Theoretical Epidemiology

PRINCIPLES OF OCCURRENCE RESEARCH IN MEDICINE

Olli S. Miettinen, M.D., Ph.D., M.D. (hon.)
Professor of Epidemiology and Biostatistics
School of Public Health, Harvard University
Boston, Massachusetts
Professor of Epidemiology and Biostatistics
Professor of Medicine
Faculty of Medicine, McGill University
Montreal, Quebec

A WILEY MEDICAL PUBLICATION
JOHN WILEY & SONS
New York • Chichester • Brisbane • Toronto • Singapore
writing in the summer months of 1981 and 1982, complete with secretarial help, word processing, statistical computing, and other amenities. For computing collaboration there I am very indebted to Mrs. Taula Nurminen. Two other colleagues there, Dr. Markku Nurminen and Mr. Timo Partanen, read the first draft and provided valuable suggestions. Subsequently, a similar service was provided by a large number of students and other colleagues. In the final revision stage, after a couple of years of testing the text in various courses, Drs. Chung Hsieh and Markku Nurminen were devoted and most valuable collaborators.

The developmental work with previously unpublished results was supported in part by grant number 5 R 01 CA 06373 from the National Cancer Institute of the United States of America.

For long-term support and encouragement within Harvard I am indebted to Dr. Alexander Nadas, the towering clinician and recognized patron of epidemiologic scholarship in the environment of academic medicine.

O.S.M.

Contents

1. THE STUDY OF OCCURRENCE PROBLEMS IN MEDICINE: INTRODUCTION

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

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. The Ideal Reference Entity and Its Source 79
      4.4.2.2. Inclusions vs. Exclusions 82
      4.4.2.3. The Validity of Siblings and Neighbors 82
      4.4.2.4. Reference Series from Outside the Base 82

5. DESIGN OF INFORMATION ON THE SAMPLES 84
5.1. Presampling Provisions for Information 84
5.2. Time Referent of Information 85
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.3</td>
<td>Conditional Sufficiency of Information</td>
<td>87</td>
</tr>
<tr>
<td>5.4</td>
<td>Content of Information</td>
<td>88</td>
</tr>
<tr>
<td>5.5</td>
<td>Structure of Information</td>
<td>89</td>
</tr>
<tr>
<td>5.6</td>
<td>Comparability of Information</td>
<td>91</td>
</tr>
<tr>
<td>5.7</td>
<td>Accuracy of Information</td>
<td>94</td>
</tr>
<tr>
<td>6.</td>
<td>DESIGN OF ANALYSIS</td>
<td>96</td>
</tr>
<tr>
<td>7.</td>
<td>INTERDEPENDENCIES IN DESIGN</td>
<td>99</td>
</tr>
<tr>
<td>8.</td>
<td>STUDY PROTOCOL</td>
<td>101</td>
</tr>
</tbody>
</table>

**PART 2  DATA ANALYSIS**

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

**Contents**

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.</td>
<td>ANALYSIS OF A SINGLE RATE</td>
<td>129</td>
</tr>
<tr>
<td>10.1</td>
<td>Single Proportion</td>
<td>130</td>
</tr>
<tr>
<td>10.1.1</td>
<td>Chi Square Function</td>
<td>130</td>
</tr>
<tr>
<td>10.1.2</td>
<td>Asymptotic LR Function</td>
<td>134</td>
</tr>
<tr>
<td>10.1.3</td>
<td>Exact P-Value Function</td>
<td>135</td>
</tr>
<tr>
<td>10.1.4</td>
<td>Exact LR Function</td>
<td>137</td>
</tr>
<tr>
<td>10.2</td>
<td>Single Density</td>
<td>137</td>
</tr>
<tr>
<td>11.</td>
<td>COMPARATIVE ANALYSIS OF TWO RATES I: NULL CHI SQUARE</td>
<td>140</td>
</tr>
<tr>
<td>11.1</td>
<td>Two Proportions</td>
<td>141</td>
</tr>
<tr>
<td>11.1.1</td>
<td>Unstratified Series</td>
<td>141</td>
</tr>
<tr>
<td>11.1.2</td>
<td>Stratified Series in General</td>
<td>143</td>
</tr>
<tr>
<td>11.1.2.1</td>
<td>The Cochran–Mantel–Haenszel Statistic</td>
<td>143</td>
</tr>
<tr>
<td>11.1.2.2</td>
<td>LR Statistics</td>
<td>146</td>
</tr>
<tr>
<td>11.1.3</td>
<td>Matched Series</td>
<td>149</td>
</tr>
<tr>
<td>11.2</td>
<td>Two Densities</td>
<td>152</td>
</tr>
<tr>
<td>12.</td>
<td>COMPARATIVE ANALYSIS OF TWO RATES II: CHI SQUARE FUNCTION FOR RATE DIFFERENCE</td>
<td>155</td>
</tr>
<tr>
<td>12.1</td>
<td>Contrast-Based $\chi^2$ Function of RD</td>
<td>156</td>
</tr>
<tr>
<td>12.1.1</td>
<td>Unstratified Series</td>
<td>156</td>
</tr>
<tr>
<td>12.1.2</td>
<td>Stratified Series in General</td>
<td>162</td>
</tr>
<tr>
<td>12.1.3</td>
<td>Matched Series</td>
<td>164</td>
</tr>
<tr>
<td>12.2</td>
<td>LR-Based $\chi^2$ Function of RD</td>
<td>165</td>
</tr>
<tr>
<td>12.3</td>
<td>Simple Analysis for RD</td>
<td>166</td>
</tr>
<tr>
<td>12.4</td>
<td>Analysis for Modification of RD</td>
<td>168</td>
</tr>
<tr>
<td>13.</td>
<td>COMPARATIVE ANALYSIS OF TWO RATES III: CHI SQUARE FUNCTION FOR RATE RATIO</td>
<td>172</td>
</tr>
<tr>
<td>13.1</td>
<td>Contrast-Based $\chi^2$ Function of RR</td>
<td>172</td>
</tr>
<tr>
<td>13.1.1</td>
<td>Unstratified Series</td>
<td>172</td>
</tr>
</tbody>
</table>
13.1.2. Stratified Series in General 176
13.1.3. Matched Series 177
13.2. LR-based $X^2$ Function of $RR$ 179
13.3. Simple Analysis for $RR$ 180
13.4. Analysis for Modification of $RR$ 181

14. COMPARATIVE ANALYSIS OF TWO RATES IV: CHI SQUARE FUNCTION FOR ODDS RATIO 184
14.1. Score-Based $X^2$ Function of OR 184
14.1.1. Unstratified Series 184
14.1.2. Stratified Series in General 189
14.1.3 Matched Series 189
14.2. LR-Based $X^2$ Function of OR 191
14.3. Simple Analysis for OR 192
14.4. Analysis for Modification of OR 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 $r^2$ 204
16.2.2. $X^2$ or $r^2$ Function of Mean Difference 206
16.3. Several Means 207
16.3.1. Null $X^2$ or $r^2$ 207
16.3.2. $X^2$ or $r^2$ Function of Regression Slope 210

