Epidemiologic Analysis With A Programmable Calculator

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Public Health Service
National Institutes of Health

Kenneth J. Rothman
Department of Epidemiology
School of Public Health
Harvard University
Boston, Massachusetts 02115

and

John D. Boice, Jr.
Environmental Epidemiology Branch
National Cancer Institute
Bethesda, Maryland 20205

NIH Publication No. 79-1649
June 1979

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C. 20402
Stock No. 017-042-00143-9
CONTENTS

PREFACE ........................................................................................................................................ v

ACKNOWLEDGMENT .................................................................................................................. vii

1 INTRODUCTION ....................................................................................................................... 1-4

2 CASE-CONTROL STUDIES ...................................................................................................... 5-10
   Program 1: Odds ratio estimation and testing (7 or fewer 2 × 2 tables) ................................................................................. 6
   Program 2: Odds ratio estimation and testing (unlimited number of 2 × 2 tables) ................................................................. 8
   Program 3: Mantel-extension test and standardized risk ratios for 2 × R tables ................................................................. 9

3 FOLLOW-UP (COHORT) STUDIES ........................................................................................... 11-17
   Program 4: Estimation and testing for follow-up data with count denominators (6 or fewer 2 × 2 tables) ........................................ 13
   Program 5: Estimation and testing for follow-up data with count denominators (unlimited number of 2 × 2 tables) ....................... 14
   Program 6: Estimation and testing for follow-up data with person-time denominators (6 or fewer strata) ........................................ 15
   Program 7: Estimation and testing for follow-up data with person-time denominators (unlimited number of strata) ....................... 16

4 MATCHED CASE-CONTROL STUDIES .................................................................................... 19-24
   Program 8: Case-control matched analysis ......................................................................................................................... 20
   Program 9: Case-control matched analysis with variable matching ratio .................................................................................. 21
   Program 10: Case-control matched pair analysis for 3 exposure levels ................................................................................... 22

5 EXACT TESTING AND ESTIMATION ...................................................................................... 25-32
   Program 11: Exact testing and interval estimation for the odds ratio in a 2 × 2 table ................................................................ 26
   Program 12: Exact testing and interval estimation for the incidence density ratio in a single table .................................................. 28
   Program 13: Exact p-value for a Poisson variate ...................................................................................................................... 29
   Program 14: Exact confidence limits for an SMR .................................................................................................................... 30
   Program 15: Exact (and approximate) confidence limits for a proportion .................................................................................. 31

6 ADDITIONAL ANALYSES AND CALCULATIONS .................................................................. 33-41
   Program 16: Study size and power ............................................................................................................................... 33
   Program 17: P-value from normal deviate ......................................................................................................................... 34
   Program 18: P-value from chi-square ............................................................................................................................. 35
   Program 19: Linear trend analysis ............................................................................................................................... 36
   Program 20: Seasonal analysis of monthly frequencies ..................................................................................................... 37
   Program 21: Logistic curve fit ........................................................................................................................................... 38
   Program 22: Survival and relative survival curves ........................................................................................................... 39

REFERENCES ...................................................................................................................................... 43-44

APPENDIX PROGRAM LISTINGS ................................................................................................. 45-142
PREFACE

Our purpose in writing this book is to present a collection of pocket calculator programs developed to handle the range of analyses that most epidemiologists face routinely. The programs are accompanied by step-by-step descriptions of their use and examples taken from epidemiologic studies. We have assumed a basic familiarity with epidemiologic principles, this book being intended as a handbook or reference work for epidemiologists rather than as a textbook. Most of the examples are drawn from the area of chronic disease epidemiology and are meant to exemplify the application of calculator programs and not epidemiologic findings.

Before the advent of moderately priced programmable pocket calculators, complicated numerical calculations could be accomplished only by relying upon a sizeable computer or upon a willingness to plow through tedious computations, recording intermediate results along the way. The disadvantages of the large computer are inaccessibility and expense. The pencil and paper approach is often frustrating and unreliable. Fortunately, recent advances in electronic technology have introduced literally into the hands of the general public the capability to solve complex computational problems quickly, accurately and inexpensively.

