1987

Exclusion Bias

Todd Q. Miller
Loyola University Chicago

Exclusion Bias

by

Todd Q Miller

A Thesis Submitted to the Faculty of the Graduate School of Loyola University of Chicago in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

April

1987

ACKNOWLEDGMENTS

I wish to express my sincere appreciation to the members of my thesis committee. My thanks to R. Scott Tindale Ph.D. for his suggestions and direction aided greatly in increasing the clarity of many of the ideas presented in this thesis. I also gratefully acknowledge Emil J. Posavac Ph.D. for his many helpful comments and encouragement.

I wish to express my thanks to Charles W. Turner Ph.D. who contributed greatly to most of the ideas and concepts presented in this thesis. His enthusiasm, friendship, and advice have been invaluable.

Finally, I would like to thank my friends and family for their encouragement and support.

FORWARD

Using unrepresentative sampling techniques to study unusual, or rare populations creates statistical problems that reduces the likelihood of obtaining statistically significant results. This paper refers to these problems as exclusion bias. The term "exclusion" is used to connote that potential observations from the population are filtered or selected into or excluded from the study sample.

This paper presents a description of how exclusion bias is related to statistical power. Decision Theory is used to develop a statistical model of exclusion bias and several computer simulations are presented that demonstrate how exclusion bias reduces statistical power. These simulations can be used by researchers to determine the statistical power of their studies just as Cohen's (1977) power tables are used to assess the sample size required to obtain a certain degree of statistical power. Ways in which exclusion bias may have produced misleading study findings for research on the relationship between Type A behavior and arteriosclerosis, and techniques for assessing and controlling for exclusion bias, are discussed.

VITA

The author, Todd Q Miller, is the son of Ralph Howard Miller and Janet Arlene (Mitchell) Miller. He was born June 25, 1957, in Salt Lake City, Utah. His elementary education was obtained in the public schools of Salt Lake City, Utah. His secondary education was completed in 1975 at Skyline High School, Salt Lake City, Utah. In September, 1975, Mr. Miller entered the University of Utah, receiving the degree of Bachelor of Science in psychology in June, 1984. In September, 1984, Mr. Miller enrolled in the applied social psychology program at Loyola University of Chicago and completed his Master of Arts in 1987.

TABLE OF CONTENTS

LIST OF FIGURES

INTRODUCTION

Using unrepresentative sampling techniques to study unusual, or rare populations creates statistical problems that reduces the likelihood of obtaining statistically significant results. This paper refers to these problems as exclusion bias. Exclusion bias is present whenever the probability of selecting certain observations for study is associated with the predictor and/or the criterion. The term "exclusion" is used to connote that potential observations from the population are filtered or selected into or excluded from the study sample. The purpose of this paper is to present a detailed quantitative description of the influence exclusion bias can have on statistical power. This paper attempts to demonstrate that exclusion bias is an important and often overlooked problem in many areas of research.

Exclusion bias is often present in applied research where the purpose is to find predictors of differences between normal and unusual, rare, or abnormal individuals. For example, personnel psychologists have constructed psychological tests to identify the most qualified individuals for a particular job from an applicant pool. For research on depression, the focus has been on finding differences (e.g., cognitions and/or biochemical abnormalities) between depressives and non-depressives. Similarly, health psychology researchers have attempted to find predictors (e.g., research on stress, Type A behavior, and hardiness) of disease.

In many cases, these applied researchers have used convenience or purposive sampling techniques because of practical and/or ethical problems associated with using representative sampling techniques. For example, to conduct studies using representative sampling techniques to determine the degree of statistical association between job performance and personnel selection tests is very difficult. In most cases, hiring all available job applicants in the general population would be too costly. Similarly, college students have commonly been used for research on the etiology of depression because of the inconvenience associated with obtaining large samples of individuals that are actually depressed.

A common result of studies that use unrepresentative sampling techniques is that extreme scores on the variables of interest are over-represented in the study sample. In the case where depressed individuals are sampled from outpatient clinics, most individuals would receive high scores on a measure of depression and few low scores would be present in the study sample. Conversely, most depression scores would be in the low range if a college student population was examined. Similarly, only a selected few individuals with exceptional qualifications are hired for most jobs. Therefore, studies on the predictors of job performance only use individuals with high job performance test scores.

This paper addresses a number of issues concerned with obtaining statistical significance from studies mainly sample individuals with extreme scores on the study variables. Some researchers have referred to the problem of selecting study samples from a narrow range of test values as a problem of restriction of range (viz. Pearson, 1903). One section of this paper discusses previous work on range restriction and its implications for exclusion bias. However, this paper demonstrates that sampling from extreme ranges of values has different consequences than if the researcher selects observations from the middle of the population. Therefore, statistical parameters in addition to a reduction in variance determine whether a study finds statistical significance. Thus, the term exclusion bias as opposed to restriction of range is used so as not to suggest that the only effect of sampling from extreme ranges is a reduction in variance.

Cohen (1977) has illustrated how sample size influences statistical power. This paper uses an approach similar to Cohen's (1977) to illustrate how other statistical parameters influenced by exclusion bias can reduce statistical power. One section of this paper defines exclusion bias in terms of Decision Theory (DT). Several computer simulations are used to illustrate the extent to which various statistical parameters influenced by exclusion bias can reduce statistical associations between study variables. These simulations can be used by researchers to determine the degree to which their study's statistical

power has been influenced by exclusion bias just as Cohen's work is used to assess the sample size needed for a given degree of statistical power.

In addition, this paper discusses ways that exclusion bias can confuse or obfuscate research findings and attempts to illustrate why exclusion bias is an important and often underated problem in many areas of research. Another section of this paper discusses methods of controlling for problems of exclusion bias. Finally, a summary section discusses the implications of controlling exclusion bias for future research endeavors.

TYPE A BEHAVIOR AND ARTERIOSCLEROSIS

To ease exposition, the statistical definitions and computer simulations of bias are presented in the context of a research problem. The research problem chosen for this paper is determining whether Type A behavior is related to coronary arteriosclerosis. This section of this paper is an introduction to relevant theory and research on the relationship between Type A behavior and arteriosclerosis.

The idea that Type A behavior might be predictive of heart disease has been given some notice because traditional risk factors predict only about half of the new cases of coronary heart disease each year (Jenkins, 1976). Type A behavior (Friedman & Rosenman, 1974) has been defined as "any person who is aggressively involved in a chronic, incessant struggle to achieve more and more in less and less time, and if required to do so against the opposing efforts of other things and persons." Type A behavior is considered to have three core components: (a) hostility/aggressiveness, (b) a sense of time urgency, and (c) competitive/achievement striving (Glass, 1977). Type A behavior has been described as a set of behaviors elicited by a challenging or threatening environment (Matthews, 1982). In addition, Type A's exhibit behaviors that would appear as typical reactions to continuous stress whether stressers are present in the environment or not. Type A's may actually seek out challenging and threatening environments that produce stress (Smith & Anderson, 1986).

One theory of how Type A behavior induces heart disease suggests that Type A's exhibit elevated blood pressure in response to challenge. The increased lability in blood pressure results in micro fine tears in the endothelial lining and/or smooth muscle wall of the artery. When tears in the endothelial lining and/or smooth muscle wall heal, atheromatous plaque remains in the walls of the arteries. Presumably, repeated vessel injury leads to a build up of plaque. This build up of plaque is referred to as arteriosclerosis. Presumably, when severe arteriosclerosis leads to complete occlusion of one or more of the coronary arteries a heart attack occurs. Severe occlusions of at least one artery are present in over 90% of all heart attacks (cited in Pearson, 1984, pp. 142). For a more detailed summary of this theory of how Type A behavior produces heart disease, see Williams (1979).

A diagnosis of arteriosclerosis requires validation through a surgical procedure known as a coronary angiography (Conti, 1977). The procedure involves inserting a catheter into an artery in the patient's arm or thigh. Next, the catheter is advanced until reaching the heart where contrast medium is injected into the heart and monitored by fluoroscopy. The presence of fibrous plaque in the heart appears as a narrowing of the diameter of the image of the dye column appearing on the fluoroscope. The actual measurement of the dye column is often quite subjective. Pearson (1983) reports that angiographies agree with actual degree of occlusion determined by autopsy from 61 to 84 percent of the

time. In general, angiographies tend to underestimate the degree of occlusion present.

After correction for traditional risk factors (i.e., age, blood pressure, smoking, and serum cholesterol) Type A's have a risk 1.97 times greater than Type B's (Brand, Rosenman, Sholtz, & Friedman, 1976) for having a first myocardial infarcation (MI) and are five times more likely to experience a second MI. Brand (1977) found evidence that traditional risk factors serve as moderator variables. In other words, the presence of Type A behavior characteristics combined with other risk factors multiplies one's risk for coronary heart disease. For a more thorough review of research on Type A behavior, see Matthews and Haynes (1986).

The most common measures of Type A behavior have been the Jenkins Activity Survey (JAS; Jenkins, Zyzanski, & Rosenman, 1971) and the Structured Interview (SI; Rosenman, 1978). The JAS is a 52-item questionnaire that yields four subscale scores and an overall score. Scoring is based on optimal weights generated from a discriminant function analysis that predicted SI classification from the white collar men that participated in the Western Collaborative Group Study (Rosenman, Friedman, Straus, Wurm, Kositchek, Hahn, & Werthessen, 1964).

The SI classifies individuals into one of five categories: A1, the subject strongly indicates the Type A personality; A2, displays some Type A characteristics; X, displays some Type A and some Type B quali-

ties: B3, displays some Type B characteristics, B4 displays mostly Type B characteristics. The SI was developed in a middle class nondiseased male population by content validity judgments made by Rosenman and Friedman (see Rosenman, 1978).

STATISTICAL MODELS OF THE SELECTION PROCESS

Previous Research on Range Restriction

Research on the problem of range restriction is directly related to the problem of exclusion bias because exclusion bias produces range restriction. That is, when researchers select observations that represent an extreme range of scores on a predictor and/or criterion a restriction in range and thus variance, occurs.

The problem of range restriction was first identified in personnel selection research when employers began to use psychological tests to select employees. Researchers became interested in knowing how well psychological test scores predicted job performance. When researchers began to compare scores from the group of individuals that were hired on the basis of their psychological test scores, they found low correlations between job performance and their test scores.

The first published report of the problem of restriction of range was by Thorndike (1947). For pilot trainees whose psychological test scores indicated they would be successful, the correlation between a composite aptitude test score and a measure of pilot trainee performance was a most unimpressive .18. For a measure of complex motor coordination, the correlation with job preformance was -.03. However, the Thorndike (1947) study was different from previous studies. The test scores were not used to select applicants. Instead, all applicants to

the training program were admitted. The correlation between all applicants and measures of pilot performance was .64 while the correlation between pilot performance and the complex motor skills test was .40. These correlations suggested that the psychological tests were highly predictive of pilot performance.

Thorndike (1949) and many others have since attributed the rather striking differences in correlations between the selected and unselected groups as due to "range restriction." That is, they attributed the reduction in correlation between performance and predictive test score for the selected group to the restriction in variance in test scores. In the Thorndike study, only pilot trainees with test scores in the top 10% were predicted to be successful. Therefore, only a narrow range of test scores were present in the selected group.