17. ANALYSIS OF CASE-REFERENT DATA 211
17.1. Null $X^2$ 211
17.2. $X^2$ Function of RR 213
17.3. Simple Analysis for RR 214

18. REGRESSION ANALYSIS 216
18.1. Linear Regression: The Concept 217
18.2. Independent Variables: Specification and Result Interpretation 218
18.2.1. Determinant Representation 218
18.2.1.1. Representation of Nominal Determinant 218
18.2.1.2. Representation of Ordinal or Quantitative Determinant 221
18.2.1.3. Determinant Characterized in Two or More Dimensions 223
18.2.1.4. Determinant History 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 Model 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
| Contents |
|------------------|-----|
| A.1.3. Risk | 249 |
| A.1.4. Proportionate Rate | 250 |

**APPENDIX 2. MEASURES OF OCCURRENCE RELATION**

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.2.1. Fundamental Measures of Susceptibility and Resistance</td>
<td>251</td>
</tr>
<tr>
<td>A.2.2. Susceptibility/Resistance and Measures of Relation</td>
<td>252</td>
</tr>
<tr>
<td>A.2.2.1. Rate Difference</td>
<td>252</td>
</tr>
<tr>
<td>A.2.2.2. Derivatives of Rate Ratio</td>
<td>252</td>
</tr>
<tr>
<td>A.2.3. Derivatives of Observed and Expected Numbers of Cases</td>
<td>254</td>
</tr>
<tr>
<td>A.2.4. Odds Ratio</td>
<td>256</td>
</tr>
<tr>
<td>A.2.5. Correlation and Regression Coefficients</td>
<td>257</td>
</tr>
<tr>
<td>A.2.6. Proportionate Mortality Ratio</td>
<td>260</td>
</tr>
<tr>
<td>A.2.7. Deaths Prevented (Lives Saved)</td>
<td>261</td>
</tr>
</tbody>
</table>

**APPENDIX 3. CAUSAL AND PREVENTIVE INTERDEPENDENCE (SYNERGISM AND ANTAGONISM)**

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.3.1. Definitions</td>
<td>262</td>
</tr>
<tr>
<td>A.3.2. Rate Criteria</td>
<td>264</td>
</tr>
</tbody>
</table>

**APPENDIX 4. CRUDE, ADJUSTED, AND STANDARDIZED RATES**

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.4.1. Crude Rate as a Weighted Average of Specific Rates</td>
<td>266</td>
</tr>
<tr>
<td>A.4.2. Adjustment and Standardization: The Ideas</td>
<td>268</td>
</tr>
<tr>
<td>A.4.3. The Notion of &quot;Indirect&quot; Standardization</td>
<td>269</td>
</tr>
<tr>
<td>A.4.4. The Standardized Mortality (Morbidity) Ratio</td>
<td>270</td>
</tr>
<tr>
<td>A.4.5. Precision-Maximizing Weights</td>
<td>272</td>
</tr>
</tbody>
</table>

**APPENDIX 5. CENSUS VS. CASE-REFERENT APPROACH: RELATIVE INFORMATIVENESS**

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.13.1. Rate Difference</td>
<td>273</td>
</tr>
</tbody>
</table>

**APPENDIX 6. CASE-REFERENT APPROACH: SAMPLE OF BASE (REFERENT) VS. SAMPLE OF NONCASE SUBDOMAIN OF BASE**

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.8.1. Questions of Validity in the Design Stage</td>
<td>277</td>
</tr>
<tr>
<td>A.8.2. Questions of Efficiency in the Design Stage</td>
<td>280</td>
</tr>
<tr>
<td>A.8.3. Questions in the Review of &quot;Positive&quot; Studies</td>
<td>284</td>
</tr>
<tr>
<td>A.8.4. Questions in the Review of &quot;Negative&quot; Studies</td>
<td>285</td>
</tr>
</tbody>
</table>

**APPENDIX 9. DISTRIBUTION OF P-VALUE AND INTERPRETATION OF TEST RESULTS**

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.10.1. Binomial Distributions</td>
<td>286</td>
</tr>
<tr>
<td>A.10.2. Poisson Distributions</td>
<td>294</td>
</tr>
<tr>
<td>A.10.3. Hypergeometric Distributions</td>
<td>295</td>
</tr>
<tr>
<td>A.10.4. Noncentral Hypergeometric Distributions</td>
<td>297</td>
</tr>
<tr>
<td>A.10.5. Gaussian Distributions</td>
<td>298</td>
</tr>
</tbody>
</table>

**APPENDIX 11. P-VALUE FROM STANDARD GAUSSIAN STATISTIC**

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.11.1. Rate Difference</td>
<td>299</td>
</tr>
</tbody>
</table>

**APPENDIX 12. P-VALUE FROM CHI SQUARE STATISTIC**

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.12.1. Rate Difference</td>
<td>301</td>
</tr>
</tbody>
</table>

**APPENDIX 13. ANALYSIS OF MATCHED SERIES: VARIANCE ESTIMATES FOR X² FUNCTIONS OF RATE DIFFERENCE AND RATE RATIO**

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.13.1. Rate Difference</td>
<td>303</td>
</tr>
<tr>
<td>A.13.2. Rate Ratio</td>
<td>304</td>
</tr>
</tbody>
</table>
### Glossary

**Abstract** Theoretical (q.v.).

**Abstract-General** Theoretical (q.v.).

**Accuracy** Degree of lack of error (cf. Precision and Validity).

**Acquired** Postnatal in origin (as opposed to congenital).

**Acute**
1. Of short duration.
2. Of sudden onset.

**Additive Model** A regression model without any term whose $X$ is a product of two other $X$'s in the model representing separate items of information.

**Adjusted Measure of Relation** (e.g., adjusted rate difference) A measure based on conditional comparative information (cf. Crude and Standardized).

**Adjusted Rate** The projection of a crude rate to another (possibly hypothetical) population structure; that is, a weighted average of specific (conditional) rates, with the weights proportional to the sizes of the corresponding experiences in another population (cf. Crude and Standardized).

**Administratively** Having to do with actual delivery of health care in a particularistic setting (as opposed to its scientific premises).
ADMISSIBILITY (into the study base) The property of satisfying the criteria of membership (potential or actual) in the population whose experience constitutes the study base.

ALTERNATIVE HYPOTHESIS
1. In science—a hypothesis entertained as a "competitor" for the hypothesis at issue.
2. In statistics—the range of parameter values corresponding to the hypothesis at issue (as opposed to the null value or "null hypothesis").

ANALYSIS OF DATA The process of summarizing, under a statistical model, the evidence in the data with reference to the object of research.

ANALYTIC EPIDEMIOLOGY Misnomer for epidemiologic hypothesis testing (whose logic is synthetic, not analytic). In terms of a classical misapprehension, "analytic epidemiology" is viewed as the alternative to descriptive epidemiology (q.v.).

ANTAGONISM One factor's inhibition of the effect of another.

APPLIED HEALTH SCIENCE Scientific research with direct implications for health care (i.e., assessment of relative merits of alternative algorithms of care) (cf. BASIC HEALTH SCIENCE).