In the past, the most appropriate analytic procedures for epidemiologic data have often not been applied either because they were too complicated for routine use or not widely known among epidemiologists. Today programmable pocket calculators are available which are essentially low-priced computers that can be programmed to handle nearly all calculations routinely encountered in epidemiologic analysis. Programs can be permanently stored on thin magnetic cards which are kept with the calculator, eliminating the need to renew familiarity with analytical formulas each time a particular set of equations is to be solved. These small, self-contained, battery-operated calculators are transported conveniently from office to home, to conferences, or to classrooms.

The epidemiologic programs developed and presented here are for a particular calculator, the Hewlett-Packard 67 (HP-67). The individual programs, however, could be rewritten for other comparable machines. With improvements in pocket computer technology, one or two programs might profitably be revised, but today's computer technology has already placed sufficient power into a pocket calculator to cope effectively with nearly the full range of epidemiologic analyses. We therefore believe that present-day calculators, purchased to perform the computations described in this book, are not likely to be obsolete in the near future.

In preparing this book we have tried to delineate the conditions for which all programs and program formulas are applicable. Generally the computations performed conform to what is accepted as theoretically most desirable by methodologists, although our personal biases have undoubtedly influenced our choice of certain epidemiologic methods. We have attempted to avoid controversial or esoteric methods, but it is inevitable that objections will be raised over our choice in the selection of some computational techniques. Many methods have been described, for example, for large-sample interval estimation, and there is little unanimity among statisticians as to which is preferable. Disagreement about the desirability of adjusting statistical tests of discrete data for noncontinuity has persuaded us to follow the dictates of parsimony and omit such an adjustment. In choosing among these and other techniques, we were guided mainly by theoretical considerations but partly by technical issues such a programming space, storage requirements and computational time.

Kenneth J. Rothman
John D. Boice, Jr.

June 1978
Chapter 1
INTRODUCTION

1-1 Choice of Analytic Methods

The analysis of epidemiologic studies generally has two distinct phases: first, in the data-processing or data-reduction phase, the information on individual subjects is summarized into a substantially reduced format, such as a set of 2×2 contingency tables or perhaps only a single 2×2 table; in the second phase, that of statistical analysis, the reduced data are further summarized into estimates of the epidemiologic measures of interest. The data-processing phase usually warrants the use of electronic data-processing equipment, though mechanical sorters or even hand sorting can often accomplish the task efficiently for moderately sized series. The statistical analysis can range in complexity from a simple pencil and paper computation to a complex iterative solution of a set of high order equations, depending on the data layout, the parameters being estimated, and the choice of estimation technique.

From the epidemiologist’s point of view, methods for the statistical analysis of epidemiologic data fall into two major classes: methods which are theoretically preferable but computationally difficult, and methods which are theoretically acceptable, though not ideal, but computationally simple. Occasionally a statistic which has desirable characteristics is also easy to obtain, but commonly the ideal statistic involves iterative solution of a complicated set of equations. Thus there seems to be an inverse relation between theoretical preferability of formulas and their computational ease. Whatever the choice of analytic methods, the investigator has always faced the necessity for having the formulas committed to memory or filed conveniently so that the appropriate calculations could be performed when desired. Because the more complicated techniques are often difficult to remember, such methods tend to be applied only as the final analysis for large studies in which much effort has been invested, if at all.

Soon all this will change. Already, for several hundred dollars, the epidemiologist can purchase a programmable handheld calculator which can perform complicated analyses. Instructions for the analyses can be coded into programs stored on magnetic strips that can be easily carried with the calculator. These machines have the capacity to perform quickly and accurately the difficult analyses which previously took either considerable time or required access to a large-scale computer. Although it is certain that calculator technology will continue to improve rapidly, currently available machines already have the capability of revolutionizing the analysis of epidemiologic data by providing fast, accurate and theoretically preferable results.

1-2 Elements of Epidemiologic Analysis

The two basic types of epidemiologic study are the follow-up or cohort study, and the case-control study. In follow-up studies, groups which are identified with regard to the presence or absence of some exposure or characteristic of interest are followed through time to assess and compare the disease rates in each group. In case-control studies, people with disease are compared with those without disease for the relative frequency of exposure history. In addition to follow-up and case-control studies, epidemiologists occasionally conduct cross-sectional studies, in which disease and exposure status are recorded simultaneously. Cross-sectional studies can be treated as either case-control or follow-up studies, the choice depending on the objectives. More detailed descriptions of particular study types can be found elsewhere (Miettinen, 1976b).

