

---

---

## Contents

1.	THE STUDY OF OCCURRENCE PROBLEMS IN MEDICINE: INTRODUCTION	1
1.1.	Theoretical Epidemiology as the Discipline of Occurrence Research in Medicine	1
1.2.	Generality of the Epidemiologic Discipline	4
1.3.	Occurrence Relations as the Actual Objects: Occurrence Parameters, Their Determinants, and Modifiers of Relations	5
1.4.	Descriptive vs. Causal Relations	11
1.5.	Conditional Relations: Confounders	12
1.6.	Referent of the Relation: Particularistic vs. Abstract Problems	16
1.7.	Levels of Study	17
1.7.1.	Qualitative Exploration and Hypothesis Testing	17
1.7.2.	Quantification, Relative and Absolute	18

## PART 1 STUDY DESIGN

OVERVIEW OF ISSUES OF DESIGN	21
2. DESIGN OF THE OCCURRENCE RELATION	25
2.1. Outcome Phenomenon: Definitions, Conceptual and Empirical	25
2.2. Outcome Parameter: Prevalence or Status Distribution vs. Incidence or Change Distribution	28
2.3. Determinant Scale: Validity of Contrasts	29
2.3.1. Comparability of Effects	30
2.3.2. Comparability of Populations	31
2.3.3. Comparability of Information	34
2.3.4. Epilogue	35
2.4. Relation: Cross-Sectional vs. Longitudinal	36
2.5. Modifiers: Treatment	37
2.6. Confounders: Identification in Concept	39
2.7. Confounders: Operational Imperfections	42
2.8. Domain of the Relation	44
3. DESIGN OF THE STUDY BASE	46
3.1. Introduction	46
3.2. Membership of the Base Population	48
3.2.1. Admissibility	48
3.2.1.1. Cohort Admissibility	48
3.2.1.2. Dynamic Population Admissibility	52
3.2.1.3. Admissibility in Population Cross Sections	53
3.2.1.4. Base Membership Secondary to Case Selection	54
3.2.1.5. Place and Time in Admissibility	55
3.2.2. Distribution Matrix	56
3.2.2.1. Distribution(s) of Determinant(s)	56
3.2.2.2. Distributions of Modifiers	58

3.2.2.3.	<i>Distributions of Confounders</i>	59
3.2.2.4.	<i>Joint Distribution</i>	59
3.2.3.	Size	60
3.3.	Coverage of Risk Period	62
3.3.1.	Follow-up of Entry Cohort	62
3.3.2.	Follow-up of Oblique Cohort	63
3.3.3.	Dynamic Population Coverage	64
3.3.4.	Retrospective Coverage	64
3.4.	Biased Base	66
3.4.1.	Censorship	66
3.4.2.	Result-Dependent Investigation	67
3.4.3.	Outcome-Dependent Admissibility	68
3.4.4.	Outcome-Dependent Termination of Follow-up	68
4.	DESIGN OF SAMPLING OF THE BASE	69
4.1.	The Case-Referent Approach: Fundamentals	69
4.2.	Stratified Sampling of Cases	73
4.3.	Stratified Sampling of the Base	74
4.4.	Valid Sampling	78
4.4.1.	Valid Sampling in the Context of a Primary Base	78
4.4.2.	Valid Sampling in the Context of a Base Secondary to Cases	79
4.4.2.1.	<i>The Ideal Reference Entity and Its             Source</i>	79
4.4.2.2.	<i>Inclusions vs. Exclusions</i>	82
4.4.2.3.	<i>The Validity of Siblings and             Neighbors</i>	82
4.4.2.4.	<i>Reference Series from Outside the             Base</i>	82
5.	DESIGN OF INFORMATION ON THE SAMPLES	84
5.1.	Presampling Provisions for Information	84
5.2.	Time Referent of Information	85

5.3.	Conditional Sufficiency of Information	87
5.4.	Content of Information	88
5.5.	Structure of Information	89
5.6.	Comparability of Information	91
5.7.	Accuracy of Information	94
6.	DESIGN OF ANALYSIS	96
7.	INTERDEPENDENCIES IN DESIGN	99
8.	STUDY PROTOCOL	101