ASSESSMENT Classification according to a scale of quantity (cf. EVALUATION).

ASSOCIATION Occurrence relation (q.v.).

AT RISK Not known to have zero risk (cf. CANDIDATE).

ATTACK RATE In communicable-disease epidemiology, the proportion of candidates who come down with the illness in a particular (usually short) period of time, as in "secondary attack rate" (the proportion of the "contacts" of cases that comes down with the disease, presumably as a consequence of contact with a case).

ATTRIBUTABLE RISK Misnomer for rate difference (q.v.).

BACKGROUND LEVEL (of a parameter of occurrence) Level determined by characteristics other than that under study.

BASE See STUDY BASE.

BASE POPULATION The population whose experience (in a defined segment of time) constitutes the study base (q.v.).

BASELINE The starting point of the risk period under study, or of follow-up.

BASIC HEALTH SCIENCE Science (q.v.) motivated by intent to provide new theoretical foundations for innovations in health care (cf. APPLIED HEALTH SCIENCE).

BAYESIAN STATISTICS Statistics concerned with the use of data in the modification of the probabilities (credibilities) of various values of the parameter of interest (cf. FREQUENTIST STATISTICS).

BIAS
1. MEAN BIAS Mean error, that is, difference between the mean of empirical values (over a large number of hypothetical replications) and the theoretical value.
2. MEDIAN BIAS Median error, that is, difference between the median of empirical values (over a large number of hypothetical replications) and the theoretical value.
3. Tendency for error on the mean or median.

BIASED BASE A study base formed conditionally on what it shows, or with membership that is outcome-selective (conducive to biased result).

BIASED RESULT A result (empirical value) from a biased study or procedure, that is, a result with median bias.

BIASED SAMPLING Sampling in which the sampling probability depends in an unknown way on a determinant of the observation.

BIASED STUDY A study with invalidity of design (conducive to a biased result).

BINARY Consisting of (or viewed as if consisting of) two categories.

BIOMETRY The discipline of how to study statistical problems in biology and medicine.

BIOSTATISTICS Biometry (q.v.).

BLINDED STUDY Study characterized by arrangements designed to assure the observers' and/or the subjects' lack of awareness of whether the subject is a member of the index or the reference series.

BLOCKING Restriction of randomization; that is, the use of separate randomizations within categories of a covariate, with the aim of assuring similarity of the distribution of the covariate between the index and reference series (in the study base).

CANDIDATE An individual who, by virtue of not having the state (illness) at issue, could logically experience its inception (even if, in light of substantive facts, the risk is known to be zero).

CANDIDATE TIME The amount of candidate experience (in terms of the integral of the size of the candidate population over the period of observation).
CASE
1. In medicine—episode of illness, as in "a case of gonorrhea."
2. In epidemiology—a person representing a case (in the medical sense) of some state or event, as in "the interview of cases."

CASE-CONTROL STUDY Case-referent (case-base) study (q.v.).

CASE-REFERENT STUDY A study involving the sampling design of enrolling all of the cases in the study base (in the direct referent of the empirical result) and, separately, a sample (not census) of the study base itself (or of its noncase segment).

CAUSAL RELATION Occurrence relation reflecting a causal connection between the determinant and the outcome phenomenon (as opposed to confounding by extraneous determinants of the outcome parameter).

CAUSE A category of a determinant, in relation to a particular reference category, capable of completing a sufficient cause in some instances in which the reference category is incapable of such completion. When, in this sense, a category (relative to a particular reference category) can be both a cause (in some instances) and a preventive (in some others), it is commonplace to mean by "cause" that causal instances are more common than preventive ones.

CENSUS 100% sampling, or 100% sample.

CENTILE A fractile expressed in terms of a scale from 0 to 100% for the fraction involved, as in "the median is the 50th centile, or the 0.5 fractile."

CHARACTERISTIC (of person)
1. A dimension in which persons (units of observation) may be characterized, such as any risk indicator.
2. A realization in a dimension in which persons are characterized, such as any risk indication (q.v.).

CHI SQUARE The square of a standard Gaussian variate ($\chi^2$ with 1 df) or a statistic whose distribution conforms to that of the sum of two or more independent $\chi^2$ 1 df variates ($\chi^2$ with 2 or more df).

CHRONIC Long-lasting or long-term.

CHRONIC-DISEASE EPIDEMIOLOGY Misnomer for epidemiology of noninfectious diseases.

CLINICAL Having to do with patient care in a direct way (originally bedside).

COACTION (of factors) Joint action.

COEFFICIENT OF CORRELATION See Correlation Coefficient.

COEFFICIENT OF REGRESSION See Regression Coefficient.

COEFFICIENT OF VARIATION Standard deviation divided by mean.

COHORT A closed population (from Latin, cohors, enclosure); that is, a population whose membership is defined on the basis of some event, for ever after; that is, a population with fixed membership (cf. dynamic population).

COHORT INCIDENCE Proportion-type incidence (cf. incidence density)

COHORT STUDY A study (of incidence or change) whose base is the experience of a cohort over time (as opposed to a dynamic-population study or a cross-sectional study).

COMMUNICABLE Subject to transmission from one person to another, that is, contagious.

COMORBIDITY Presence of associated illness.

COMPARABILITY OF EFFECTS Identity of extraneous effects of the compared empirical categories of the determinant, as in "the treatment under study and the 'placebo' treatment have comparable effects (in studying the drug effect) if the extraneous (nondrug) aspects of the two treatments have identical effects, and if, in addition, the 'placebo' has no effect."

COMPARABILITY OF INFORMATION
1. Absence of differential errors in the outcome information between the compared populations (in the study base).
2. In case-referent studies, also absence of differential errors in the determinant information between the case series and the base sample.

COMPARABILITY OF POPULATIONS Balanced distributions between populations representing the compared categories of the determinant, specifically distributions with respect to extraneous determinants of the outcome parameter (i.e., the property that randomization and "blocking" are designed to achieve in experiments).

COMPARATIVE STUDY A study involving separate representations of the compared categories of the determinant (as opposed to a before-after study).

CONDITIONAL STATISTIC (estimate or test statistic) Statistic derived with a restriction as to either the realization for an ancillary statistic (e.g., total number of cases), parameter space (e.g., all stratum-specific values of comparative parameter equal to a given theoretical value), or covariate.

CONFIDENCE INTERVAL See Estimate.

CONfounder An extraneous determinant of the outcome parameter in terms of which there is lack of comparability (q.v.) of effects and/or
DESIGN MATRIX  Distribution matrix (q.v.).

DETERMINANT (of occurrence)  A characteristic of individuals (constitutional, behavioral, or environmental) on which a parameter (q.v.) of occurrence depends (causally or noncausally).

DIAGNOSIS
1. The inferred (rather than necessarily the actual) state of, or event in, a person (or an object).
2. The process of arriving at an inference as to the state of, or event in, a person (or an object).

DICHTOMOUS  Divided in two categories (cf. POLYTOMOUS and TRICHOTOMOUS).

DIFFERENTIAL (selection, errors, etc.)  Operating differently between/ among the compared categories or series.