The type of study determines the type of epidemiologic analysis. Follow-up studies (chapter 3) permit the direct estimation of incidence rates for exposed and nonexposed subjects. The incidence rates may be expressed simply as the proportion of the initial group becoming ill after a fixed period of follow-up; this proportion is termed cumulative incidence. Ideally, incidence is expressed as the ratio of cases of illness to the amount of person-time experience, and referred to as the incidence density, force of morbidity, or hazard function. The choice between the type of incidence measure depends on the format of the data: when the denominators are simply the counts of the number of subjects, the rate is cumulative incidence; when the denominators are the sums of the time at risk for each subject, the rate is incidence density. Though incidence density is theoretically preferable for analytic work, sometimes cumulative incidence data are preferred for their simplicity and easy interpretability. The effect of the exposure is assessed by a comparison of the incidence rates, either cumulative incidence or incidence density, for exposed and nonexposed groups. The difference and the ratio of the compared rates may both be of interest. Case-control studies (chapter 2) do not usually permit the direct estimation of incidence rates, but the ratio of incidence rates for exposed and unexposed can be estimated from the exposure odds ratio (Miettinen, 1976a).

The major concern in the analysis of epidemiologic studies is usually control of confounding (Miettinen, 1974a; Rothman, 1977). In simple terms, confounding can be defined as the mixing of effects of two or more factors, one of which is the exposure of interest. The control of confounding amounts to the separating of the study effect from the effect of the confounding factors. Confounding can be controlled in the analysis by one of two strategies: stratification by category of the confounding factor(s), or multivariate modeling. With stratification, the comparison of exposed with nonexposed subjects or of cases with controls occurs within narrow categories of the confounding factor(s). Within these strata, the confounding factor is held constant and consequently no confounding occurs. An overall measure of the effect of exposure can be obtained by combining the stratum-specific results. Stratification is generally preferred to multivariate analysis because it permits the investigator to retain a familiarity with the data and, for related reasons, usually offers more clarity of interpretation for readers. Multivariate modeling, on the other hand, reduces the “feel” for the data, involves a set of mathematical assumptions which in practice are rarely met, and leads to results which are often difficult to translate into meaningful epidemiologic measures. The main motivation for multivariate analysis...
sis has been the need-to-control many factors simultaneously, a situation in which stratification may not be feasible. In such situations, a good solution is a combined approach recently proposed by Miettinen (1976b), which exploits the advantages of both strategies.

When confounding is controlled by stratification, an overall estimate of the effect is obtained by taking a weighted average of the stratum-specific effect estimates. There are two approaches to the combination of stratum-specific effect estimates: pooling and standardization. With pooling, the weights used to get a weighted average are chosen, explicitly or implicitly, in proportion to the precision of each estimate, thereby obtaining the most stable overall estimate. An underlying assumption for pooling is that the effect is uniform over all strata and that differences among stratum-specific estimates are attributable to sampling error. This assumption may be assessed by appropriate tests of heterogeneity of effect. The maximum-likelihood estimator of a uniform effect, though not employing explicit weights and, in fact, calling for an iterative calculation, is widely accepted to be the most desirable estimator. Other pooled estimators, however, use explicit weights which are inversely proportional to the variance of the stratum-specific estimate, are often equally acceptable. With standardization, the weights used to combine the stratum-specific effects are derived from a standard distribution (for a ratio measure of effect, the weights involve both the standard distribution and the category-specific rates among nonexposed) (Miettinen, 1972). Standardization is employed to get an overall estimate of effect when the assumption of homogeneity is not tenable, or if the overall estimate is to be compared with another overall estimate, as, for example, in the evaluation of dose-response effects.

Hypothesis testing is usually a part of the formal analysis of the data, sometimes preempting estimation as the main focus in the analysis. We view hypothesis testing as an issue of secondary interest, less important than estimation because hypothesis testing reveals nothing about the magnitude of effect (Rothman, 1978b). Hypothesis testing is almost always accomplished by calculating a p-value from the data according to some statistical model which corresponds to the null or no-effect situation. The larger the p-value, the greater the compatibility between the data and the hypothesis of no-effect. A small p-value indicates that chance is a remote explanation for the discrepancy between the actual data and what would be expected if the null situation were the true state of nature. A small p-value, however, can never rule out chance as an explanation, and even when the p-value is small, chance may be the best explanation of the findings.