Pearson (1903) presented a formula to correct for the range restriction problem. The correction for range restriction when selection is based solely on the predictor variable is

$$R_{xy} = \frac{S_x / s_x \, r_{xy}}{(1 + (S_x^2 / s_x^2 - 1) r_{xy}^2)}$$

where R_{xy} = the estimate of the correlation in the population, S_x = the standard deviation of the predictor variable in the population, s_x = the standard deviation of the predictor variable in the sample, and r_{xy} = the sample correlation. The key parameter in the formula that is difficult to estimate from most studies is S_x. The formula illustrates

that range restriction formulas only correct for a reduction in variance.

Range restriction formulas have been applied to a number of areas of research including evaluating college entrance examinations (Linn & Dunbar, 1982) and determining the monetary impact of valid selection procedures (Schmidt, Hunter, McKenzie, & Muldrow, 1979). Gulliksen (1950) developed formulas for correcting for range restriction on the criterion variable and multiple predictor variables. These correction formulas are endorsed by the American Psychological Association (1980) and are presented without criticism in many standard works on measurement and testing (e.g., Ghiselli, Campbell, & Yedeck, 1981).

However, there is another body of work that criticizes the use of these formulas. These criticisms are concerned with the plausibility of the various assumptions underlying these correction formulas. In particular, this literature suggests that correction formulas will be the least accurate when the study sample includes mostly extreme scores. Unfortunately, this is the situation that occurs most frequently in applied research. The review of the literature below suggests range restriction formulas do not take into account the effects that sampling from extreme ranges has on the sample correlation.

Range restriction formulas are based upon six basic assumptions: (a) linearity, (b) homoscedasticity, (c) symmetry in the shapes of the distributions of scores for the predictor and criterion variables, (d)

that either the predictor or the criterion completely determines how observations are selected into the study sample, (e) the population variance is known, and (f) the variables are continuous. First, the correction formulas assume that the relationship between the predictor and criterion variable is linear and homoscedastic. Greener and Osburn's (1979, 1980) computer simulation studies found that the correction formulas are somewhat robust to violations of the assumption of homoscedasticity, but are sensitive to violations of nonlinearity.

A study by Lee and Foley (1986) demonstrated what Lord and Novick (1968) suggested that because violations of homoscedasticity and linearity are likely to occur in the tails of bivariate test data, range restriction formulas are least appropriate when applied to extreme score groups. Lee and Foley found that samples taken from extreme scores on an armed services vocational battery test did not accurately reflect the population validity coefficient.

Although range restriction formulas do not depend on the variables being normally distributed (Rydberg, 1963), Brewer and Hill's (1969) computer simulation study found that these formulas were sensitive to a lack of symmetry in the distributions of the criterion for different values of a dichotomous predictor variable. Asymmetry occurs in many areas of research. For example, asymmetry is likely to occur when aptitude tests are used that are too difficult for examinees, or when employment tests are designed to optimally discriminate at a point where

most examinees will fail the test (Brewer & Hill, 1969). In particular, asymmetry is likely to be present when scores from an extreme range are selected. Brewer and Hill recommended not using range correction formulas when the difference in skewness for different values of the predictor is greater than one.

Another finding of Brewer and Hill (1969) was that a large part of the range of the population sample must be included in the study sample for the range correction formula to be accurate. For example, the corrected correlation coefficient can vary from .26 to .77 if the study sample includes 37% of the original sample and the correlation in the population is .51. In most cases in applied research, one would suspect that the study sample would represent less than 37% of the sample. The study sample in the aforementioned study by Thorndike (1947) represented only 10% of the total population. Thus, corrected correlation coefficients may be very inaccurate for study samples found in applied research.

Another assumption of restriction of range formulas is that complete truncation occurs at some point on the test. Olson and Becker (1983) have pointed out that in most cases incomplete truncation occurs. With incomplete truncation, observations are present at any point along the range of test values but the probability that an observation is "lost" from the sample is associated with the observation's values on the predictor and/or criterion variable.

For a variety of reasons, incomplete truncation is more likely than complete truncation in most areas of applied research. For example, voluntary quits and promotions in personnel selection research are likely to lead to some incomplete truncation. Moreover, unmeasured variables such as personality, race, or personal finances may be correlated with the selection process and so may produce incomplete truncation. For example, the admission procedures of the health organization, the willingness of the patient to seek medical attention, and the nature of the disease influence who becomes part of a medical research study.

Becker and Olson (1983) have demonstrated that using range restriction formulas in samples that violate correction formulas' assumptions of complete truncation can lead to seriously misleading results. A range formula that assumes incomplete truncation on the predictor and criterion was given by Thorndike (1949, p. 174). Unfortunately, this formula requires knowledge of the variable or variables that completely determine the selection process. Olson and Becker (1983) discuss a more practical procedure for estimating the population parameters. This procedure is discussed in further detail in the section of this paper that discusses ways to control for problems associated with exclusion bias.

Another problem associated with range restriction formulas is that they assume the population that one corrects for in restriction of range formulas is a constant. In personnel selection research, the population

of job applicants can vary with extent of advertising, characteristics of the job preview, changes in numbers and quality of available applicants due to demographic economic factors etc... These factors may affect the population so that the population variance may vary widely for different times and circumstances. Thus, assuming that the population variance is a constant may be unreasonable. In fact, correction formula estimates may reflect nothing more than changes in the variance of available job applicants rather than unbiased estimates of the validity of the tests. The true practical validity of a test is concerned with whether the test can discriminate between individuals that will perform successfully and those that will not. That is, there must be some score or cutoff point on the test that will accurately divide applicants into two groups; those hired and those not hired. The cutoff point on the test should divide the applicants so that the number it indicates should be hired is approximately equal to the number that the company wants to hire. Thus, the practical validity of a test is not related to the variance of test scores in the population but the location of the cutoff score.

A final assumption of range restriction formulas is that the predictor and criterion are continuous variables. The majority of studies concerning range restriction have come from the industrial psychology literature where the correlational approach is dominant. There is a problem with assuming that study variables are continuous. The correla-

tional approach is not very informative in many applied research situations. That is, predictor variables in most applied research are used to make dichotomous as opposed to continuous decisions such as to hire or not to hire. An approach that identifies optimal decision points based on scores on the predictor variable is needed. The value of the dichotomous variable approach is that one can easily see how predictor variables can be used to make decisions (i.e., hire or not hire, is the patient diseased or not diseased). Thus, the approach taken in medical research has been to treat variables as dichotomous. Moreover, the assumptions of constant test validity and homoscedasticity which are often difficult to meet with medical variables are not required.

Perhaps the approach that medical researchers have taken to addressing problems associated with exclusion bias is quite different from the range restriction approach taken by industrial psychologists because medical researchers treat variables as dichotomous. For example, Kleinbaum, Morgenstern, and Kupper (1981) demonstrated the effects that different probabilities of diseased versus nondiseased subjects in the study population can have on the direction of statistical associations found. The advantage of the dichotomous approach is that it becomes very easy to see how different probabilities in extreme score groups affect the results of studies.

A problem with the dichotomous variable approach is that researchers typically assume that the cutpoint chosen is optimal. The first

computer simulation presented in this paper challenges this assumption and demonstrates that choosing a suboptimal cutpoint can severely reduce statistical associations.

This paper uses a DT approach to describe the sample selection process. The DT approach treats the predictor variable as a dichotomous variable so that assumptions of constant test validity and homoscedasticity are not required. The criterion variable is treated as a continuous variable so as to be able to assess problems associated with using a suboptimal cutpoint. Thus, the DT approach can be used to assess effects of different proportions of diseased versus nondiseased populations being included in the sample. In addition, problems associated with suboptimal cutpoints can be assessed. Thus, the DT approach has several advantages over treating all study variables as either correlational or dichotomous.

The approach taken in this paper differs in two other respects from the literature on range restriction. The focus of the range restriction literature has been on obtaining an accurate estimation of the correlation between a predictor and a criterion in the general population. In contrast, this paper is concerned with whether studies of selected populations can find statistically significant results. Thus, where the range restriction literature has focused on correcting the study sample correlation, this paper attempts to show how the degree of statistical association (as measured by the χ^2 statistic) is influenced by exclusion bias.

Finally, the degree of extremity of scores is ignored in range correction formulas. This paper demonstrates that by using DT researchers more accurately estimate the degree to which extremity of the range of scores in the study sample reduces statistical significance.

Decision Theory

Terminology. The purpose of this section is to define explicitly how exclusion bias reduces statistical significance. To this end, a DT model of exclusion bias is presented. An understanding of a DT model of exclusion bias requires a knowledge of the terminology and assumptions of DT. Therefore, the next few paragraphs of this paper are a brief introduction to DT; for a more detailed discussion, see Raiffa (1968).

DT uses several terms to describe the statistical association between a predictor variable and a criterion. A predictor variable is presumed to be the cause of the criterion variable. For research on Type A behavior and arteriosclerosis, Type A behavior is the predictor variable and arteriosclerosis is the criterion. Units of analysis are referred to as observations. For research on Type A behavior and arteriosclerosis, observations are patients that have undergone a coronary angiography. A sample is the collection of all observations included in a single research study.

The attribute or set of attributes that a measurement instrument uses to classify observations for a criterion variable are referred to as decision criteria. For example, the decision criteria is a physi-

cian's judgment of arteriosclerosis based upon the results of a coronary angiography. A decision rule is used to categorize all observations into two and only two mutually exclusive and exhaustive groups. A cutpoint is defined as the value on a measuring instrument associated with the decision rule that is used to categorize observations. Scores that are lower than the value of the cutpoint on the measuring instrument are designated as negatives, and higher values than the value associated with the cutpoint are designated as positives. For example, the cutpoint for the JAS would be the score where all who received higher scores would be considered Type A's and all who received a lower score would be considered Type B's.

All possible categories that result from using decision rules on the predictor and criterion are given in the contingency table illustrated in Figure 1. Actual negatives are observations that the decision criterion indicates are negative. In Figure 1, actual negatives are located in the two squares on the left-hand side of the graph. Actual positives are observations that the criterion indicates are positive. Actual positive observations are represented in the two squares on the right-hand side of Figure 1. For research on Type A behavior and arteriosclerosis, actual negatives would be all subjects that the physician decides do not have arteriosclerosis and actual positives would be all observations the physician labels as possessing arteriosclerosis. Predicted positives are observations that the predictor variable indicates

Figure 1

A Decision Theory Contingency Table.

		Criterion	
		Actual Negatives	Actual Positives
Predictor	Positive	False Positives (FP)	True Positives (TP)
	Negative	True Negatives (TN)	False Negatives (FN)

N = Sample size

$$\chi^2 = \frac{N(|(TN)(TP) - (FP)(FN)| - N/2)}{(TN + FP)(FN + TP)(TN + FN)(FP + TP)}$$

should be actual positives. Predicted negatives are observations that the predictor variable indicates should be actual negatives. For research on Type A behavior and arteriosclerosis, predicted positives would be all individuals classified as Type A's and predicted negatives would be Type B's.