DIRECTLY STANDARDIZED  Mismner for standardized (q.v.). (Cf. INDIRECTLY STANDARDIZED)

DISCIPLINE  Field learning and practice.

DISEASE  A process of illness, as in “neoplastic disease” (in Latin, morbus; cf. DEFECT and ILLNESS)

DISTRIBUTION MATRIX  The distribution of the base experience according to the determinant(s), modifiers, and confounders in the occurrence relation.

DOMAIN OF STUDY  The type of situation in which the object of the study (occurrence relation) will be (was) explored, as in “the effects of Indomethacin in the domain of the premature neonate with patent ductus arteriosus.”

DOSE-RESPONSE
1. Dependence between the level of a determinant (causal or preventive) and the magnitude of effect, as in “there was evidence of dose-response.”
2. The relation between the level of a determinant and the magnitude of the effect, as in “linear dose-response.”

DOUBLE BLINDED  Blinded with respect to both the observer(s) and the subjects.

DYNAMIC POPULATION  An open population; that is, a population whose membership is defined on the basis of some state, for the duration of that state; that is, a population with turnover of membership (cf. COHORT).

DYNAMIC-POPULATION STUDY  A study (of incidence) whose base is the experience of a dynamic population (of candidates) over time [as opposed to a cohort study (q.v.) or a cross-sectional study (q.v.).]

EFFECT
1. In science—the change in the outcome parameter brought about by a particular cause or preventive.
2. In statistics—the value of a parameter of relation, as in “the main effect of X.”

EFFECTIVENESS  Efficacy (q.v.)

EFFICACY (of study)  Ability to bring about the intended change in the outcome or outcome parameter.

EFFICIENCY (of study)
1. Informativeness with a given cost (cost-efficiency).
2. Informativeness with a given size (size-efficiency).

EMPIRICAL
1. Operational (as opposed to conceptual), as in “empirical scale.”
2. Based directly on experience, as in “empirical rate.”

ENDPOINT  An event that terminates follow-up, especially the event whose occurrence is under study.

ENDEMIC OCCURRENCE  The usual rate of occurrence (cf. EPIDEMIC).

ENTRY COHORT  A cohort defined as of the beginning of the risk period under study (i.e., at “baseline”).

ENVIRONMENTAL  Having to do with the settings in which individuals (units of observation) exist [as opposed to constitutional (q.v.) or behavioral].

EPIDEMIC  An episode, or relating to an episode, of unusually common occurrence.

EPIDEMIOLOGY
1. That aspect of a medical science or discipline which deals with the (frequency of the) occurrence of phenomena of interest in that field, as in “the epidemiology of oncology” (neoplastic or cancer epidemiology), or “health care epidemiology.”
2. The (frequency of) occurrence aspect of a phenomenon, as in “the epidemiology of sudden death.”
3. Research into the occurrence (epidemiologic) aspect of a phenomenon, as in “practice of epidemiology.”

EPistemologic  Having to do with theoretical (rather than procedural) aspects of how to learn about the objects of research, that is, with theory of methodology (cf. ONTOLOGIC).
Glossary

**Estimate**

1. **Interval estimate** (confidence interval) An interval (for a parameter) constructed (from a particular set of data) in a way designed to assure that in a specified proportion of such applications (out of a large number of them) it covers the actual value of the parameter.

2. **Point estimate** A 0% two-sided confidence interval; that is, a 50% one-sided confidence bound; that is, a value (for a parameter) computed (from a particular set of data) in a way designed to assure that in 50% of such applications (out of a large number of them) it exceeds, and in the other 50% it falls short of, the actual value of the parameter (cf. Bias).

**Estimator** A function of data, designed to provide estimates; that is, a way (conceptual) of computing an estimate (q.v.) from data.

**Etiologic fraction** Of the actual total rate of occurrence (total load of cases), the proportion that is attributable to the cause at issue. It equals the proportion of actual cases that is attributable to the cause if there are no instances of preventive susceptibility (cf. Preventive fraction.)

**Etiology** Causal origin, or causal explanation.

**Evaluation** Classification according to a scale of quality or preference (cf. Outcome evaluation, Practice evaluation, Premise evaluation, and Process evaluation). Also used (as a misnomer) for mere assessment (q.v.).

**Event**

1. An episode.
2. The transition (rapid) from one state to another.

**Exact model/distribution** A model/distribution that represents exactly the sampling distribution of data on the premise that the study experience represents a probability sample of the abstract reality ("superpopulation") under study.

**Expectation** See Expected value

**Expected value** Mean of sampling distribution (q.v.), often, implicitly, on the "null hypotheses" (q.v.).

**Experiment** A study in which a determinant is intentionally perturbed for reasons none other than the goals of the study itself.

**Explanatory study** (of exposure or intervention) Study that addresses the effect of a particular agent in the exposure or intervention (cf. Pragmatic study).

**Exposure**

1. A particular (index) category of a (potential) environmental (as opposed to constitutional or behavioral) determinant, entertained as a cause or preventive for the phenomenon of interest, relative to a particular other (reference) category.
2. Experiencing the index (exposure) category of an environmental determinant (potential) of the outcome.
3. Environmental determinant.
4. Misnomer for any determinant (i.e., constitutional or behavioral as well as environmental), or its index category.

**Exposure odds** In a case-referent study, the theoretical proportion of cases or members of the reference series representing the index category of an environmental determinant divided by the proportion representing the reference category.

**Exposure-odds ratio** In a case-referent study, the exposure odds of cases divided by that of reference subjects.

**External validity** Generalizability beyond the experience of the study (cf. Internal validity).

**Extraneous characteristic/factor** A characteristic/factor that is not part of the object of research (occurrence relation) per se, but that may have to be considered in the interest of validity or efficiency (as a conditioning factor); that is, a characteristic/factor that a covariate (q.v.) represents.

**Factor**

1. A cause or a preventive (e.g., a category of blood pressure relative to another).
2. A causal or preventive determinant (e.g., blood pressure).
3. Misnomer for determinant (q.v.) or a category of it.

**False negative rate** The complement of sensitivity; that is, one minus sensitivity (q.v.).

**False positive rate** The complement of specificity; that is, one minus specificity (q.v.).

**Fractile** The value of a characteristic that corresponds to a given cumulative fraction (proportion) of its realizations, as in "the median is the 0.5 fractile of a distribution."

**Frequentist statistics** Statistics concerned with the frequency behavior (sampling distribution) of quantities (statistics) derived from data, without addressing the probabilities (credibilities) of various values of the parameter of interest (cf. Bayesian statistics).
FUNCTION  Relation between a dependent quantity and its determinant(s).

GENERAL EPIDEMIOLOGY  Theoretical epidemiology (q.v.).

GENERAL POPULATION  A broadly defined (and thus ill-defined) population (and a concept of inherently no value in science).

GENERALIZATION
1. In science—indeference (inductive) from the empirical (particularistic) to the theoretical (abstract-general).
2. In statistics—indeference (usually "frequentist"-mechanistic) from a sample to a population (particularistic).

GENETIC  Having to do with the genome (chromosomes).

