

Contents

Preface	xiii
Part I The Methods	1
1 What can the reader expect from this book?	3
1.1 A calibration scale for evidence	4
1.1.1 <i>T</i> -values and <i>p</i> -values	4
1.1.2 How generally applicable is the calibration scale?	6
1.1.3 Combining evidence	7
1.2 The efficacy of glass ionomer versus resin sealants for prevention of caries.	8
1.2.1 The data	8
1.2.2 Analysis for individual studies	9
1.2.3 Combining the evidence: fixed effects model	10
1.2.4 Combining the evidence: random effects model	10
1.3 Measures of effect size for two populations	11
1.4 Summary	13
2 Independent measurements with known precision	15
2.1 Evidence for one-sided alternatives	15
2.2 Evidence for two-sided alternatives	18
2.3 Examples	19
2.3.1 Filling containers	19
2.3.2 Stability of blood samples	20
2.3.3 Blood alcohol testing	20
3 Independent measurements with unknown precision	23
3.1 Effects and standardized effects	23
3.2 Paired comparisons	26
3.3 Examples	27
3.3.1 Daily energy intake compared to a fixed level	27
3.3.2 Darwin's data on <i>Zea mays</i>	28

4	Comparing treatment to control	31
4.1	Equal unknown precision	31
4.2	Differing unknown precision.	33
4.3	Examples	35
4.3.1	Drop in systolic blood pressure	35
4.3.2	Effect of psychotherapy on hospital length of stay	37
5	Comparing K treatments	39
5.1	Methodology	39
5.2	Examples	42
5.2.1	Characteristics of antibiotics	42
5.2.2	Red cell folate levels.	43
6	Evaluating risks	47
6.1	Methodology	47
6.2	Examples	49
6.2.1	Ultrasound and left-handedness	49
6.2.2	Treatment of recurrent urinary tract infections	49
7	Comparing risks	51
7.1	Methodology	51
7.2	Examples	54
7.2.1	Treatment of recurrent urinary tract infections	54
7.2.2	Diuretics in pregnancy and risk of pre-eclamsia	54
8	Evaluating Poisson rates	57
8.1	Methodology	57
8.2	Example.	60
8.2.1	Deaths by horse-kicks	60
9	Comparing Poisson rates	63
9.1	Methodology	64
9.1.1	Unconditional evidence	64
9.1.2	Conditional evidence.	65
9.2	Example.	67
9.2.1	Vaccination for the prevention of tuberculosis	67
10	Goodness-of-fit testing	71
10.1	Methodology	71
10.2	Example.	74
10.2.1	Bellbirds arriving to feed nestlings	74
11	Evidence for heterogeneity of effects and transformed effects	77
11.1	Methodology	77
11.1.1	Fixed effects	77
11.1.2	Random effects.	80

11.2	Examples	81
11.2.1	Deaths by horse-kicks	81
11.2.2	Drop in systolic blood pressure	82
11.2.3	Effect of psychotherapy on hospital length of stay	83
11.2.4	Diuretics in pregnancy and risk of pre-eclamsia	84
12	Combining evidence: fixed standardized effects model	85
12.1	Methodology	86
12.2	Examples	87
12.2.1	Deaths by horse-kicks	87
12.2.2	Drop in systolic blood pressure	88
13	Combining evidence: random standardized effects model	91
13.1	Methodology	91
13.2	Example.	94
13.2.1	Diuretics in pregnancy and risk of pre-eclamsia	94
14	Meta-regression	95
14.1	Methodology	95
14.2	Commonly encountered situations.	98
14.2.1	Standardized difference of means	98
14.2.2	Difference in risk (two binomial proportions).	99
14.2.3	Log relative risk (two Poisson rates).	99
14.3	Examples	100
14.3.1	Effect of open education on student creativity	100
14.3.2	Vaccination for the prevention of tuberculosis	101
15	Accounting for publication bias	105
15.1	The downside of publishing	105
15.2	Examples	107
15.2.1	Environmental tobacco smoke	107
15.2.2	Depression prevention programs	109
Part II	The Theory	111
16	Calibrating evidence in a test	113
16.1	Evidence for one-sided alternatives	114
16.1.1	Desirable properties of one-sided evidence	115
16.1.2	Connection of evidence to p -values.	115
16.1.3	Why the p -value is hard to understand	116
16.2	Random p -value behavior.	118
16.2.1	Properties of the random p -value distribution.	118
16.2.2	Important consequences for interpreting p -values	119
16.3	Publication bias.	119
16.4	Comparison with a Bayesian calibration	121
16.5	Summary	123