True positives are observations that the decision rule of the predictor variable indicates are positive and are actual positives. That is, true positives are observations where the predictor variable correctly predicts the observations are actual positives. For research on Type A behavior and arteriosclerosis, the true positive cell in Figure 1 would include all patients that are Type A's and that have arteriosclerosis. True negatives are observations where the decision rule correctly indicates are actual negatives. For research on Type A behavior and arteriosclerosis, true negatives would be all observations where the physician decides the patient does not have arteriosclerosis and the JAS score indicates the patient is a Type B. False positives are observations where the decision rule of the predictor variable indicates the observations are positive when the criterion variable indicates the observations are actual negatives. The percentage of false positives among all actual negatives is commonly referred to as the probability of making a Type I error. False positives would be all observations where the physician decides the patient does not have arteriosclerosis but the patient's JAS score indicates they are Type A's. False negatives are

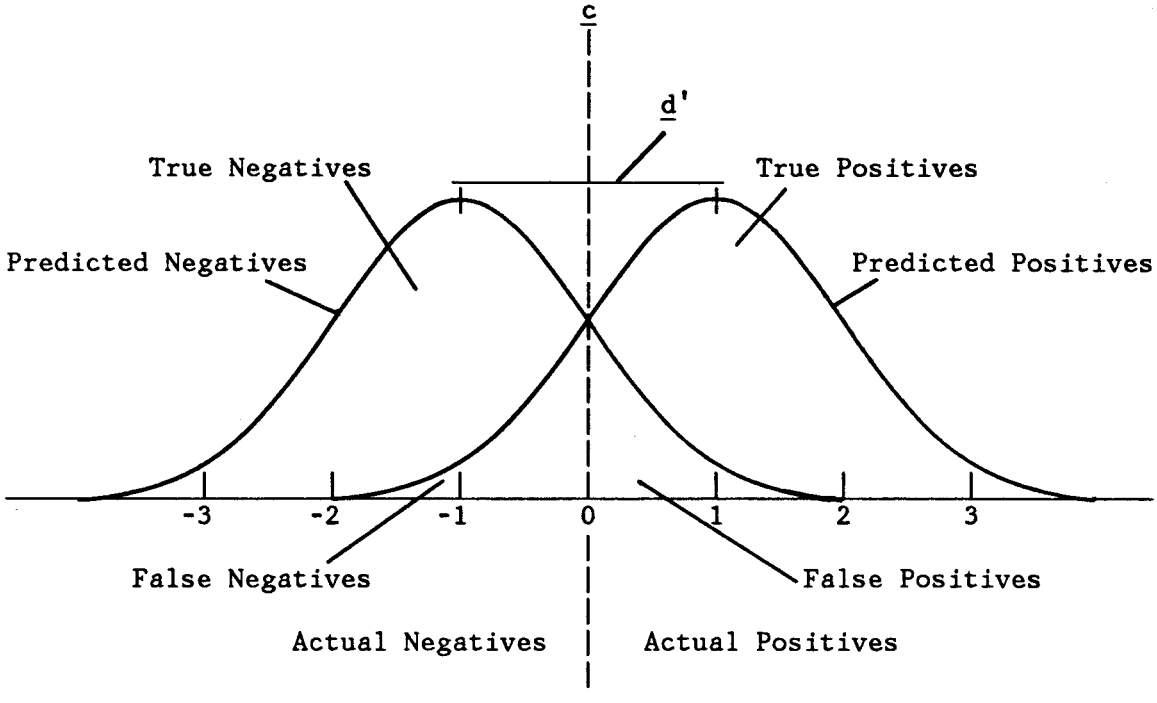
observations where the predictor variable indicates the observations are negative although the observations are actual positives. The percentage of false negatives from among all actual negatives is associated with the risk of making a Type II error. False negatives would be all observations where the physician decides the patient has arteriosclerosis but the patient's JAS score indicates they are Type B's.

Parameters that determine statistical significance. The extent to which the predictor variable is able to accurately classify observations as actual positives or negatives can be evaluated by a χ^2 statistic. The formula given at the bottom of Figure 1 indicates that the size of χ^2 depends upon two parameters (a) sample size, and (b) the ratio of false positives and false negatives to true positives and true negatives. Therefore, larger χ^2's are more likely to occur with larger sample sizes and/or fewer false positives and negatives.

The most well known statistical parameter that can reduce the power of a study is a small sample size. However, parameters other than sample size can influence statistical power. In this respect, the contingency table in Figure 1 is misleading because three additional parameters that influence the size of χ^2 have been implicitly defined a priori within the contingency table. Figure 2 can be used to illustrate how these other parameters influence the χ^2 statistic. Figure 2 represents all the information in the contingency table and includes additional information about several other parameters that can affect the

Figure 2

Terminology Used by Decision Theory.



size of χ^2 including: (a) the magnitude of the standardized difference between the mean score of actual positives and negatives on the predictor variable--$\underline{d}'$, (b) the decision rule that defines the location of the cutpoint--$\underline{c}$, and (c) the extremity and range of scores on the criterion variable.

Figure 2 illustrates the case where the criterion variable is continuous and the predictor variable is dichotomous. The x-axis in Figure 2 represents the continuum of values associated with the criterion variable. The x-axis could represent various degrees of coronary occlusion among patients that have been administered an angiography. The y-axis represents the frequency of occurrence of values on the criterion variable for a corresponding value on the x-axis. For example, the height of the curves in Figure 2 could indicate the number of patients associated with various degrees of arteriosclerosis indicated by different values on the x-axis. The normal distribution on the left hand side of Figure 2 represents the frequency of various degrees of arteriosclerosis for all Type B's in the sample. The distribution on the right hand side of Figure 2 represents all predicted positives Type A's.

Figure 2 displays several parameters that influence the size of χ^2. First, the symbol $\underline{d}'$ is the standardized distance between the means of the distributions of actual positives and negatives and is an indicator of the extent to which actual positives can be distinguished from actual negatives by the predictor variable. For research on Type A

behavior and arteriosclerosis, $\underline{d}'$ would be the standardized difference between the mean arterial disease score for all patients that are Type B's and the mean arterial disease score for all Type A patients. As $\underline{d}'$ becomes large, the proportion of false negatives and positives diminishes and the statistical association between the predictor and criterion variables becomes stronger. Thus, larger values of $\underline{d}'$ indicate a strong association between the predictor and criterion while smaller $\underline{d}'$s are associated with smaller χ^2s.

The size of the χ^2 statistic is often interpreted to be an estimate of the degree of statistical association between two variables but χ^2 is a sample biased statistic. That is, the magnitude of χ^2 is, in part, based upon sample size and, in part, based on $\underline{d}'$. In contrast, $\underline{d}'$ is not dependent upon sample size and, therefore, is a purer measure of statistical association uninfluenced by sample size (Glass, 1976). It should be noted that $\underline{d}'$ is also a direct function of the tetrachoric $\underline{r}$ (Davidoff & Goheen, 1953). Thus, $\underline{d}'$ is directly related to the $\underline{r}$ used in range restriction formulas.

Another feature revealed by Figure 2 is the location of the decision rule. The vertical line labeled $\underline{c}$ is the "cutpoint" associated with the decision rule that determines whether values on the criterion are categorized as actual positives or negatives. Scores on the left side of $\underline{c}$ are observations the decision rule indicates are actual negatives on the criterion--nondiseased. Observations on the right hand

side of c are observations the decision rule indicates are actual positives--a diagnosis of arteriosclerosis.

In Figure 2 observations that are Type B's are predicted negatives and Type A's are predicted positives. Type B's located to the right of c are false positives. Observations located to the left of c that are Type A's are false negatives. Type A's located to the right of c are true positives and Type B's located to the left of c are true negatives.

In Figure 2, the numbered markings on the x-axis indicate the distance in standard deviations from the optimal cutpoint that is located at the point on the x-axis designated by a zero. Note that the optimal cutpoint that maximizes the value of the χ^2 statistic is located where the distributions of Type A's and Type B's intersect (Cureton, 1957). The point is located where an equal number of Type A's and Type B's are present on the x-axis and is the zero point on the x-axis in Figure 2.

Cohen (1977) has presented a series of power tables that allow researchers to determine when the sample size is too small to detect important statistical relationships. Parameters other than sample size have not been subjected to analyses to determine their influence on statistical power. One purpose of this paper is to present some computer simulations that illustrate how other parameters (i.e., the location of the cutpoint and the numbers of predicted positives versus negatives included in the sample) influence statistical significance. These simulations can be used by researchers to determine the statistical power of

their studies just as Cohen 's work has been used to assess the sample size associated with a given degree of statistical power.

Many other characteristics of a sample can influence the statistical power between variables including, degree of error variance, and unequal variances and/or asymmetrical distributions for different levels of the predictor variable. These problems are not discussed in this paper because exclusion bias is not hypothesized to influence these parameters. What is discussed is the degree to which parameters influenced by exclusion bias can reduce statistical significance.

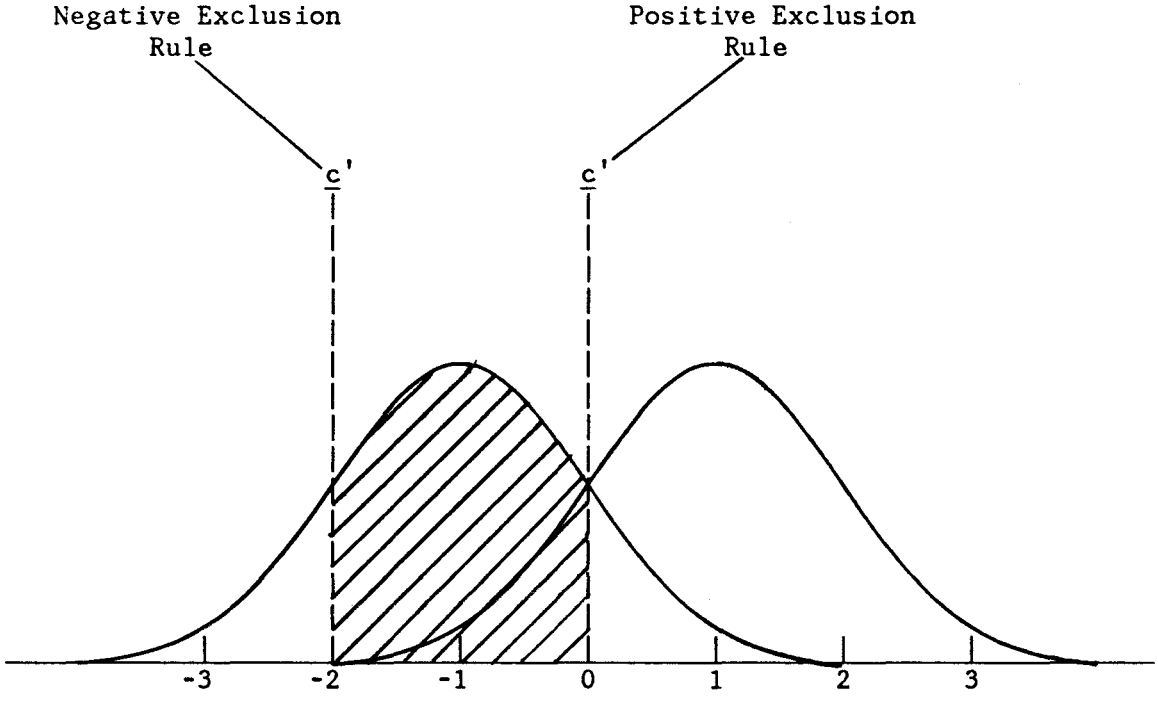
A Statistical Model of Exclusion Bias

Previous work with DT has used symbols and terminology that describe a single sample or universe. The purpose of this paper is to describe how statistical associations change for different subsamples of a population. Therefore, additional symbols and terminology are required to describe the relationship between a sample and the population from which the sample was obtained.