HEALTH  The state of being, or the extent to which one is, as functional and symptom-free as is commonly attainable at the age at issue, and freedom from unusual (for age) constitutional indications of future disability, symptoms or death; in other words, freedom from illness (q.v.) or the extent to which one is free from illness.

HEALTH CARE  Any activity or arrangement aimed at the maintenance or improvement of health (whether education, regulation, or service; whether preventive or therapeutic) or adaptation to illness (rehabilitative).

HEALTH-CARE RESEARCH  Research into the realities (as opposed to premises) of health care (cf. HEALTH SCIENCE).

HEALTH SCIENCE  Science (q.v.), either basic (q.v.) or applied (q.v.), relevant to health care (its premises).

HEALTHY WORKER EFFECT  Tendency of occupationally defined populations to show lower mortality than "the general population," conditionally on controllable covariates (usually age, gender, race, and calendar time).

HOMOSCEDASTICITY  Constancy of the variance (theoretical) of the outcome measure over the determinants at issue.

HYPOTHESIS  A tentative piece of scientific (theoretical) knowledge; that is, a scientific idea (cf. ALTERNATIVE HYPOTHESIS and NULL HYPOTHESIS).

ILLNESS  Ill-health; that is, disease (q.v.) or defect (q.v.).

IMPRECISION  Lack of total precision (q.v.).

INCIDENCE  The appearance of events (q.v.) of a particular kind in a population (of candidates over time) (cf. PREVALENCE).

INCIDENCE DENSITY  The ratio of the number of events to the corresponding population time (candidate time).

INCIDENCE PROPORTION  See COHORT INCIDENCE and CUMULATIVE INCIDENCE.

INCIDENT CASES  Cases that appear (as against those that exist or prevail).

INCONCLUSIVE STUDY  A study whose result is reasonably consistent with both the hypothesis at issue and its denial (cf. NEGATIVE STUDY and POSITIVE STUDY).

INDEPENDENCE  Lack of association (relation) or of synergism and antagonism.

INDEPENDENT VARIATE  A variate (statistical) representing a determinant of the dependent parameter in a regression model.

INDEX  A measure that provides for ranking.

INDEX CATEGORY  (of a determinant) A category that is of express interest, for example, a particular new treatment under study (as opposed to the reference category of the "standard" treatment or, alternatively, no treatment).

INDEX EXPERIENCE/GROUP/POPULATION/SERIES
1. The experience/group/population/series representing an index category (in the study base).
2. The experience/group/population/series around which a study base is built (as is commonplace in case-referent studies, for example).

INDICATION
1. Realization of an indicator, indicative of the state or event at issue, as in "symptomatic improvement as an indication of therapeutic efficacy," or "test positivity as an indication of disease."
2. Situation that prompts a particular type of action, as in "the small size of the index series as an indication for matching."

INDICATOR  A characteristic (variates) whose realization, or a test whose outcome, conveys information about the presence or future occurrence of the phenomenon at issue.

INDICATOR VARIATE  Statistical variate representing a binary characteristic, usually with values 1 and 0 corresponding to the two categories, respectively.

INDIRECTLY STANDARDIZED  Misnomer for standardized according to the distribution of the index experience. (This means standardization between any given index rate and the reference rate, but not among two or more index rates.)
**Induction period** The time lag from the beginning of the causal process to the manifestation of the effect.

**Infectious** Due to a microorganism.

**Inference (in statistical science)** The movement, presumptive, from the results of data analysis toward new knowledge (q.v.) about the object of study; that is, interpretation of statistics.

**Information**

1. In statistics— with reference to a particular parameter, the inverse of the sum of variance and the square of mean bias of its point estimator; that is, the inverse of the average of squared deviations of point estimates from the parameter value itself.
2. In statistical science—data and statistics serving as evidence for the advancement of knowledge about the object of study.

**Informativeness (of study)** The degree of evidence from a study (as determined by validity, efficiency, and size).

**Interaction** Mutual influence between two factors (the factors’ influence on each other). Also used, as a misnomer, for interdependence (q.v.) of coaction and the modification (q.v.) of relation (need for product term in regression analysis).

**Intercept** The value (actual or hypothetical) of the dependent parameter in a regression function that corresponds to all the independent variates in the function equaling zero.

**Interdependence (of coaction)** Dependence of the effects of two factors on each other; that is, synergism (q.v.) or antagonism (q.v.).

**Internal validity** Validity with reference to the study base itself (the direct referent of the study result; cf. External validity).

**Interval estimate** See Estimate.

**Intervention** Willful perturbation of a determinant, aimed at influencing the outcome.

**Knowledge (scientific)** Strong expert belief, especially a consensus belief, about a scientific proposition.

**Latency period** The span of time for which a state remains inapparent.

**Level of test** An a priori ‘‘critical’’ level (α) for a P-value, such that P < α is deemed ‘‘significant’’ (at level α), and P > α is considered ‘‘non-significant,’’ in the sense of calling for ‘‘acceptance’’ and ‘‘rejection’’ of the hypothesis, respectively.

**Link** Lay location for causal connection.

**Linked** Lay location for causally connected.

**Logit** Logarithm (natural) of odds (q.v.).

**Longitudinal relation (between outcome parameter and its determinant)** A relation in which the time referent of the information on the determinant is antecedent to that on the outcome parameter by an amount consonant with the nature of the theoretical relation under study (e.g., length of the induction period) (cf. Cross-sectional).

**Longitudinal study** A study whose base is a population’s experience over time (as opposed to a population cross-section).

**Main effect** Regression coefficient of a nonproduct term.

**Malformation** An anatomic defect of developmental origin.

**Matching**

1. Selection of the base sample in such a way that its distribution according to a covariate becomes more similar to that of the case series than would tend to be the case if it were selected independently of the case series.
2. Selection of the comparison group for the study base in such a way that its distribution according to a covariate becomes more similar to that of the other group than would tend to be the case if it were selected independently.

**Matching ratio** The ratio of the size of a matched base sample (casereferent matching) or matched comparison group (base matching) to that of the index series/group.

**Median bias** See Bias.

**Median unbiased estimator** Estimator whose sampling distribution has the actual parameter value as its median (cf. Bias).

**Medical**

1. Having to do with medicine, that is, with the prevention, diagnosis, or treatment of, or rehabilitation for, illness.
2. Having to do with medicinal (as opposed to surgical) therapy.

**Medicine**

1. The health field, as in ‘‘preventive medicine.’’
2. The field of diagnosis and medicinal treatment of internal, somatic illness.

**Metameter (of a measure)** A transform of a measure.

**Method** A planned means of achieving a preset end (as opposed to serendipity) (cf. Principle).
Glossary

**Methodology** The body of methods in a discipline (cf. **Epistemology**).

**Model** A formal, simplified representation of, or the terms in which to consider, the object of study.

**Modification** Inconstancy of a parameter of occurrence relation over another subject characteristic.

**Modifier (of a relation)** A characteristic (of individuals) on which a parameter of occurrence relation depends.

**MOR** See **Mortality Odds Ratio**.

**Morbidity** The occurrence of illness.

**Mortality** The occurrence of death.

**Mortality Odds** Occurrence odds (q.v.) for a given cause of death relative to another (or a series of others).

**Mortality Odds Ratio** Ratio of mortality odds between two categories of a determinant.

**Negative Confounding** Confounding that reduces, or even reverses, the empirical relation (relative to the theoretical one).