With sparse data, p-values should be calculated exactly from the appropriate probability model. With a large number of observations, simpler asymptotic procedures suffice. For most epidemiologic data, the Mantel-Haenszel test or one of its extensions or modifications is an appropriate, if not the optimal asymptotic test. The best guide to deciding between the use of exact and asymptotic tests is experience; whenever the p-values obtained exactly and asymptotically fail to agree, it is preferable to use the exact p-value. Fortunately, the programs in this book simplify the calculation of exact p-values for many situations so that, when there is a serious question about the applicability of an asymptotic test, an exact p-value can often be obtained. (The main exception is with stratified data, which calls for extremely complicated and lengthy computations for exact p-values).

Exact p-values usually are calculated by taking the probability of the observed outcome, under an assumed probability model, and adding to it the probability of all outcomes more extreme than that observed. Recently Miettinen has suggested an alternative method of computation which takes one-half of the probability of the observed data and adds to that the probability of all more extreme outcomes (Miettinen, 1974d). The p-value computed in this way has several theoretical advantages over the usual type of p-value. Programs in chapter 5 deal with exact p-values and generally give the user an option for either type of p-value.

Confidence limits for an overall measure of effect can be derived in a variety of ways. We have employed the test-based approach proposed by Miettinen (1976a). This technique combines the point estimate with the chi resulting from hypothesis testing to obtain the limits. This simple approach appears to be reasonably accurate, and agrees well with more complicated procedures. Limits in the vicinity of the null value of the parameter are highly accurate with this approach, as long as the asymptotic test is applicable.

Most investigators are aware that special analytic procedures are called for whenever individual matching is employed in subject selection. These special "matched analyses" are essential for case-control studies; for which a failure to employ the proper matched analysis can introduce a substantial bias into the results. In essence, matched analyses are stratified analyses with each matched set being one stratum. Testing and estimation are accomplished using the same tests and estimation procedures ordinarily applicable to stratified data, but adapted to the different data layout using matched sets. Recently, a technique was proposed to compute consistent rate ratio estimates for several exposure levels with matched case-control pairs (see chapter 4).

For follow-up studies using general population rates as a comparison with the observed rates in the cohort, the analysis usually involves comparing the observed number of events in the cohort to the number expected if the population rates had prevailed for the study group. The ratio of this observed and expected frequency is the SMR, or standardized mortality ratio. Hypothesis testing and interval estimation for SMR estimates can be performed by assuming that the observed frequency follows a Poisson distribution, and that the expected frequency is so stable that it has a variance of zero. Exact p-values and confidence limits for SMR estimates from follow-up studies are obtainable with programs described in chapter 5.

Frequently the focus of an analysis is on the evaluation of a dose-response relationship. With unstratified data, hypothesis testing of a dose-response relationship usually amounts to an evaluation of linear trend in a 2 × K table, with scores assigned to the K levels of exposure in a way that reflects the prior judgment about the relationship of exposure to disease. The dose-response relationship can be quantified by comparing effect estimates at each level of exposure or by computing the slope of the linear regression of exposure to disease. With stratified data, the same procedures can be employed, the appropriate procedures for testing and slope estimation described by Mantel (1963) in his paper discussing extensions of the Mantel-Haenszel procedure. Programs 3 and 19 deal with the evaluation of trend.
Several additional computational problems are addressed in chapter 6. One problem often confronting epidemiologists is the calculation of size or power for a contemplated study. Power is the probability that a specific study design will lead to a p-value less than some arbitrary value. Power curves can be of help in determining the appropriate size for a study. Another frequent computation well suited to programmable calculators is the computation of a survival or relative-survival curve from life-table data. Analyses of clinical trials, animal studies, or even the retrospective clinical experience of a number of patients on a given treatment regime can be quickly and effectively analyzed with the description of the survival curves. Other programs in chapter 6 include the analysis of seasonal trends, logistic curve-fitting, and utility programs to convert a normal deviate or a chi-square to a p-value.