Figure 3 illustrates a statistical model of exclusion bias. Figure 3 includes the same parameters illustrated in Figure 2. As in Figure 2, the x-axis in Figure 3 represents various values on the criterion variable and the y-axis indicates the frequency of observations associated with each value on the criterion variable. In addition, Figure 3 includes some other parameters not illustrated in Figure 2. The shaded portion in Figure 3 represents a subsample selected from the population.

Figure 3

A Statistical Model of Exclusion Bias.



The two lines drawn around the shaded portion of Figure 3 indicate the range of values that have been excluded from the study sample. Shaded observations between the two lines are considered to be observations that have been included in the study sample. The symbol c' associated with each of the lines represents the location of an exclusion rule. A cutoff point or c' is defined as an endpoint associated with a value on the criterion variable where observations are either included or excluded from the study sample. All observations outside the shaded portion of the figure represent observations excluded from the study.

For research on Type A behavior and arteriosclerosis, extreme scores may have been excluded at the c' located on the right hand side of the figure because no person could survive total occlusion of all his/her coronary arteries. Such patients would be excluded from the study a priori because they would have already become ill and so would have either received treatment or previously expired. Another exclusion rule is located on the left hand side of Figure 3. This c' could indicate where patients with little coronary occlusion were excluded. Not surprisingly, some people would never be in a study involving an angiography because they are healthy.

This paper refers to the exclusion rule on the left hand side of the figure as the negative exclusion rule because mostly actual negatives are excluded from the sample. Similarly, exclusion rules located to the right of the study sample are referred to as positive exclusion

rules because actual positive observations are mostly excluded. As decision rules determine the proportion of false positives and negatives within a sample, exclusion rules determine the extent of exclusion bias within a sample.

COMPUTER SIMULATIONS OF EXCLUSION BIAS

This paper proposes that there are three ways that exclusion bias can influence the statistical power of a study. First, exclusion bias may lead researchers to use suboptimal decision rules. In many cases, researchers use a median split or some other arbitrary means to determine the location of the decision rule. Moreover, even when the optimal cutpoint is found in one sample other researchers may find that the same cutpoint is suboptimal in their sample. A computer simulation is used to demonstrate the degree to which using a less than optimal cutpoint can severely reduce a study's statistical power.

Second, this paper demonstrates that exclusion bias can produce unequal numbers of observations for different values on either the predictor or criterion variable. The more extreme the range of values included in the sample, the more disproportionate the numbers of positives and negatives will be. This paper refers to this type of problem as unequal ratios of positives and negatives. Unequal numbers of observations on either the criterion or predictor variable can reduce χ^2. A computer simulation is used to demonstrate that when total sample size is held constant, statistical power decreases as the ratio of predicted and/or actual positives to negatives becomes more disproportionate.

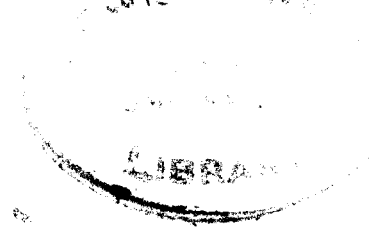
A related problem is that observations surrounding the optimal cutpoint may be excluded from the sample if a disproportionate ratio of

positives to negatives is due to the degree of extremity of scores included in the sample. This paper demonstrates that obtaining statistical significance is impossible when the optimal decision rule is excluded from the study sample.

A third way that exclusion bias reduces the power of a study is by restricting the range of values on the criterion and predictor variables. In Figure 3, the distance between the negative and positive exclusion rules is an indicator of the degree of variance reduction in the sample. The relative importance of restriction of range as compared with other parameters that reduce statistical significance is discussed.

In this section, several computer simulations are used to illustrate how χ^2 is influenced by suboptimal decision rules, unequal ratio of positives and negatives, range restriction, and combinations of all three. The computer simulations presented in this paper were produced by SAS/GRAPH (1984) software. The computer programs that produced the figures are given in Appendix A.

Computer simulations are used in this paper to investigate hypotheses concerning exclusion bias. A computer simulation approach was used because one purpose of this paper is to demonstrate just how much of an effect exclusion bias can have on statistical significance. The results of data from a single study would be less convincing because the results could be attributed to idiosyncrasies in the data. Moreover, the computer simulation approach allows researchers to assess the degree to



which various statistical parameters (e.g., sample size, d', sample variance, and unequal numbers of positives and negatives) influence statistical significance across a wide range of possible conditions. Thus, the simulations can illustrate under what conditions exclusion bias has important consequences for empirical researchers.

Assumptions of the Computer Simulations

Before discussing the computer simulations, the assumptions underlying these simulations are presented. The computer simulations are based upon four assumptions. First, the simulations assume that distributions of values on the criterion for predicted negatives and positives have the same degree of skewness. This assumption was made because asymmetrical distributions alter tests of significance. One purpose of this paper is to demonstrate that all other things being equal, exclusion bias will reduce statistical significance. Exclusion bias is not hypothesized to influence the degree of skewness between different levels of the predictor variable. Therefore, the computer simulations assume that the frequency distributions of predicted positives and negatives are symmetrically distributed across the range of values on the criterion variable. The variables were assumed to be a normally distributed to ease the computational formulas used by the computer simulation program.

One would expect that the distribution of various degrees of arteriosclerosis would be highly positively skewed because most individ-

uals in the general population would have very little arteriosclerosis present. However, normalizing transformations are usually used before data analysis takes place for such medical variables (see Steel & Torrie, 1980).

A second assumption of the simulations is that the variances of the distributions of actual positives and negatives are equal. Rorer et al. (1966a) demonstrated that unequal variances change the location of optimal cutpoints and can alter the potential accuracy of a predictor variable. The variances of Type A and B scores for various degrees of coronary arteriosclerosis is not known because researchers have operationalized arteriosclerosis as a discrete variable. Thus, somewhat arbitrarily, the simulations assume that the variances of predicted positives and negatives are equal.

To ease interpretation of the simulations the variances of predicted positives and negatives for the simulations were set to equal to 1.0 so that the simulations can be reported as if the results are being reported in standardized scores. For example, the extremity of ranges included in various samples can then be expressed as differences in the number of standard deviations from the population mean.

A third assumption for most of the simulations is that the numbers of predicted and/or actual positives and negatives in the sample are equal. Unequal numbers of positives and negatives change the location of the optimal cutpoints (Rorer et al., 1966a). The degree to which

unequal numbers of positives and negatives can influence statistical significance is discussed later.

A fourth assumption of the simulations is that statistical significance is only obtained when there is a less than 1 in 20 twenty chance of committing a Type I error and the value of committing a Type II error is dependent upon sample size. The assumptions were made because that is the accepted standard for tests of significance in research studies. For this paper, the χ^2 statistic is used to assess statistical significance. In applied research, Type II errors may be more serious. For example, Type II errors would be associated with the JAS suggesting that patients do not have arteriosclerosis when they actually do. Rorer, Hoffman, and Hsieh (1966b) have demonstrated how to locate optimal cut-points when Type I and Type II errors are to be weighted in some other fashion.

Figure 2 represents the distributional characteristics of predicted negatives and positives based upon the assumptions used in this paper. That is, the distributions of predicted positives and negatives are normally distributed and have equal sample sizes with variances equal to one.

Suboptimal Decision Rules

Reasons for the use of suboptimal decision rules. DT can be used to evaluate problems associated with the use of suboptimal decision rules. DT has been used in medical research and is the basis for the

well known medical concepts of sensitivity and specificity (Metz, 1978) and has occasionally been used to identify the optimal decision rules for diagnostic tests (Swets Pickett, Whitehead, Getty, Schnur, Swets, & Freeman, 1979). DT also provides the basis for clinical decision analysis (Weinstein & Fineberg, 1980). Nevertheless, suboptimal decision rules continue to be used in many areas of medical research (Christensen-Szlankski, Diehr, Bushyhead, & Wood, 1978).

Assuming that false positives and negatives are to be considered equally costly, the point that maximizes the value of the χ^2 statistic is located where the distribution of predicted positives and negatives intersect. Although Rorer et al. (1966a) have recommended that researchers use optimal cutpoints in their samples, most researchers continue to choose their cutpoints arbitrarily. The use of suboptimal decision rules is a serious problem in many areas of research. For example, an important part of medical research is concerned with evaluating the efficacy of diagnostic tests. Typically, the actual cutoff point used to determine which individuals are diseased and which are nondiseased is chosen somewhat arbitrarily. Therefore, the researcher may falsely conclude a diagnostic test is of little value when a suboptimal cutoff point is used. However, the same test may have had great diagnostic value if the appropriate cutoff point had been chosen.

Many researchers appear to choose their study sample cutpoint by using the point that equally divides study sample observations into

equal numbers of actual positives and negatives. However, this practice may lead to the use of suboptimal decision rules. For example, if the sample illustrated in Figure 3 selected a cutpoint using a median split the point chosen would be located -1 standard deviations from the optimal cutpoint.

Note that the proportions of false negatives, false positives, true negatives and true positives changes as $\underline{c}$ changes. For example, when $\underline{c}$ is one standard deviation to the left so that $\underline{c}$ is located above the -1 mark on the x-axis in Figure 3, the figure has many more false positives and slightly fewer false negatives. Overall there are more false positives and negatives so the value of χ^2 is less. Because χ^2 is determined by the frequency of false positives and negatives (see formula displayed in Figure 1), as $\underline{c}$ varies so will χ^2. Thus, researchers that arbitrarily use a median split may be reducing the statistical power of their studies.

Note that the location of the optimal cutpoint does not change if some observations are excluded from the study sample. Therefore, the optimal cutpoint is located at the same point in Figure 3 regardless of where the exclusion rules are located.