**Negative Study** A study that detracts from the hypothesis at issue (cf. **Inconclusive Study and Positive Study**).

**Nominal P-Value** A P-value that is not subject to the usual interpretation.

**Nominal Scale** A scale without any quantitative implications (not even those of ranking) (cf. **Ordinal Scale and Quantitative Scale**).

**Nonexperimental Study** A study that is not experimental (q.v.).

**Nosocomial** Having to do with hospitals.

**Nuisance Parameter** An extraneous parameter that needs to be estimated (in assessing a parameter of express interest).

**Null Distribution** The distribution that corresponds to the null hypothesis (q.v.).

**Null Expectation** The mean of null distribution (q.v.).

**Null Hypothesis** The denial, parsimonious, of a hypothesis (e.g., the proposition of no relation between an outcome parameter and its potential determinants, given a hypothesis of such a relation).

**Null Value** Parameter value corresponding to null hypothesis

**Null X²** Statistic designed to have chi square (q.v.) distribution on the null hypothesis (q.v.).

**Object of Study** That which is under study (in epidemiology, an occurrence relation).

**Objective of Study** That which is to be achieved by a study (i.e., advancement of knowledge about the object of study).

**Objectivity** The quality of being subject to agreement by independent qualified examiners.

**Oblique Cohort** A cohort whose zero time is distributed across the risk period under study (cf. **Entry Cohort**).

**Observational Study** Misnomer for nonexperimental study (q.v.). All empirical studies, experimental as well as nonexperimental, are observational (i.e., based on observation).

**Occurrence**
1. A particular instance of a phenomenon, as in "this occurrence of complication."
2. The frequency (statistical) aspect of a phenomenon, as in "the occurrence of complications."

**Occurrence Function** Occurrence relation (q.v.).

**Occurrence Odds**
1. For a single theoretical rate of the form of a proportion (and for risk), the ratio of that rate (risk) to its complement (i.e., to one minus that rate or risk).
2. For two theoretical rates, having to do with alternative types of state or event in the same base, the ratio of the rate for one type to that for the other.

**Occurrence Relation** The relation of a parameter of occurrence (e.g., incidence rate) to one or more characteristics of persons (or other units of observation).

**Odds** Ratio of probability (P) to its complement (1-P) [i.e., P/(1-P)]. (See **Exposure Odds** and **Occurrence Odds**).

**Odds Ratio** The ratio of two odds (q.v.).

**O/E Ratio** Observed-to-expected ratio (as to number of cases; for expected number, see **Null Number**).

**Ontologic** Having to do with theoretical aspects of how to view nature and formulate the objects of research (cf. **Epistemologic**).

**OR** Odds ratio (q.v.).

**Ordinal (Quasi-Quantitative) Scale** A scale with implications for ranking, but not for absolute quantification of differences or ratios (cf. **Nominal Scale** and **Quantitative Scale**).
OUTCOME The end stage of a process, as in "the health outcome following treatment" (distinct from the effect of treatment).

OUTCOME EVALUATION (of health care) Premise evaluation (q.v.), or indirect process evaluation (q.v.).

OVERMATCHING Matching that reduces the amount of information per subject; that is, matching of base sample to case series on a covariate that will not be controlled as a confounder nor explored as a potential modifier but is a correlate of the determinant in the study base.

P-VALUE A statistic so derived that its sampling distribution is on the "null hypothesis" uniform in the zero-to-one range, with a shift to the left in this range on the hypothesis itself.

PARAMETER
1. In medicine—a characteristic of interest in individual patients, as in "diagnostic parameter."
2. In mathematics—a constant (as opposed to a variable).
3. In mathematical statistics—a constant in a statistical (distribution) model.
4. In epidemiology—a population measure of occurrence, or occurrence relation, empirical or theoretical.

PARSIMONIOUS Conforming to the principle of parsimony.

PARSIMONY (principle of)
1. Regarding concepts—one is to restrict the invocation of concepts to the minimum sufficient.
2. Regarding ideas—one is not to accept hypotheses (alternatives to the parsimonious null outlook) except in the face of reasonably compelling evidence.

PARTICULARISTIC Specific to a particular time and/or place (as opposed to abstract-general).

PARTICULARISTIC STUDY A study whose ultimate object is particularistic (administrative, rather than scientific).

PERCENTILE Centile (q.v.).

PHENOMENON A state or event.

PMR See PROPORTIONATE MORTALITY RATIO.

POINT ESTIMATE See ESTIMATE.

POLYCHOTOMOUS Misnomer for polytomous.

POLY TOMOUS Divided in several categories (cf. DICOTOMOUS and TRICHOTOMOUS).

POPULATION See COHORT, DYNAMIC POPULATION, and CROSS-SECTIONAL POPULATION.

POPULATION ATTRIBUTABLE RISK Etiologic fraction (q.v.).

POPULATION-BASED STUDY Misnomer for study with primary base (q.v.).

POPULATION CROSS-SECTION See CROSS-SECTIONAL POPULATION.

POPULATION TIME The amount of population experience in terms of the integral of population size over the period of observation (cf. CANDIDATE TIME).

POSITIVE CONFOUNDING Confounding that exaggerates the empirical relation (cf. NEGATIVE CONFOUNDING).

POSITIVE STUDY A study that supports the hypothesis at issue (cf. INCONCLUSIVE STUDY and NEGATIVE STUDY).

POSTERIOR PROBABILITY Credibility in light of the evidence from the study (cf. PRIOR PROBABILITY).

POTENTIAL CONFOUNDER An extraneous characteristic that, in a priori terms, cannot be deemed not to be a confounder (q.v.).

POWER OF TEST (statistical) Probability that the P-value will be less than the chosen level of the test conditionally on a particular (nonnull) value of the parameter at issue.

PRACTICE (of health care) Activity undertaken on the presumption that it preserves or promotes health (rather than for learning whether it does) (cf. PRACTICE EVALUATION and PROCESS EVALUATION).

PRACTICE EVALUATION Research (administrative) aimed at description (classification) of practice in terms of a scale of quality (cf. PROCESS EVALUATION).

PRAGMATIC STUDY (of exposure or intervention) Study that addresses the effect of the exposure/intervention in terms of its operational definition, without regard to the particular agent(s) involved (cf. EXPLANATORY STUDY).

PRECISION Reproducibility; that is, degree of lack of random error (as opposed to bias) in quantification; that is, degree of constancy of error in quantification.

PRECURRENCE PERIOD The waiting time until the occurrence of the event at issue.

PREMISE EVALUATION Research (scientific) into the tenability of a premise (scientific) of health care.
PREVALENCE The existence (as opposed to the inception or termination) of a particular state among the members of a population.

PREVALENCE RATE The proportion of a population that is in a particular state.

PREVALENT CASES Cases that exist (as of given point in time).