## PART 2 DATA ANALYSIS

9.	ELEMENTS OF DATA ANALYSIS AND INFERENCE	107
9.1.	Introduction	107
9.2.	Hypothesis vs. Its Denial	109
9.3.	The $P$ -Value	110
9.3.1.	Sampling Behavior of the $P$ -Value	110
9.3.2.	Basic Interpretation of $P$ -Values	111
9.3.3.	$P$ -Values for Data-Suggested Hypotheses	113
9.3.4.	$P$ -Values for Multiple Hypotheses	114
9.3.5.	$P$ -Values in Repeat Testing	116
9.4.	Testing vs. Estimation	117
9.5.	Interval Estimates	117
9.5.1.	One-Sided Interval Estimate	117
9.5.2.	Two-Sided Interval Estimate	120
9.6.	Point Estimate	122
9.7.	The Likelihood Ratio	122
9.7.1.	Essence of the Likelihood Ratio	122
9.7.2.	Interpretation of the Likelihood Ratio	124
9.8.	Review	127

10.	ANALYSIS OF A SINGLE RATE	129
10.1.	Single Proportion	130
10.1.1.	Chi Square Function	130
10.1.2.	Asymptotic <i>LR</i> Function	134
10.1.3.	Exact <i>P</i> -Value Function	135
10.1.4.	Exact <i>LR</i> Function	137
10.2.	Single Density	137
11.	COMPARATIVE ANALYSIS OF TWO RATES I: NULL CHI SQUARE	140
11.1.	Two Proportions	141
11.1.1.	Unstratified Series	141
11.1.2.	Stratified Series in General	143
11.1.2.1.	<i>The Cochran–Mantel– Haenszel Statistic</i>	143
11.1.2.2.	<i>LR Statistics</i>	146
11.1.3.	Matched Series	149
11.2.	Two Densities	152
12.	COMPARATIVE ANALYSIS OF TWO RATES II: CHI SQUARE FUNCTION FOR RATE DIFFERENCE	155
12.1.	Contrast-Based $X^2$ Function of <i>RD</i>	156
12.1.1.	Unstratified Series	156
12.1.2.	Stratified Series in General	162
12.1.3.	Matched Series	164
12.2.	<i>LR</i> -Based $X^2$ Function of <i>RD</i>	165
12.3.	Simple Analysis for <i>RD</i>	166
12.4.	Analysis for Modification of <i>RD</i>	168
13.	COMPARATIVE ANALYSIS OF TWO RATES III: CHI SQUARE FUNCTION FOR RATE RATIO	172
13.1.	Contrast-Based $X^2$ Function of <i>RR</i>	172
13.1.1.	Unstratified Series	172

	13.1.2. Stratified Series in General	176
	13.1.3. Matched Series	177
13.2.	<i>LR</i> -based $X^2$ Function of <i>RR</i>	179
13.3.	Simple Analysis for <i>RR</i>	180
13.4.	Analysis for Modification of <i>RR</i>	181
14.	COMPARATIVE ANALYSIS OF TWO RATES IV: CHI SQUARE FUNCTION FOR ODDS RATIO	184
14.1.	Score-Based $X^2$ Function of <i>OR</i>	184
	14.1.1. Unstratified Series	184
	14.1.2. Stratified Series in General	189
	14.1.3. Matched Series	189
14.2.	<i>LR</i> -Based $X^2$ Function of <i>OR</i>	191
14.3.	Simple Analysis for <i>OR</i>	192
14.4.	Analysis for Modification of <i>OR</i>	193
15.	COMPARATIVE ANALYSIS OF SEVERAL RATES: NULL CHI SQUARE	194
15.1.	Basic Null $X^2$	194
15.2.	Angular Null $X^2$	198
16.	ANALYSIS OF MEANS OF QUANTITATIVE OBSERVATIONS	201
16.1.	Single Mean	201
16.2.	Two Means	204
	16.2.1. Null $X^2$ or $t^2$	204
	16.2.2. $X^2$ or $t^2$ Function of Mean Difference	206
16.3.	Several Means	207
	16.3.1. Null $X^2$ or $t^2$	207
	16.3.2. $X^2$ or $t^2$ Function of Regression Slope	210
17.	ANALYSIS OF CASE-REFERENT DATA	211
17.1.	Null $X^2$	211
17.2.	$X^2$ Function of <i>RR</i>	213
17.3.	Simple Analysis for <i>RR</i>	214