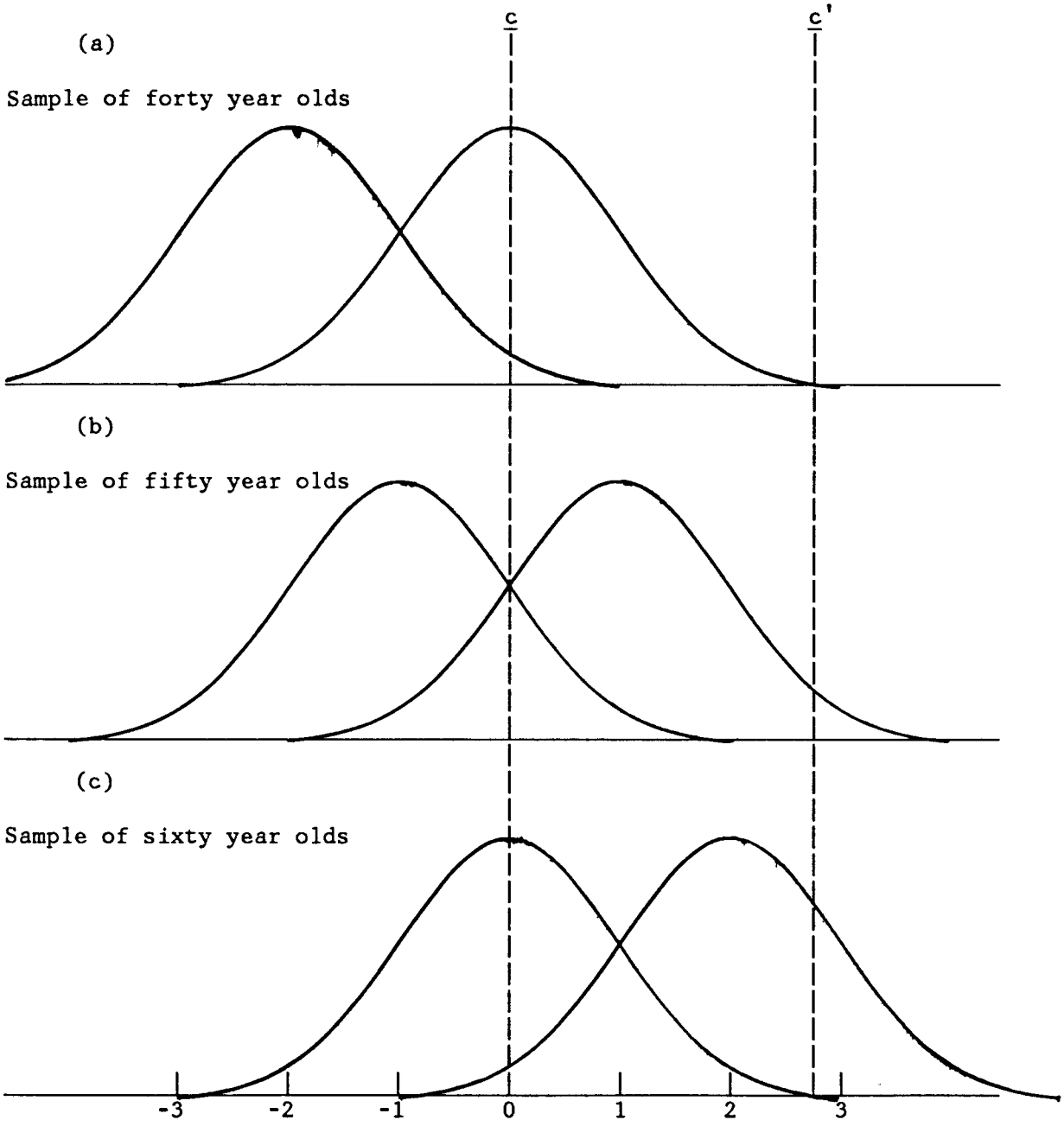
Nevertheless, even when a cutpoint is used that finds statistically significant results, the cutpoint may not be appropriate for another study sample. Thus, other researchers may not be able to replicate a previous study's significant results. The best way to describe

how the location of optimal cutpoints can vary from study sample to study sample is to use a hypothetical example. Suppose a physician were to conduct coronary angiographies on the same group of men at three points in time: Once when the men were all 40 years old, once when they were 50 years old, and again when they were 60 years old. Also, suppose that all of these men have a family history of heart disease, are heavy smokers and drinkers, and have high cholesterol diets. In other words, the hypothetical study sample includes a group of men that are at high risk to develop arteriosclerosis and so most eventually will. Therefore by the time the physician conducts coronary angiographies at 60 years of age, most of them have at least one artery that is more than 50% occluded. Also assume that half of the men are Type A's and the other half are Type B's.

The series of graphs in Figure 4 represent the hypothetical study. As with previous figures, the x-axis in Figure 4 represents values on the criterion variable (e.g., percentage of occlusion in most severely diseased artery) and the y-axis indicates the frequency of observations for any given value on the x-axis. The line labeled c' indicates where the patients' disease is severe enough that death or medical intervention (e.g., a coronary bypass operation) has occurred. Thus, observations to the right of c' are excluded from the study. The diagram located at the top of Figure 4 labeled '(a)' represents the sample of men when they are forty year old, the middle diagram labeled

Figure 4

A Hypothetical Longitudinal Study of Arteriosclerosis and Type A Behavior.



'(b)' represents the sample when the men are fifty years old, and the bottom diagram labeled '(c)' represents the sample when the men are sixty years old. As one might expect, as the age of the men increases the proportion of men that have arteriosclerosis increases. In Figure 4, the distributions of Type A's are located to the right of the Type B distributions because Type A's should develop arteriosclerosis sooner than Type B's if Type A behavior really does cause heart disease. Therefore, Type A's in Figure 4 are illustrated as developing arteriosclerosis at earlier ages than Type B's.

Note that in the graphs in Figure 4 that as the range of values in the sample changes as the men become older, the location of the optimal cutpoint also changes. Thus, the location of the optimal cutpoint may vary from sample to sample depending upon the range of values included in the sample for other variables (i.e., age) that are correlated with either the predictor or criterion. Thus, researchers that use cutpoints that were optimal in one sample may not be optimal for another sample. For example, researchers in Type A research may find a statistically significant result in one sample but not in another because the ages of the patients in the sample may differ.

Rorer et al. (1966a) have demonstrated how researchers can determine the optimal cutpoints for their samples; however, Rorer et al. did not demonstrate how <u>much</u> of a problem using less than optimal cutpoints can be. The purpose of the first two computer simulations is

to illustrate the degree to which using a suboptimal cutpoint can influence the statistical significance of a study.

The relationship between χ^2 and c. The purpose of the first two simulations is to demonstrate the degree or extent of influence that c has on χ^2. For the first simulation, a computer program was written that calculated χ^2 values given a value of c and a value of d'. Values of d' were varied from .3 to .7 by intervals of .1 and values of c were varied from 2.5 standard deviations below to 2.5 standard deviations above the optimal cutpoint at intervals of .2 standard deviations. Values of d' were varied from .3 to .7 because this range of values represents an average range of values found in social science research (Glass, 1976). The actual size of d' in Type A behavior and arteriosclerosis research is not known. Previous researchers have not reported their d's and these would probably be biased estimates anyway because of the presence of exclusion bias.

The range of values of c was varied from -2.5 to 2.5 standard deviations because this is the range of values that the χ^2 statistic is an accurate indicator of statistical significance. For values outside the -2.4 to 2.4 range the cell counts are likely to become less than 5 per cell and the χ^2 statistic is no longer an appropriate indicator of statistical significance.

For the first simulation, sample size was held at a constant value of 200. The value of 200 was chosen because the sample size is slightly

larger than what most studies on the relationship between Type A behavior and arteriosclerosis have used. A slightly larger value was chosen to demonstrate that using a less than optimal cutpoint can have an effect on χ^2 even when the researcher uses a large sample size.

Table 1 gives a list of the sample sizes used in previous Type A behavior and arteriosclerosis research that used the SI to assess Type A behavior. The list of published studies in Appendix B was obtained from a computer assisted search of the past ten years of Psychological Abstracts and Medicus Index. The search revealed 27 published articles and one dissertation on the relationship between Type A behavior and coronary arteriosclerosis that used the results of a coronary angiography as a criterion (see Appendix B). The fifth column from the right hand side of Table 1 gives the sample sizes of previous studies. On the average, most studies have used sample sizes between 100 and 150.

For the first simulation, no truncation of the study sample was assumed because the purpose was to demonstrate the influence of suboptimal cutpoints on χ^2 and not other factors influenced by exclusion bias that reduce statistical significance (i.e., a reduction in variance).

Figure 5 illustrates the results of the computer simulation. The y-axis indicates the size of χ^2 while the x-axis indicates the location of c. For research on Type A behavior and arteriosclerosis c would be the cutpoint where the physician decides that enough coronary occlusion

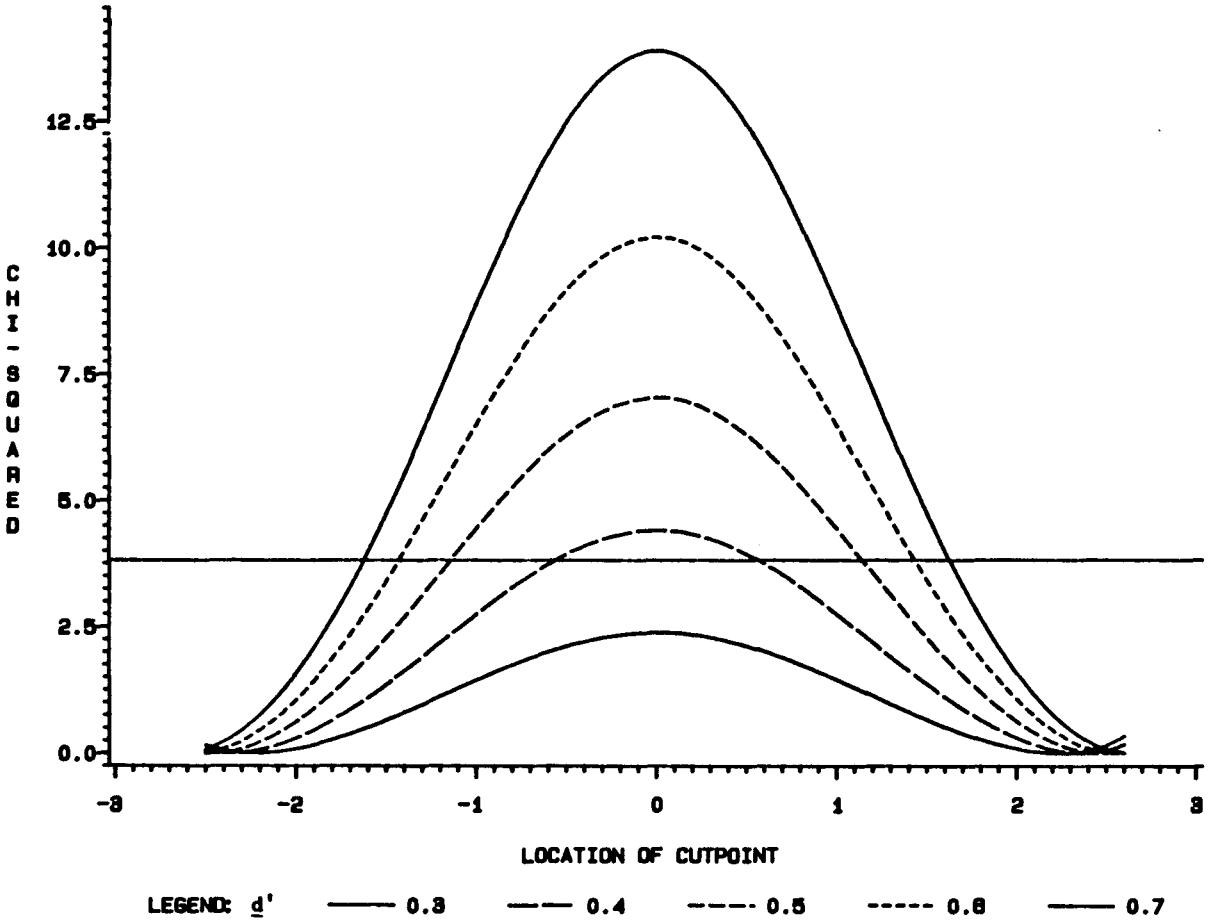
Table 1

Results of Angiography Studies that Used the SI.

AUTHOR	%A's	Scoring method	% ill	Age range	N	p< .05	%A's ill	%B's ill	χ^2
Blumenthal et al. (1978)	60	TOTCI	45	15-69	142	+	60	23	17.29
Krantz et al. (1981)	75	TOTCI	59	30-67	83	-	61	52	.001
Blumenthal et al. (1985)	65	TOTCI	73	20-71	281	-			
Dembroski et al. (1985a)	63	TOTCI	69		132	-			
Williams et al. (1980)	75	>75%	70		424	+	71	56	7.35
Arrowood et al. (1982)	61	>75%			75	-			
Frank et al. (1978)	73	>50%	80	29-65	147	+	87	59	12.22
Krantz et al. (1981)	79	>50%	78	30-67	83	-	73	76	.34
Schwertner et al. (1982) *	42	>50%			50	+	52	14	
Scherwitz et al. (1983)	70	>50%	>85	35-69	52	-			
Dimsdale et al. (1979a,b, 1980)	64	>50%	84	18-70	103	-	85	84	.02

TOTCI = Total Coronary Index
* = This study used only Type A1's and excluded A2's from analysis.
+ = Study found a statistical significant association.
- = Study did not obtain statistical significance.