PREVENTIVE 1. A category of a determinant, in relation to a particular reference category, incapable of completing a sufficient cause in some instances in which the reference category is capable of it (particularly if these instances are more common than those of causation); that is, the reference category of a cause.

2. A category of a determinant, in relation to a particular reference category, capable of blocking (“neutralizing,” or “frustrating”) a sufficient cause in some instances in which the reference category is incapable of it (particularly if these instances are more common than the converse ones).

PREVENTIVE FRACTION Of the hypothetical total rate of occurrence (total load of cases) that would obtain were the preventive to be totally absent, the proportional reduction that is attributable to the preventive (cf. ETIOLOGIC FRACTION).

PRIMARY BASE A base defined directly (rather than indirectly through a series of cases) (cf. SECONDARY BASE).

PRIMARY DIAGNOSIS Diagnosis related to the reasons for, or causes of, the patient’s coming to attention.

PRIMARY INFORMATION Information in the sense of elementary data (as opposed to derived data or information).

PRIMARY OBJECTIVES (of a study) Objectives that justify the study (as opposed to those that the study justifies, i.e., secondary objectives).

PRINCIPLE (of research) A guideline for decision-making in research (ontologic as well as epistemologic or methodologic).

PRIOR PROBABILITY Credibility in the absence of the evidence from the study (cf. POSTERIOR PROBABILITY).

PROBABILITY 1. Relative frequency, as in “the probability that a 95% confidence interval will cover the value of the parameter is 95%.”

2. Credibility, as in “the prior probability of the hypothesis.”

PROBABILITY SAMPLE A sample resulting from probability sampling (q.v.).

PROBABILITY SAMPLING Sampling in such a way that each member of the sampled population has a known and independent probability (chance) of selection into the sample.

PROCESS EVALUATION (of health care) Practice evaluation (q.v.). [Note: quality scales have to do with process, not outcome, so that outcome evaluation (q.v.) addresses premises of practice rather than practice (health care) itself, except indirectly.]

PROGNOSIS Expected future course in the sense of expected utility (disutility). It is determined by the utilities of the various possible outcomes together with their respective probabilities.

PROGRAM A regimented undertaking of indefinite duration (cf. PROJECT).

PROJECT A regimented undertaking of a priori limited duration (cf. PROGRAM).

PROPORTIONAL RATE Proportionate rate (q.v.).

PROPORTIONATE MORTALITY RATIO Ratio of proportionate rate for mortality between two categories of a determinant.

PROPORTIONATE RATE The proportion that (the rate for) a given type of state or event represents out of (the rate for) two or more types of state or event.

PROSPECTIVE BASE A study base whose experience is concurrent with the execution of the study.

PROSPECTIVE EXPERIENCE (of a cohort) The experience subsequent to the time as of which the cohort is defined (zero time).

PROSPECTIVE STUDY A study with a prospective base.

PROTOCOL OF STUDY A written documentation of a study plan.

PUBLIC HEALTH 1. The health of a community of people.

2. Care for the health of a community of people.

3. Social medicine; that is, public (social) aspects of health and health care (personnel, organization, finance, program policy, etc.).

QUALITATIVE SCALE A nominal scale.

QUALITATIVE STUDY A study whose objective is modification of the credibility of a hypothesis.

QUANTITATIVE SCALE A scale that provides not only for ranking but also for quantification of differences (interval scale) and/or ratios (ratio scale).
Glossary

QUANTITATIVE STUDY A study whose object is the magnitude of a parameter (e.g., of relation) in a general way (rather than in the limited sense of a qualitative study).

QUANTITATIVE METHODS Statistical methods (a recent euphemism).

QUASI-EXPERIMENTAL Noneexperimental.

QUASI-QUANTITATIVE (scale) Ordinal, with scores assigned to categories.

RANDOM ERROR The difference between error and its average (in quantification).

RANDOM SAMPLE A probability sample (q.v.).

RATE (of occurrence) Frequency in population experience (empirical or theoretical).

RATE DIFFERENCE Difference of rate between two categories of a determinant, especially the rate in an index category expressed as its difference from that in a reference category.

RATE RATIO Ratio of rate between two categories of a determinant, especially the rate in an index category expressed as its rate to that in a reference category. (cf. RELATIVE RATE.)

RD Rate difference (q.v.).

REALIZATION Value or category occurring in a particular instance.

REFERENCE CATEGORY The category (of a determinant) that is viewed as the alternative to a category of direct interest; that is, the category that provides comparative, or reference, information for the experience in the index category (often about the null value of the outcome parameter), as in "the rate ratio for the heavily exposed, with never exposed" as the reference category."

REFERENCE GROUP/POPULATION The group/population representing the reference category (in the study base).

REFERENCE SAMPLE A series of individuals (units of observation) enrolled to supply reference information for a case series; that is, a sample of the study base (in a case-referent study).

REFERENT That to which something refers, as in "the referent of the empirical occurrence relation, in direct, technical terms, is the study base.""

REGRESSION The relation of a parameter to its determinants(s).

REGRESSION COEFFICIENT (in linear regression) The ratio of the "effect" of a unit increase in a determinant (on the dependent parameter) to that unit increase.

REGRESSION TOWARD THE MEAN A tendency of a repeat observation, after an appreciable period of time, to be closer to the mean than the original one.

RELATION See OCCURRENCE RELATION.

RELATIVE RATE The magnitude of a rate (an index rate) expressed in terms of its ratio to a reference rate.

RELATIVE RISK The magnitude of a risk (an index risk) expressed in terms of its ratio to a reference risk; also a misnomer for relative rate (q.v.).

RELIABILITY Misnomer for reproducibility.

REPRESENTATIVE OF THE STUDY BASE A member of the base sample.

REPRESENTATIVE SAMPLE
1. Unbiased sample.
2. Sample whose distribution conforms to that of the sampled population.

REPRODUCIBILITY Precision (q.v.).

RESEARCH Activity aimed at the advancement of knowledge (scientific or nonscientific).

RESISTANT Not susceptible (q.v.).

RESPONSIVENESS Susceptibility.

RESTRICTED See CONDITIONAL.

RESULT OF STUDY
1. A product of data analysis; that is, a statistic, as in "the result section of a paper."
2. A consequence of a study (change in view).

RETROSPECTIVE BASE A study base whose experience is antecedent to the implementation of the study.

RETROSPECTIVE EXPERIENCE (of a cohort) The experience preceding the time as of which the cohort is defined (zero time).

RETROSPECTIVE STUDY A study with a retrospective base. (Some use the term, as a misnomer, for a case-referent study).

RISK The probability that an event (unfavorable) will occur.

RISK FACTOR Causal risk indicator; also used (as a misnomer) as a synonym for risk indicator (q.v.).

RISK FUNCTION A formal (mathematical) expression of risk in relation to its determinants (indicators of risk).
Glossary

**Risk indication** Realization of risk indicator (q.v.).

**Risk indicator** A characteristic on which risk depends; that is, a determinant of risk; that is, a determinant of expected (theoretical) incidence for people of the type at issue.