1-3 Comprehensive Example

To exemplify the applicability of a few of the programs in this book, consider a case-control study to evaluate the association between smoking history and lung cancer risk. To plan the size of such a study, the investigator would use program 16 to compute some power curves. If an overall ratio of one control to one case were planned, with an arbitrary one-sided alpha level of 0.05, and the proportion of smokers in the population is assumed to be 50 percent, the power curves for rate ratios of 2, 5, and 10 can be quickly determined to be:

<table>
<thead>
<tr>
<th>Total Size</th>
<th>RR = 2</th>
<th>RR = 5</th>
<th>RR = 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>0.22</td>
<td>0.60</td>
<td>0.85</td>
</tr>
<tr>
<td>50</td>
<td>0.31</td>
<td>0.83</td>
<td>0.98</td>
</tr>
<tr>
<td>100</td>
<td>0.51</td>
<td>0.99</td>
<td>1.00</td>
</tr>
<tr>
<td>200</td>
<td>0.77</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>500</td>
<td>0.99</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Smoking History</th>
<th>Yes</th>
<th>No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases</td>
<td>668</td>
<td>21</td>
<td>709</td>
</tr>
<tr>
<td>Controls</td>
<td>650</td>
<td>59</td>
<td>709</td>
</tr>
</tbody>
</table>

In one of the first studies concerning the association between smoking and lung cancer, Doll and Hill (1954) interviewed 709 hospitalized patients with lung cancer and an equal number of hospitalized controls. The crude results are given in table 1.1. Program 1 or 2 could be used to analyze the data in this single 2x2 table. The results, using either program, are

Mantel-Haenszel chi-square p-value 4.4 (hypothesis test)
RR estimate 6.2x10^-6 (from Program 17)
95% Confidence Limits 1.8, 4.8

Table 1.2

<table>
<thead>
<tr>
<th>Sex</th>
<th>Male Smoking History</th>
<th>Female Smoking History</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases</td>
<td>Yes 647 No 2</td>
<td>Yes 41 No 19</td>
</tr>
<tr>
<td>Controls</td>
<td>622 27</td>
<td>23 32</td>
</tr>
</tbody>
</table>

The crude data, however, were confounded by sex. To assess and control sex confounding, the data in table 1.1 should be stratified by sex (table 1.2). These stratified data can then be entered into program 1 with the following results:

<table>
<thead>
<tr>
<th>Test Statistic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mantel-Haenszel chi-square</td>
<td>5.0 (hypothesis test)</td>
</tr>
<tr>
<td>Heterogeneity test chi-square</td>
<td>3.0x10^-7 (from Program 17)</td>
</tr>
<tr>
<td>P-value</td>
<td>5.7 (1 d.f.)</td>
</tr>
<tr>
<td>SMR</td>
<td>11.0</td>
</tr>
<tr>
<td>R (Mantel-Haenszel estimate)</td>
<td>4.5</td>
</tr>
<tr>
<td>R (maximum likelihood estimate)</td>
<td>4.3</td>
</tr>
<tr>
<td>95% confidence limits for maximum likelihood estimate</td>
<td>2.4, 7.5</td>
</tr>
</tbody>
</table>

The sex-specific estimates of the rate ratio, 14.0 for males and 2.5 for females, are considerably different. The heterogeneity test provides further evidence that it would be inappropriate to assume that the rate ratio were identical for the two sexes. Therefore neither the maximum likelihood nor Mantel-Haenszel estimates are appropriate. If a summary measure of the rate ratio is desired, the SMR would still be appropriate.

Table 1.3

<table>
<thead>
<tr>
<th>Smoking History</th>
<th>Male No. Cigs. Smoked Daily</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases</td>
<td>No 15+</td>
</tr>
<tr>
<td>Cases</td>
<td>0 15+</td>
</tr>
<tr>
<td>Controls</td>
<td>2 5</td>
</tr>
<tr>
<td>Controls</td>
<td>27 250 299 274</td>
</tr>
<tr>
<td>Total</td>
<td>28 66 543 638</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Smoking History</th>
<th>Female No. Cigs. Smoked Daily</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases</td>
<td>0 5+</td>
</tr>
<tr>
<td>Cases</td>
<td>19 5+</td>
</tr>
<tr>
<td>Controls</td>
<td>32 10</td>
</tr>
<tr>
<td>Controls</td>
<td>32 19 19 15</td>
</tr>
<tr>
<td>Total</td>
<td>51 29 21</td>
</tr>
</tbody>
</table>

* Standardized Risk Ratio (Miettinen, 1972) with the standard taken as the minimally exposed patients smoking less than 5 cigarettes per day.