FIGURE 5
THE RELATIONSHIP BETWEEN CHI-SQUARED
AND THE CUTPOINT FOR VARIOUS EFFECT SIZES



is present to conclude that the patient has arteriosclerosis. The straight horizontal line that runs across the middle of Figure 5 indicates where χ^2 values reach statistical significance. The χ^2 values that appear above the line are statistically significant and values below the line are nonsignificant.

Each point on the curved lines in Figure 5 corresponds to the single χ^2 value on the y-axis that is associated with a c on the x-axis. Each of the curves is associated with a single value of d'. The value of d' associated with each curve is indicated on the legend located at the bottom of Figure 5. Thus, each curved line represents relationships between χ^2s and cs for a given d'. For research on Type A behavior and arteriosclerosis d' would represent the difference between the average degree of coronary occlusion in Type A's and B's. The curves that are located towards the top part of the graph are curves associated with larger d's. The curves with larger d's are located above the curves with smaller d's because larger d's increase the size of the χ^2 statistic and so are generally associated with larger χ^2 values. Note for a d' of .3 it is virtually impossible to obtain statistical significance and that the larger the value of d' the greater the range of values that are statistically significant.

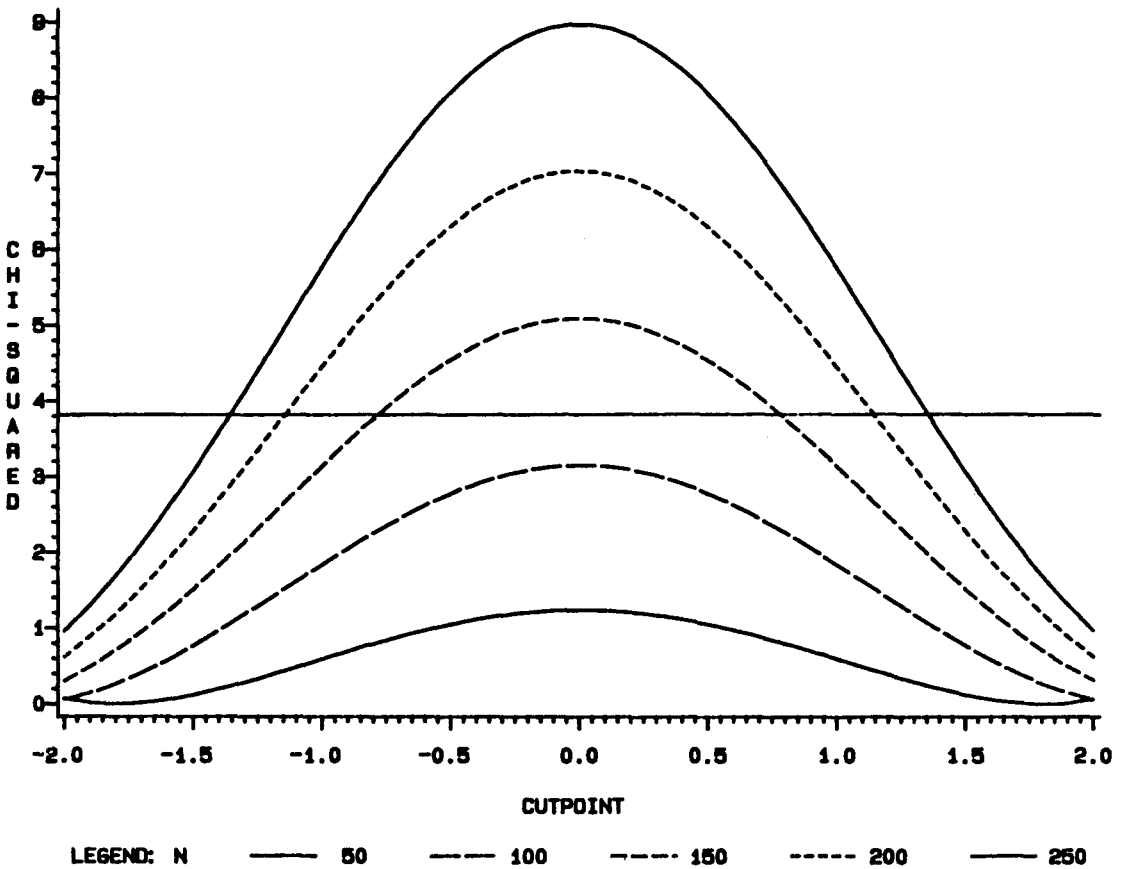
Note that cs located further away from the optimal cutpoint (the zero point) are associated with smaller χ^2s. This occurs regardless of whether the cutpoint is located to the left or the right of the optimal

point. Note that, statistical significance is never achieved if the cutpoint is located two standard deviations or more from the optimal c. For approximately half of the values presented in Figure 5 statistical significance was not obtained. The slope of each curve indicates the degree of influence that the location of the cutpoint has on statistical significance. Note that the decrease in statistical significance is more dramatic for larger d's. For example, when d' is equal to .6, the χ^2 statistic decreases approximately 50 percent per standard deviation increase in distance from the optimal cutpoint. Thus, for larger d's extremely deviant cutpoints are still not statistically significant. Therefore, even for very strong relationships between a predictor and criterion variable the use of a highly deviant cutpoint can insure nonsignificant findings.

The influence of sample size on the relationship between c and χ^2. Another simulation was conducted to demonstrate the effects of suboptimal decision rules for different sample sizes. For this simulation, d' was fixed at a constant value of .5. For most areas of research a d' of .5 would be considered a moderately large effect (Glass, 1976). Sample size was varied from 250 to 50 by intervals of 50 and the cutpoint was varied from -2.0 to 2.0.

Figure 6 illustrates the results of the simulation. As in Figure 5, the x-axis in Figure 6 indicates the location of the cutpoint and the y-axis indicates the value of χ^2. Again, the values above the

FIGURE 8
CHI-SQUARED BY CUTPOINT
FOR SEVERAL DIFFERENT SAMPLE SIZES



horizontal line that runs across the middle of the figure are statistically significant and values below the line are nonsignificant. Each curved line in Figure 6 represents a different sample size. The sample size associated with each line is indicated on the legend below the figure. Figure 6 is identical in all respects to Figure 5 except that while each curved line in Figure 5 was related to a different $\underline{d}'$. Each curved line in Figure 6 is associated with a different sample size.

In Figure 6, the five curves each correspond to a different sample size. The curve associated with the highest χ^2 values corresponds to a sample size of 250, the next highest curve represents a sample size of 200, and so forth. As one would expect, Figure 6 illustrates that for smaller sample sizes fewer values are significant. For sample sizes less than 150, statistical significance is never obtained.

As in Figure 5, the steepness of the curves indicates the influence that the location of $\underline{c}$ has on χ^2. For cutpoints located further from the optimal cutpoint statistical significance decreases. Statistical significance is never obtained for cutpoints that are more than 1.5 standard deviations from the optimal cutpoint. The range of statistically significant values increases for greater samples. Thus, for a sample size of 250 the range of statistically significant values covers 3 standard deviations. For a sample size of 150, the range of statistically significant values is 1.6 standard deviations. However, the curves become steeper for greater sample sizes. Thus, the range of

statistically significant values does not increase rapidly for larger sample sizes. The results of the simulation suggest that even for large sample sizes a researcher that uses extremely deviant cutpoints may not obtain statistically significant results.

Implications for research on Type A behavior and arteriosclerosis. It is difficult to say how much the actual results of research on Type A behavior and arteriosclerosis have been affected by suboptimal decision rules because the actual d's, and cs found in Type A research may be much different. However, the simulations do show that across a wide range of circumstances suboptimal decision rules can dramatically reduce statistical significance. There is some reason to suspect that suboptimal decision rules have been used because research on arteriosclerosis has used many different decision rules. Some researchers have only considered patients diseased if 50% stenosis of one artery is present while others have used 75% as the decision rule (see Table 1). Blumenthal, Williams, Kong, Schanburg, and Thompson (1978) used a Total Coronary Index (TOTCI) score where each major coronary artery vessel is rated on a four point scale. Young, Barboriak, Anderson, and Hoffman (1980) used a coronary occlusion score. Each of the three major coronary arteries are given a score from 0 to 100% occlusion. The left main coronary artery score is weighted double because it is larger and more important than the other coronary arteries.

Despite the abundance of decision rules, previous researchers have not reported using any systematic techniques to locate optimal cutpoints. For example, the 50% occlusion score was chosen because patients with one artery that is 50% occluded usually report that they only experience angina after physical exertion (Pearson, 1983). Moreover, research has demonstrated that physicians in clinical practice use different decision criteria for determining what patients have arteriosclerosis (Hlatky et al., 1983).

Similarly, studies to determine the optimal cutpoints for measures of Type A behavior have not been conducted and researchers have not produced any evidence suggesting that the cutpoints obtained from the Western Collaborative Group Study (Rosenman et al., 1964) were optimal for that sample or for any other.

In sum, the wide variety of cutpoints used by researchers without any reports of attempting to locate optimal cutpoints suggests that suboptimal cutpoints have been used. The simulations presented suggest that under many circumstances using suboptimal cutpoints can lead to nonsignificant findings even when a strong relationship between a predictor and a criterion does exist. Thus, the negative findings found in many of the studies reported in Table 1 may be due to the use of suboptimal decision rules.

Unequal Numbers of Positives and Negatives

Unequal numbers of predicted positives and negatives. Another factor that can influence statistical significance is the degree to which the study sample includes more or less actual positives than negatives. Similarly, the degree to which the study sample includes more or less predicted positives than negatives also influences statistical significance.

The Taylor-Russell Tables (Taylor & Russell, 1939) give the percentage of true positives for a particular ratio of actual positives to negatives and a given size of the correlation between the predictor and criterion variable. This is useful to determine the practical value of a test. In personnel research, the number of true positives represents job candidates that would be successful and have passed the test. The Taylor-Russell tables can be used to determine the increase in the percentage of successful employees a valid selection would produce. Although the table may be useful to applied researchers, it is of limited usefulness when the researcher is interested in determining statistical power. The following simulations illustrate the influence that different ratios of positives and negatives can have on statistical significance.

