested repeatedly, and that rather than looking at any study in isolation, we need to look at the body of evidence. While not all systematic reviews carry the urgency of SIDS, the logic of looking at the body of evidence, rather than trying to understand studies in isolation, is always compelling.

Meta-analysis refers to the statistical synthesis of results from a series of studies. While the statistical procedures used in a meta-analysis can be applied to any set of data, the synthesis will be meaningful only if the studies have been collected systematically. This could be in the context of a systematic review, the process of

systematically locating, appraising, and then synthesizing data from a large number of sources. Or, it could be in the context of synthesizing data from a select group of studies, such as those conducted by a pharmaceutical company to assess the efficacy of a new drug.

If a treatment effect (or effect size) is consistent across the series of studies, these procedures enable us to report that the effect is robust across the kinds of populations sampled, and also to estimate the magnitude of the effect more precisely than we could with any of the studies alone. If the treatment effect varies across the series of studies, these procedures enable us to report on the range of effects, and may enable us to identify factors associated with the magnitude of the effect size.

FROM NARRATIVE REVIEWS TO SYSTEMATIC REVIEWS

Prior to the 1990s, the task of combining data from multiple studies had been primarily the purview of the narrative review. An expert in a given field would read the studies that addressed a question, summarize the findings, and then arrive at a conclusion – for example, that the treatment in question was, or was not, effective. However, this approach suffers from some important limitations.

One limitation is the subjectivity inherent in this approach, coupled with the lack of transparency. For example, different reviewers might use different criteria for deciding which studies to include in the review. Once a set of studies has been selected, one reviewer might give more credence to larger studies, while another gives more credence to ‘quality’ studies and yet another assigns a comparable weight to all studies. One reviewer may require a substantial body of evidence before concluding that a treatment is effective, while another uses a lower threshold. In fact, there are examples in the literature where two narrative reviews come to opposite conclusions, with one reporting that a treatment is effective while the other reports that it is not. As a rule, the narrative reviewer will not articulate (and may not even be fully aware of) the decision-making process used to synthesize the data and arrive at a conclusion.

A second limitation of narrative reviews is that they become *less useful as more information becomes available*. The thought process required for a synthesis requires the reviewer to capture the finding reported in each study, to assign an appropriate weight to that finding, and then to synthesize these findings across all studies in the synthesis. While a reviewer may be able to synthesize data from a few studies in their head, the process becomes difficult and eventually untenable as the number of studies increases. This is true even when the treatment effect (or effect size) is consistent from study to study. Often, however, the treatment effect will vary as a function of study level covariates, such as the patient population, the dose of medication, the outcome variable, and other factors. In these cases, a proper synthesis requires that the researcher be able to understand how the treatment effect varies as a function of these variables, and the narrative review is poorly equipped to address these kinds of issues.