* Fleiss would argue that SMR is unwise if stratum-specific estimates of the association are wildly different (e.g., opposite sign).
between smoking and lung cancer risk, the data are entered into program 3. The smoking levels should be assigned scores, here taken to be 0, 2, 10, and 25. The results from program 3 are

<table>
<thead>
<tr>
<th>Test</th>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mantel-extension chi-P-value</td>
<td>6.4 (1 d.f.)</td>
<td>7×10⁻¹¹ (from Program 17)</td>
</tr>
<tr>
<td>Chi-square for heterogeneity P-value</td>
<td>1.2 (1 d.f.)</td>
<td>0.3 (from Program 18)</td>
</tr>
<tr>
<td>SRR estimates</td>
<td></td>
<td>(listed in table 1.3)</td>
</tr>
</tbody>
</table>

The results strongly support the hypothesis that smoking is associated with lung cancer and that the risk of lung cancer appears to be related to the quantity of tobacco smoked. The different sex-specific effects seen in table 1.2 appear to have resulted from the lighter smoking habits of women in this study.

These detailed analyses can be performed in less than 10 minutes using the calculator programs to be described in subsequent chapters.
Chapter 2
CASE-CONTROL STUDIES

2.1 Introduction

In case-control studies, groups with different disease and health characteristics are identified and prior exposure histories determined. Exposure rates or exposure distributions for both case and control series are then estimated and compared. For the usual type of case-control study, the ratio of exposure rate odds, i.e., the common odds ratio, serves as an estimate of the incidence density ratio ("relative risk"). Incidence density, the fundamental measure of disease occurrence, is defined as the ratio of new cases of illness to the relevant person-time experience (Miettinen, 1976a). The incidence density ratio, which is the incidence density for a group exposed to some agent or possessing some characteristic of interest divided by the incidence density for some referent group, is a measure of the effect of a particular exposure. If, however, a case-control study is conducted after the period of disease risk has passed, as with case-control studies of congenital defects, the exposure rate odds ratio approximates, assuming low disease rates, either the cumulative incidence ratio or the prevalence ratio. Cumulative incidence is defined as the proportion of individuals who develop the illness in question during a specified period of time, and prevalence is the proportion of a population with illness. Becoming a case v. being a case.

Data from case-control studies can be arrayed into a $2 \times 2$ table or a $2 \times R$ table with $R$ categories of exposure (table 2.1). For the $2 \times 2$ table, the exposure rate odds ratio is specified as $(a/c)/(b/d) = ad/bc$. With several levels of exposure, estimates of effect are obtained for each level of exposure by contrasting each level separately with the reference level. The point estimate of the exposure effect at each level is the exposure rate odds ratio, and the set of effect estimates for all levels are the ratios $a_1/d_1,$ $a_2/d_2,$ etc.

<table>
<thead>
<tr>
<th>Table 2.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Notation and Data Arrangement for Case-Control Studies</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cases</th>
<th>Exposed</th>
<th>Unexposed</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Controls</td>
<td>$a_i$</td>
<td>$c_i$</td>
<td>$M_i$</td>
</tr>
<tr>
<td>Total</td>
<td>$N_i$</td>
<td>$N_e$</td>
<td>$T_i$</td>
</tr>
</tbody>
</table>

Exposure level

<table>
<thead>
<tr>
<th>Cases</th>
<th>$a_i$=c=0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4 ..</th>
</tr>
</thead>
<tbody>
<tr>
<td>Controls</td>
<td>$b_i$=d=0</td>
<td>$b_1$</td>
<td>$b_2$</td>
<td>$b_3$</td>
<td>$b_4$ ..</td>
</tr>
<tr>
<td>Total</td>
<td>$N_i$</td>
<td>$N_i$</td>
<td>$N_e$</td>
<td>$N_0$</td>
<td>$N_e$ ..</td>
</tr>
</tbody>
</table>

Statistical hypothesis testing for the $2 \times 2$ table, evaluating the consistency of the data with the null hypotheses that the exposure has no effect, is generally accomplished either by an exact test (see chapter 5) or by an asymptotic (large sample) test based on the hypergeometric distribution. Interval estimation, i.e., the computation of confidence limits, can be performed by combining the point estimate of the odds ratio with the results from the hypothesis test (Miettinen, 1976a).