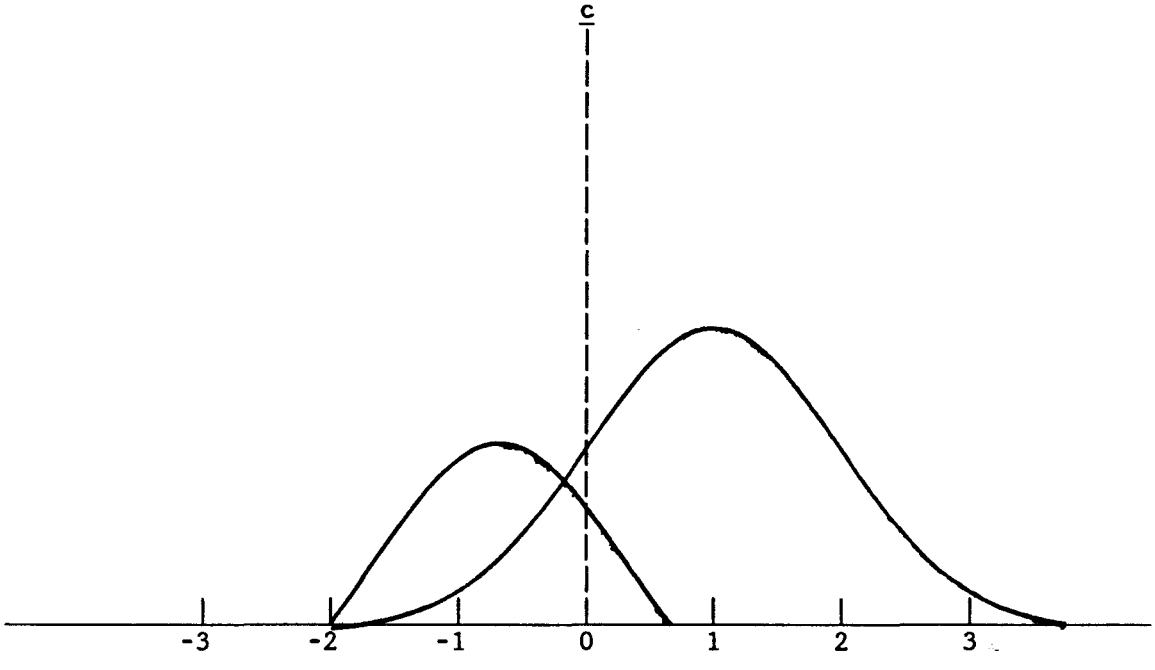
The relationship between the ratio of predicted positives to negatives and χ^2 for various d's. The decision rule used by the predictor variable to classify observations as Type A's and B's is not presented in Figure 2 or any of the other figures. However, the effects of the decision rule can be inferred because the location of the decision rule influences the ratio of positives and negatives in the study sample. For example, the distribution of predicted positives would be smaller and the distribution of Type B's would be larger if the cutpoint on the JAS is raised so that only very high scores are considered to be Type A's. Conversely, fewer observations would be diagnosed as being Type B's and more as Type A's if the cutpoint was lowered.

The extent to which the study sample includes more or less positives than negatives can be expressed as a ratio. For this paper, the ratio of positives divided by negatives is used. The extent to which a ratio of positives to negatives departs from one indicates the degree of inequality that exists. The next two computer simulations are used to illustrate that as the ratio of predicted positives to negatives departs from 1.0, a larger sample size is required for statistical significance to be achieved.

Figure 7 illustrates a situation where more positives are present in the study sample than negatives. Figure 7 is similar to Figure 2 in all respects except the distribution of predicted positives is larger

Figure 7

A Population with Greater Numbers of Predicted Positives than Negatives.

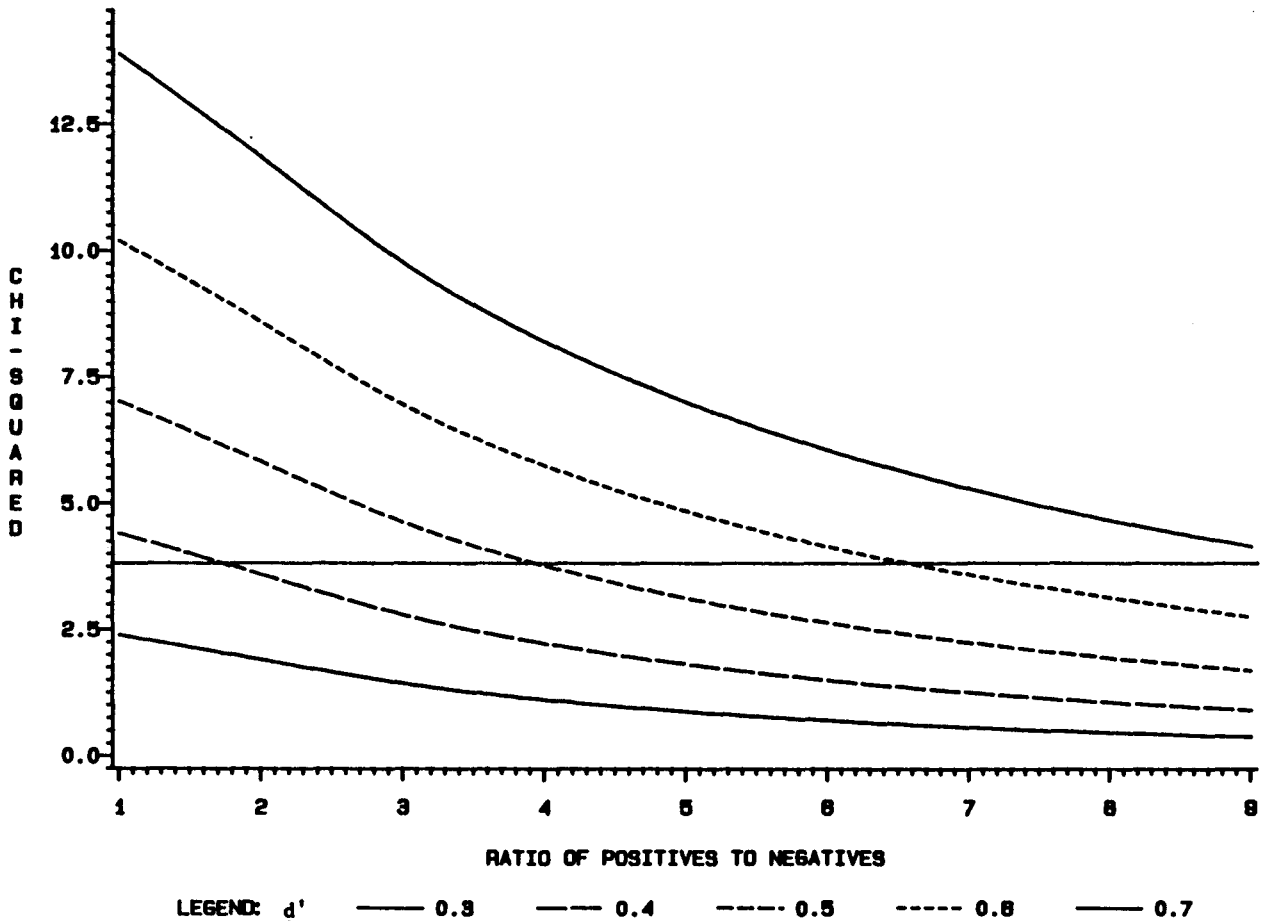


than the distribution of predicted negatives. In Figure 7, predicted negatives could be Type B's and predicted positives would be Type A's. As usual, the y-axis indicates the frequency of observations. Thus, the larger distribution on the right hand side of Figure 7 indicates there are more predicted positive observations--Type A's--than the distribution of predicted negatives--Type B's--on the left.

The next simulation is used to demonstrate that as the ratio of predicted positives to negatives departs from one, the likelihood of obtaining statistically significant results decreases if sample size is held constant. Sample size was held at a constant value of 200 and $\underline{c}$ was adjusted for all calculations so it was located at its most optimal point. The value of $\underline{d}'$ was varied from .3 to .7 by intervals of .1. The ratio of predicted positives to negatives was varied from 1 to 9 by intervals of 1. Note that a ratio of 3 to 1 corresponds to a 75 percent to 25 percent distribution of positives to negatives. This is approximately equal to the ratio of Type A's to Type B's, and diseased to nondiseased found in most research on Type A behavior and arteriosclerosis (see Table 1). As in the previous simulations, no truncation of variables was assumed so that the effects of an unequal ratio of positives to negatives could be assessed independently.

Figure 8 illustrates the relationship between $\underline{d}'$, χ^2, and the ratio of predicted positives to negatives. The y-axis in Figure 8 indicates the value of χ^2 and the x-axis represents the number of

FIGURE 8
CHI-SQUARED BY RATIO OF PREDICTED POSITIVES TO NEGATIVES FOR VARIOUS EFFECT SIZES



predicted positives divided by the number of predicted negatives. Each point on the curve corresponds to a χ^2 value on the y-axis and a ratio of positives to negatives on the x-axis. That is, each point represents the χ^2 that would be obtained for a given ratio of predicted positives to negatives.

Each curved line in Figure 8 illustrates the relationship between χ^2 and the ratio of positives to negatives for a single $\underline{d}'$. Again for large $\underline{d}$'s, the χ^2 values are greater. Thus, curves associated with larger $\underline{d}$'s are located higher in the figure. The $\underline{d}'$ associated with each curve is indicated on the legend below Figure 8.

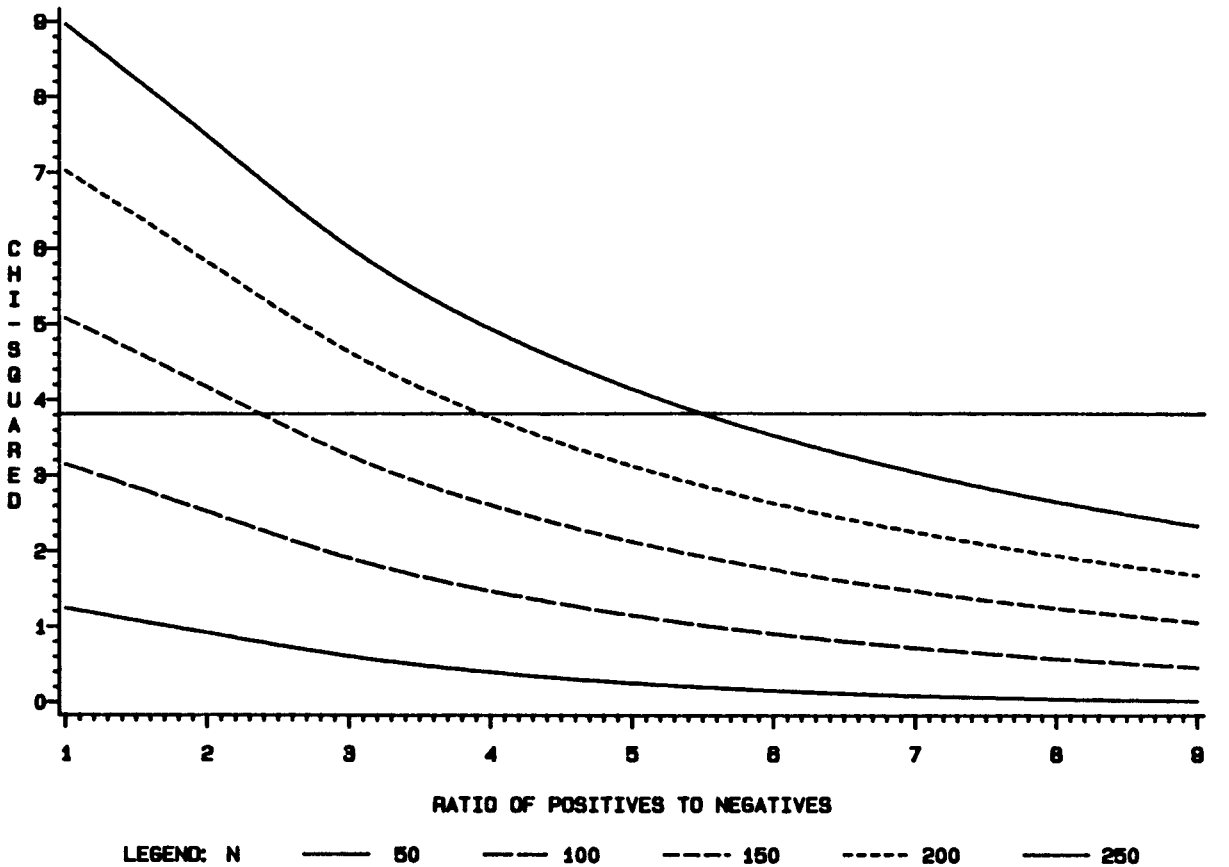
The steepness of the curves illustrates the effect that a disproportionate ratio has on statistical significance. Note that for ratios between 1 and 3 statistical significance is more sharply reduced and for ratios greater than 3 statistical significance decreases more gradually. Figure 8 illustrates that even for large $\underline{d}$'s, a disproportionate ratio of predicted positives to negatives lead to nonsignificant findings. Only a $\underline{d}'$ of .7 is significant when the ratio of positives to negatives is greater than 9 to one. For smaller $\underline{d}$'s, statistically nonsignificant results are obtained for much smaller ratios. For larger $\underline{d}$'s, the slopes of the curves are greater suggesting that a large $\underline{d}'$ cannot correct for an extremely disproportionate ratio.