**Risk ratio** Ratio of risk between two categories of a determinant, especially in the sense of relative risk (q.v.).

**RR** Rate ratio (q.v.) or risk ratio (q.v.).

**Sampling distribution** (of a statistic) Distribution in hypothetical replications of the study, infinite in number.

**Sampling frame** A list of the members of a population, used as a basis for sampling.

**Scale** A set of mutually exclusive and all-inclusive categories

**Science (natural)**
1. Scientific knowledge—a systematized body of theoretical knowledge (about a particular category of natural phenomena), as in "the insights of the science of neoplasia (oncology)."
2. Scientific research—activity aimed at the advancement of theoretical knowledge (about nature), and conforming to the rules of logic, as in "good science."

**Screening** Routine (particularly mass) examination of individuals for indications of illness or of high risk for illness.

**Secondary base** A base defined, indirectly, as the population experience, the entirety of it, from which each potential case, had it occurred, would have been enrolled in the case series of the study.

**Secondary diagnosis** Diagnosis unrelated to the reasons for the patient’s coming to attention.

**Secondary objectives** Objectives that the study justifies (as opposed to those that justify the study, i.e., primary objectives).

**Sensitivity of indication**
1. The proportion (theoretical) of instances in which the indication is manifest when the state at issue is present, or when the event at issue has taken place (cf. Specificity and False Negative Rate).
2. The proportion (theoretical) of instances in which an indication comes about (the indicator turns positive) as a reflection of the event at issue.

**Sensitivity of study** See Power of test.

**Significance level** See Level of test.

**Significant (statistically)** See Level of test.

**Simple random sampling** Probability sampling with identical selection probabilities (cf. Stratified Sampling).

**Single blinded** Blinded with respect to the observers or the subjects but not both.

**Size efficiency** See Efficiency.

**Social medicine** See Public health.

**SMR** See Standardized Mortality Ratio.

**Source population** A population from which the base population is selected (by the use of exclusions).

**Spatiotemporal** Having to do with place and time.

**Specificity (of indication)**
1. The proportion (theoretical) of instances in which the indication is absent when the state at issue is absent, or the event at issue has not taken place (cf. Sensitivity and False Positive Rate).
2. The degree to which the indication, when present, is indicative of the state or event at issue.

**Spurious**
1. Apparent (as against real).
2. Misnomer for confounded (as against causal/preventive).

**Standard deviation** Square root of variance (q.v.).

**Standard error** Estimate (point) of the standard deviation of the sampling distribution (of a point estimator of a parameter).

**Standardized rate** A rate adjusted to a structure (distribution, set of weights) viewed as a standard (common structure) for the purpose of comparison(s).

**Standardized rate difference/ratio** Difference/ratio of two mutually standardized rates. (This defines an internally standardized rate difference/ratio; two or more mutually standardized measures of association involve a common internal standard.)

**Standardized mortality ratio** Empirical rate ratio for mortality involving incidence density of death and standardization according to the distribution of the index experience: that is, 

\[ \frac{O}{E} \]

ratio of deaths, with 

\[ E \]

estimated with allowance for some covariates (usually age, gender, and calendar time).

**Stationary population** A dynamic population whose profile is invariant over time.
STATISTIC A measure summarizing (some aspect of) the evidence in the data.

STATISTICAL SCIENCE Scientific research concerned with frequency of occurrence of a phenomenon (as in epidemiologic research within various medical sciences); it is not a science unto itself.

STATISTICS
1. The discipline that deals with collecting, summarizing, and interpreting data (demographic statistics).
2. The branch of mathematics that deals with random variates (mathematical statistics).

STRATIFIED ANALYSIS Analysis in which elementary comparative information (about the studied relation) is abstracted within subdomains (strata) of some covariates, with such information then accumulated over the strata (in the form of summary statistics).

STRATIFIED SAMPLING Sampling designed to have unequal sampling fractions among subdomains (strata) of the sampled population.

STUDY (in science) A project designed to yield evidence for the advancement of knowledge (q.v.).

STUDY BASE The population experience (particularistic) captured in a study (as a basis for inference about nature in the abstract); that is, the direct referent of the empirical result of a study; specifically, in occurrence research, the study base is the population experience manifesting the occurrence under study, but the relevant realizations of its determinants may be previous to the time segment that the base represents.

STUDY DESIGN
1. A plan as to the type of evidence a study is to yield, and as to the approach to obtaining and summarizing such information.
2. The process of developing a study plan.

STUDY POPULATION The population whose experience constitutes the study base.

STUDY SUBJECT A member of the study population, especially one on whom information is obtained and incorporated in the analysis.

SUBJECT-MATTER The phenomenon under study and related phenomena (as opposed to the principles or methods of studying it).

SUBJECTIVE Not objective (q.v.); involving view(s).

SUBSTANTIVE Having to do with subject-matter (rather than principles or methods of studying it).

SUPERPOPULATION An abstract population of infinite size, viewed as the source of the studied (base) population, and also as the target population of inference.

SURVEY A particularistic and descriptive study.

SUSCEPTIBLE (person or unit of observation)
1. Causally susceptible—the event at issue would/does develop conditionally on the index but not the reference category of the determinant at issue.
2. Preventively susceptible—the event at issue would/does develop conditionally on the reference but not the index category of the determinant at issue.
3. Responsive to the index, but not the reference, category of the determinant.

SYNERGISM Cooperative action between two factors; the effect of one is enhanced by the other (and conversely).

TARGET POPULATION The population of interest in a particularistic study.

TEST (statistical)
1. Frequentist—computation of P-value.
2. Bayesian—computation of likelihood ratio, or likelihood ratio function, and then the posterior probability corresponding to a particular prior probability (or its distribution).

THEORETICAL
1. Having to do with principles (general) rather than subject-matter knowledge, as in "theoretical epidemiology."
2. Without specificity as to time or place; that is, abstract or abstract-general, as in "the theoretical proposition that smoking causes lung cancer," and "theoretical rate."

THEORETICAL EPIDEMIOLOGY The discipline, or principles, of studying epidemiologic problems, that is, frequency of occurrence of phenomena of health-care interest in population experiences. See EPIDEMIOLOGY.

THERAPEUTIC Having to do with treatment of cases of illness (as opposed to prevention or rehabilitation).

TRACKING Incomplete regression toward the mean; that is, correlation between repeat observations. (This concept has meaning in quantitative terms only.)

TRICHOTOMOUS Divided in three categories (cf. DICHOTOMOUS and POLYHOMOUS).

TYPE I ERROR The "rejection" of a correct "null hypothesis."
Glossary

TYPE II ERROR  The "acceptance" of an incorrect "null hypothesis."

UNIT OF OBSERVATION  A member of the study population.

VALIDITY OF MEASUREMENT (quantitative)  Lack of bias (q.v.).

VALIDITY OF STUDY  Lack of mediana bias (q.v.).

VARIABLE

1. Subject to variation.
2. Variate.

VARIANCE  Average of squared deviations from the mean.

VARIATE  A characteristic as represented in a statistical model, potentially
taking on differing realizations among the units of observation (the members
of the study population).

ZERO TIME (of a cohort)  The time (individual or calendar) as of which the
criteria of membership are satisfied; that is, the boundary between the
retrospective and prospective experiences of a cohort.

References