17	The basics of variance stabilizing transformations	125
17.1	Standardizing the sample mean	125
17.2	Variance stabilizing transformations	126
17.2.1	Background material	126
17.2.2	The Key Inferential Function	127
17.3	Poisson model example	128
17.3.1	Example of counts data	129
17.3.2	A simple <i>vst</i> for the Poisson model.	129
17.3.3	A better <i>vst</i> for the Poisson model	132
17.3.4	Achieving a desired expected evidence	132
17.3.5	Confidence intervals	132
17.3.6	Simulation study of coverage probabilities.	134
17.4	Two-sided evidence from one-sided evidence	134
17.4.1	A <i>vst</i> based on the chi-squared statistic	135
17.4.2	A <i>vst</i> based on doubling the <i>p</i> -value	137
17.5	Summary	138
18	One-sample binomial tests	139
18.1	Variance stabilizing the risk estimator	139
18.2	Confidence intervals for <i>p</i>	140
18.3	Relative risk and odds ratio	142
18.3.1	One-sample relative risk	143
18.3.2	One-sample odds ratio.	144
18.4	Confidence intervals for small risks <i>p</i>	145
18.4.1	Comparing intervals based on the log and arcsine transformations.	145
18.4.2	Confidence intervals for small <i>p</i> based on the Poisson approximation to the binomial	146
18.5	Summary	147
19	Two-sample binomial tests	149
19.1	Evidence for a positive effect	149
19.1.1	Variance stabilizing the risk difference	149
19.1.2	Simulation studies.	151
19.1.3	Choosing sample sizes to achieve desired expected evidence	151
19.1.4	Implications for the relative risk and odds ratio	153
19.2	Confidence intervals for effect sizes	153
19.3	Estimating the risk difference	155
19.4	Relative risk and odds ratio	155
19.4.1	Two-sample relative risk	155
19.4.2	Two-sample odds ratio	157
19.4.3	New confidence intervals for the RR and OR.	157
19.5	Recurrent urinary tract infections	157
19.6	Summary	158

20	Defining evidence in t-statistics	159
20.1	Example	159
20.2	Evidence in the Student t -statistic	159
20.3	The Key Inferential Function for Student's model	163
20.4	Corrected evidence	165
20.4.1	Matching p -values	166
20.4.2	Reducing bias	167
20.5	A confidence interval for the standardized effect	169
20.5.1	Simulation study of coverage probabilities	170
20.6	Comparing evidence in t - and z -tests	170
20.6.1	On substituting s for σ in large samples	170
20.7	Summary	172
21	Two-sample comparisons	175
21.1	Drop in systolic blood pressure	175
21.2	Defining the standardized effect	176
21.3	Evidence in the Welch statistic	177
21.3.1	The Welch statistic	177
21.3.2	Variance stabilizing the Welch t -statistic	178
21.3.3	Choosing the sample size to obtain evidence	179
21.4	Confidence intervals for δ	179
21.4.1	Converting the evidence to confidence intervals	179
21.4.2	Simulation studies	180
21.4.3	Drop in systolic blood pressure (continued)	181
21.5	Summary	181
22	Evidence in the chi-squared statistic	183
22.1	The noncentral chi-squared distribution	183
22.2	A vst for the noncentral chi-squared statistic	184
22.2.1	Deriving the vst	184
22.2.2	The Key Inferential Function	185
22.3	Simulation studies	186
22.3.1	Bias in the evidence function	186
22.3.2	Upper confidence bounds; confidence intervals	187
22.4	Choosing the sample size	190
22.4.1	Sample sizes for obtaining an expected evidence	190
22.4.2	Sample size required to obtain a desired power	192
22.5	Evidence for $\lambda > \lambda_0$	192
22.6	Summary	193
23	Evidence in F-tests	195
23.1	Variance stabilizing transformations for the noncentral F	195
23.2	The evidence distribution	199
23.3	The Key Inferential Function	202
23.3.1	Refinements	205
23.4	The random effects model	205

23.4.1	Expected evidence in the balanced case	207
23.4.2	Comparing evidence in REM and FEM.	208
23.5	Summary	208
24	Evidence in Cochran's Q for heterogeneity of effects	209
24.1	Cochran's Q : the fixed effects model	210
24.1.1	Background material	210
24.1.2	Evidence for heterogeneity of fixed effects	212
24.1.3	Evidence for heterogeneity of transformed effects	213
24.2	Simulation studies	213
24.3	Cochran's Q : the random effects model	216
24.4	Summary	220
25	Combining evidence from K studies	221
25.1	Background and preliminary steps.	221
25.2	Fixed standardized effects	222
25.2.1	Fixed, and equal, standardized effects	222
25.2.2	Fixed, but unequal, standardized effects	223
25.2.3	Nuisance parameters	223
25.3	Random transformed effects	224
25.3.1	The random transformed effects model	224
25.3.2	Evidence for a positive effect.	225
25.3.3	Confidence intervals for κ and δ : K small	226
25.3.4	Confidence intervals for κ and δ : K large	226
25.3.5	Simulation studies.	227
25.4	Example: drop in systolic blood pressure	229
25.4.1	Inference for the fixed effects model.	231
25.4.2	Inference for the random effects model.	232
25.5	Summary	232
26	Correcting for publication bias	233
26.1	Publication bias.	233
26.2	The truncated normal distribution	235
26.3	Bias correction based on censoring	237
26.4	Summary	240
27	Large-sample properties of variance stabilizing transformations	241
27.1	Existence of the variance stabilizing transformation	241
27.2	Tests and effect sizes	242
27.3	Power and efficiency	245
27.4	Summary	249
	Acknowledgements	251
	Bibliography	253
	Index	257