For the $2 \times R$ table, hypothesis testing of a monotonic trend in effect is usually based on a test for a linear trend in proportions (Armitage, 1955; Mantel 1963). Interval estimation, however, is problematic; ideally it would first involve fitting a mathematical model that relates the effect to the exposure level, but such modeling commonly necessitates stable point estimates or at least some insight into the causal relationship under study. Interval estimation for the effect at each level of exposure could be accomplished but would have the drawback of failing to account for the trend information available from the remainder of the data.

Typically one or more confounding factors must be controlled in the analysis. The classical technique to control confounding relies upon making elementary comparisons within narrow ranges or categories of the confounding factor(s). Thus, for a dichotomous exposure, there will be a $2 \times 2$ table for each category of the confounding factor(s) and for exposures classified into $R$ categories, a $2 \times R$ table for each category.

2.2 $2 \times 2$ Table Analysis

When data are stratified in sets of $2 \times 2$ tables to control confounding, the asymptotic maximum likelihood (ML) estimate of a uniform odds ratio ($R$) over all strata can be computed as the solution to the equations:

$$\Sigma a_i - \Sigma E(a_i|R) = 0$$

(Gart, 1971)

$$R = \frac{E(a_i|R) \{M_i - N_i + E(a_i|R)\}}{\{M_i - E(a_i|R)\} \{N_i - E(a_i|R)\}}$$

(2)

where $E(a_i|R)$ is the expected value of $a_i$ conditional upon the odds ratio, $R$.

For computational purposes, equation [2] can be rewritten as:

$$E(a_i|R) = \text{ABS} \left\{ \text{ABS} \left( \frac{1}{2} \left( \frac{T_i}{R-1} + N_i + M_i \right) \right) - \sqrt{\left( \frac{1}{2} \left( \frac{T_i}{R-1} + N_i + M_i \right) \right)^2 - \frac{MN_iR}{R-1}} \right\}$$

(3)

Equations [1] and [3] can be solved iteratively for $R$ by successive approximations. Information from the marginal totals of each $2 \times 2$ table must be used repeatedly in the calculation and must therefore be stored in the computer memory. For the Hewlett-Packard 67 (HP-67) calculator, the memory capacity limits the capability of the machine to a maximum of seven $2 \times 2$ tables when ML estimates are desired.

A useful approximation to the ML estimate that has received widespread application is one proposed by Mantel and Haenszel (M-H) (1959), namely

$$\hat{R}_{M-H} = \frac{\Sigma (a_i/c_i/T_i)}{\Sigma (b_i/c_i/T_i)}$$

(4)
This estimator usually agrees closely with the maximum likelihood estimator. Since the M-H estimate can be calculated directly without iteration, it provides a useful starting point for the ML iterative calculations. In addition, the M-H estimator also serves as a reasonable substitute for the maximum likelihood estimator for sets of 2 x 2 tables which exceed the storage capacity of a calculator. Separate programs (programs 1 and 2) have been developed to yield either ML estimates in the case of 7 or fewer 2 x 2 tables or M-H estimates for unlimited numbers of tables.

The Mantel-Haenszel and maximum likelihood estimators discussed above are both estimators of a uniform odds ratio. Under certain circumstances, such as heterogeneity of the odds ratio across strata or as a preliminary step to calculating the etiologic fraction (population attributable risk) (Miettinen, 1974b), a standardized estimate of the exposure odds ratio may be preferable. Taking the exposed group as the standard and assuming that the disease is rare and that the cases were identified without regard to the stratification variate(s), the standardized rate ratio (Miettinen, 1972) is obtainable as:

\[ \widehat{SMR} = \frac{\sum a_i / \sum \{b_i c_i / d_i \}}. \]  

Alternately, if the unexposed group is taken as the standard, the standardized rate ratio (Miettinen, 1972) is calculable as:

\[ \widehat{SRR} = \left( \frac{\sum a_i d_i / b_i}{\sum c_i} \right). \]  