18.	REGRESSION ANALYSIS	216
18.1.	Linear Regression: The Concept	217
18.2.	Independent Variates: Specification and Result Interpretation	218
18.2.1.	Determinant Representation	218
	18.2.1.1. <i>Representation of Nominal Determinant</i>	218
	18.2.1.2. <i>Representation of Ordinal or Quantitative Determinant</i>	221
	18.2.1.3. <i>Determinant Characterized in Two or More Dimensions</i>	223
	18.2.1.4. <i>Determinant History</i>	226
18.2.2.	Confounder Representation	227
18.2.3.	Nonconfounder–Nonmodifier Representation	229
18.2.4.	Modifier Representation	229
18.2.5.	Missing Data Representation	231
18.3.	Dependent Parameter: Specification and Result Interpretation	233
18.3.1.	Parameter for Binary Outcome: General Linear vs. Logistic Model	233
18.3.2.	Parameter for Quantitative Outcome: Mean vs. Median	237
18.4.	Regression Analysis of Case-Referent Data	239
18.5.	Implementation of Regression Analysis	241
18.5.1.	Preliminaries to Regression Analysis	241
18.5.2.	Fitting of Causal Model	242
18.5.3.	Fitting of Descriptive Model	243

## APPENDICES

APPENDIX 1.	RATE MEASURES OF OCCURRENCE	245
A.1.1.	Prevalence Rate	245
A.1.2.	Incidence Rates	246

A.1.3.	Risk	249
A.1.4.	Proportionate Rate	250
APPENDIX 2.	MEASURES OF OCCURRENCE RELATION	251
A.2.1.	Fundamental Measures of Susceptibility and Resistance	251
A.2.2.	Susceptibility/Resistance and Measures of Relation	252
A.2.2.1.	<i>Rate Difference</i>	252
A.2.2.2.	<i>Derivatives of Rate Ratio</i>	252
A.2.3.	Derivatives of Observed and Expected Numbers of Cases	254
A.2.4.	Odds Ratio	256
A.2.5.	Correlation and Regression Coefficients	257
A.2.6.	Proportionate Mortality Ratio	260
A.2.7.	Deaths Prevented (Lives Saved)	261
APPENDIX 3.	CAUSAL AND PREVENTIVE INTERDEPENDENCE (SYNERGISM AND ANTAGONISM)	262
A.3.1.	Definitions	262
A.3.2.	Rate Criteria	264
APPENDIX 4.	CRUDE, ADJUSTED, AND STANDARDIZED RATES	266
A.4.1.	Crude Rate as a Weighted Average of Specific Rates	266
A.4.2.	Adjustment and Standardization: The Ideas	268
A.4.3.	The Notion of "Indirect" Standardization	269
A.4.4.	The Standardized Mortality (Morbidity) Ratio	270
A.4.5.	Precision-Maximizing Weights	272
APPENDIX 5.	CENSUS VS. CASE-REFERENT APPROACH: RELATIVE INFORMATIVENESS	273

APPENDIX 6. CASE-REFERENT APPROACH: SAMPLE OF BASE (REFERENT) VS. SAMPLE OF NONCASE SUBDOMAIN OF BASE	277
APPENDIX 7. RELATIVE SIZES OF COMPARED SERIES	280
APPENDIX 8. REVIEW OF STUDY DESIGN	283
A.8.1. Questions of Validity in the Design Stage	283
A.8.2. Questions of Efficiency in the Design Stage	284
A.8.3. Questions in the Review of "Positive" Studies	284
A.8.4. Questions in the Review of "Negative" Studies	285
APPENDIX 9. DISTRIBUTION OF <i>P</i> -VALUE AND INTERPRETATION OF TEST RESULTS	286
APPENDIX 10. SOME PROBABILITY MODELS	294
A.10.1. Binomial Distributions	294
A.10.2. Poisson Distributions	295
A.10.3. Hypergeometric Distributions	295
A.10.4. Noncentral Hypergeometric Distributions	297
A.10.5. Gaussian Distributions	298
APPENDIX 11. <i>P</i> -VALUE FROM STANDARD GAUSSIAN STATISTIC	299
APPENDIX 12. <i>P</i> -VALUE FROM CHI SQUARE STATISTIC	301
APPENDIX 13. ANALYSIS OF MATCHED SERIES: VARIANCE ESTIMATES FOR $X^2$ FUNCTIONS OF RATE DIFFERENCE AND RATE RATIO	303
A.13.1. Rate Difference	304
A.13.2. Rate Ratio	308

APPENDIX 14. ANALYSIS OF MATCHED SERIES: $\chi^2$ FUNCTION OF ODDS RATIO	312
GLOSSARY	317
REFERENCES	345
INDEX	351