Statistical hypothesis testing for a set of 2 x 2 tables can be accomplished by applying a generalization of the test for a single 2 x 2 table. This generalized test, presented by Mantel and Haenszel (1959), is optimal when the odds ratio is uniform over the strata. The test is still useful, however, in situations in which the odds ratio varies across strata. When the odds ratio varies, the test evaluates the "average" departure from the null condition. The test statistic \( \chi_{MH} \), is computed as:

\[ \chi_{MH} = \frac{\Sigma a_i - \Sigma M_i N_i / T_i}{\sqrt{\Sigma M_i M_{ni} N_i N_{ni} / T_i (T_i - 1)}} \]  

and the probability associated with the test statistic can be obtained from tables of the standard normal distribution (see also program 17). Although the Mantel-Haenszel test is a large-sample test, large-sample conditions are not required for each 2 x 2 table, but only for the summations used to calculate the test statistic.

A general large-sample test of the hypothesis that the odds ratio is uniform across strata can be constructed by summing the one degree-of-freedom chi-square statistics computed from each 2 x 2 table and subtracting the square of \( \chi_{MH} \) (Zelen, 1971). This heterogeneity test, however, seems to apply only for near-null conditions and can be misleading in certain non-null circumstances (Miettinen et al., 1977). A more general test is an asymptotic likelihood ratio test (Miettinen, 1975) taken as

\[ \chi^2 = -2 \left[ \sum a_i \ln \left( \frac{\hat{a}_i}{\bar{a}_i} \right) + b_i \ln \left( \frac{\hat{b}_i}{\bar{b}_i} \right) + c_i \ln \left( \frac{\hat{c}_i}{\bar{c}_i} \right) + d_i \ln \left( \frac{\hat{d}_i}{\bar{d}_i} \right) \right] \]  

with \( n-1 \) degrees of freedom, in which \( \hat{a}_i, \hat{b}_i, \hat{c}_i \) and \( \hat{d}_i \) are the fitted cell entries in table i satisfying equations [1] and [2] and consistent with the observed marginal totals for each table.

Interval estimation for the odds ratio can be accomplished conveniently using the test-based interval estimation procedure proposed by Miettinen (1976a). This procedure couples the Mantel-Haenszel or the maximum likelihood point estimate with the \( \chi_{MH} \) test statistic. This technique for setting confidence limits uses the standard error of the odds ratio computed at the null value, which is implicit in the test statistic, and, therefore, the limits are more accurate if they are in the vicinity of the null value (Miettinen, 1977). The advantages of this method are extreme simplicity, agreement with more complicated procedures of interval estimation, and a high degree of accuracy near the null value, which is usually the region of greatest interest. (The technique fails, however, in the unusual situation when \( \chi_{MH} \)=0 and \( R=1 \).)

Test-based lower (\( R \)) and upper (\( \bar{R} \)) confidence limits are computed as

\[ R = \hat{R} \left(1 - Z_{1-\alpha} \right) \]  

and

\[ \bar{R} = \hat{R} \left(1 + Z_{1-\alpha} \right) \]  

where \( \hat{R} \) is a pooled point estimate and \( Z \) is the value of a standard normal deviate corresponding to the desired level of confidence (for 90 percent limits \( Z=1.645 \) and for 95 percent limits \( Z=1.960 \)).

**Program 1**

**Description and Operating Instructions**

**Title:** ODDS RATIO ESTIMATION AND TESTING (7 OR FEWER 2x2 TABLES)

**Summary:** For 7 or fewer 2x2 tables, this program computes the Mantel-Haenszel chi, the heterogeneity chi-square, the SMR, the Mantel-Haenszel pooled point estimate, the maximum likelihood pooled point estimate, and test-based confidence limits around the maximum likelihood estimate.

**Cautions:**
1. No more than seven 2x2 tables are permitted. If eight or more tables are to be analyzed, use program 2.
2. If \( d_i=0 \) for any table, the \( \widehat{SMR} \) is not calculable, and the \( \widehat{SMR} \) value displayed by the calculator should be ignored (the program divides by one, rather than zero, to avoid an error message which would halt execution).
3. No table with a zero marginal frequency should be entered. Such tables contribute no information to either testing or estimation and will halt the program.
4. If the Mantel-Haenszel chi equals zero, test-based confidence limits are not obtainable. For this infrequent situation, another interval estimation procedure must be used (see Gart (1970)).
5. If any cell entry equals zero, the asymptotic likelihood ratio test does not apply and steps 8-11 should be ignored.