The relationship between χ^2 and the ratio of predicted positives and negatives for various sample sizes. Another simulation is used to demonstrate that even for large sample sizes a statistically significant result is difficult to obtain if unequal numbers of positives and negatives are present. For this simulation, $\underline{d}'$ was held constant at a value of .5. Sample size was varied from 50 to 250 by intervals of 50. As in the previous simulations, the ratio of positives to negatives was varied from 1 to 9. The computer simulation program calculated χ^2 values for all possible combinations of ratios and sample sizes.

The results of the simulation are presented in Figure 9. Figure 9 is similar to Figure 8 in all respects except that the curved lines represent different sample sizes with $\underline{d}'$ held at a constant value of .5. Again the x-axis indicates the ratio of positives to negatives and the y-axis indicates the value of χ^2. Each point on a line indicates the χ^2 associated with a ratio of positive to negatives. Each curve in Figure 9 represents the relationship between χ^2 and the ratio of predicted positives to negatives for a different sample size.

Note that for larger sample sizes the range of χ^2 values that are statistically significant is greater. Thus, the curves located at the top of the graph correspond to larger sample sizes. The sample size that corresponds to each curve is indicated on the legend at the bottom of the figure.

FIGURE 8
CHI-SQUARED BY RATIO OF PREDICTED POSITIVES TO NEGATIVES FOR VARIOUS SAMPLE SIZES



The steepness of the curves illustrates the influence the ratio of positives to negatives has on statistical significance. Note that for each curve statistical significance is reduced as the ratio of positives to negatives becomes greater. Statistical significance is never obtained for ratios greater than 5. For larger sample sizes, the values of χ^2 decreases more rapidly. Thus, the range of statistically significant values does not increase as much as one might expect for larger sample sizes. Even for a sample size of 250, statistical significance is not possible if the ratio of positives to negatives is greater than five. Thus, very large sample sizes may be necessary if the ratio of positives to negatives is great.

Implications for Type A and arteriosclerosis research. There is abundant evidence suggesting that unequal numbers of predicted positives and negatives have played a role in Type A/arteriosclerosis research. The first column from the left in Table 1 gives the percentages of Type A's in the sample. In Table 1, the ratio of Type A's to Type B's varies from 1.5 to 3. Yet in studies that sampled normal healthy individuals (e.g., Rosenman et al., 1966) the ratio of Type A's to Type B's was approximately equal to one. This suggests that Type A/arteriosclerosis researchers are using a sampling frame that includes a greater proportion of observations to the right of the optimal cutpoint where more Type A's are present than Type B's. Thus, researchers may be underestimating the strength of the relationship between Type A behavior

and arteriosclerosis because unequal numbers of positives and negatives have reduced statistical significance in many studies.

The problem of unequal ratios of positives to negatives is even greater for prospective studies of heart disease where the vast majority of subjects remain well for the duration of the study. For example, ratio bias has drastically reduced the statistical associations found in the Western Collaborative Group studies. The Western Collaborative Group Study (Rosenman, et al., 1964) was a prospective study lasting eight and one-half years of middle aged 30 to 50 year old men. After two years the proportion of nondiseased to diseased men who were between 39 and 49 years old (Rosenman, Friedman, Straus, Wurm, Jenkins, & Messinger, 1966) was eighty-four to one, after four and one-half years (Rosenman, Friedman, Straus, Jenkins, Zyzanski, & Wurm, 1970) the ratio was forty to one and after eight and one-half years (Rosenman, Brand, Jenkins, Friedman, Straus, & Wurm, 1975) the ratio was sixteen to one. Only the studies at four and one half years and eight and on-half years reached statistical significance.

Unequal numbers of actual negatives and positives. The effects of unequal numbers of actual positives and negatives is statistically equivalent to the problem of unequal numbers of predicted positives and negatives. Therefore, no additional computer simulations need to be produced. However, unequal numbers of actual positives and negatives does have different implications for research on Type A behavior and arteriosclerosis.

Research suggests that coronary angiography studies have sampled much more diseased samples than exist in the general population. Angiography studies only include patients that agree with their physician that they should be willing to undergo an angiography. Physicians and patients will be reluctant to use the procedure unless they are fairly certain that an angiography will find some arteriosclerosis present. Therefore, most patients will have some arteriosclerosis.

The fourth column of Table 1 indicates the percentage of observations from each study that were diseased. For all of the studies, more patients were diagnosed as having arteriosclerosis than not. Most studies (see Table 1, column 3) included patients that have a high degree of disease. In contrast, most validity studies of coronary occlusion in individuals that die from violent deaths reported that approximately 20% of the population had 50% occlusion in at least one coronary artery (see Pearson, 1984). Rissanen (1975) found that for men between the ages of 45-64, approximately 40% had occlusion. One would suspect that most observations in Type A/arteriosclerosis research are located to the right of the optimal cutpoint because the percentage of occlusion in coronary angiography studies is much higher than the degree of occlusion found in autopsy studies.

Note that in the study illustrated in Figure 4 only the fifty year old sample has equal numbers of actual positives and negatives. The

forty year old sample has fewer observations diagnosed as having arteriosclerosis while the sixty year old sample has more observations diagnosed as having arteriosclerosis. Therefore, the researcher may only find a statistically significant relationship in the sample of fifty year olds because the numbers of actual positives and negatives are unequal for the forty and sixty year old samples. Type A behavior and arteriosclerosis research may best be represented by the sample of sixty year olds. That is, most individuals scheduled for angiography are high on some risk factors for disease (although the risk factor may be some factor other than age). As in the sixty year old sample, most observations will be diseased and very few patients will be nondiseased.

In sum, the results of the preceding simulations suggest that unequal numbers of positives and negatives are biasing the results of angiography studies towards failing to reject the null hypothesis.

Extremity of the Range of Sample Values

Factors that produce samples that only include extreme scores. Exclusion bias can be produced by differences in values of clinicians and applied researchers. Under most circumstances, clinicians consider Type II errors as more costly. Thus, there is a tendency to reduce the number of Type II errors by excluding negatives--either predicted or actual negatives--from the sample. For example, clinicians attempt to administer angiographies only to patients that are diseased. This tendency can be described mathematically as moving the negative

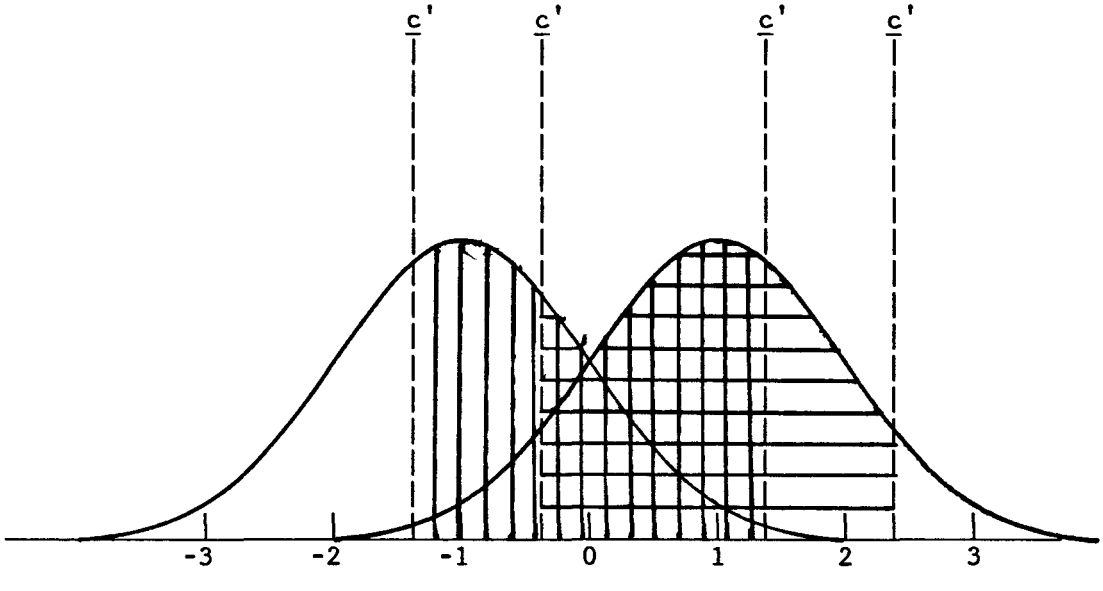
exclusion rule to the right, thereby excluding more negatives. As the exclusion rule moves further to the right, observations located around the optimal cutpoint may be excluded from the study sample and the ratio of actual positives to negatives becomes larger.

Figure 10 illustrates the case where the physician's clinical judgment improves so that the negative exclusion rule moves to the right thereby excluding more negatives from the sample. Figure 10 also illustrates the hypothetical study sample distributions that would account for why one study may find statistical significance and another would not. Figure 10 is identical in all respects to Figure 4 except that Figure 10 illustrates two as opposed to one study samples. The portion of the graph in Figure 10 shaded with horizontal lines represents the sampling frame of a study that would not find statistically significant results. The portion of the graph shaded by vertical lines represents the range of the sample distribution of the study that would find statistically significant results. Notice that the nonsignificant study sample includes more observations located to the right of the optimal cutpoint. Thus, the nonsignificant sample distribution includes a more disproportionate number of actual and predicted positives.

For research on Type A behavior and arteriosclerosis, the nonsignificant study findings may have been produced because an experienced physician did not subject many nondiseased individuals to an

Figure 10

The Hypothesized Relationship Between Significant and Nonsignificant Studies Produced By Extremity of Range.



angiography. The significant study would represent the outcomes of a study from a physician with less experience. Thus, there are equal numbers of diseased and nondiseased subjects in the sample.

As mentioned previously, statistical significance cannot be obtained when observations around the optimal cutpoint have been excluded. This represents a more extreme case of the nonsignificant study finding illustrated in Figure 10. Even though the criterion and predictor variable are strongly associated, no statistically significant association can appear in the study sample regardless of sample size. This is the case because the χ^2 statistic assesses the degree to which greater numbers of true negatives than false negatives and greater numbers of true positives than false positives are present within the study. In Figure 3, more actual positives are present throughout the entire range of the study sample regardless of where the decision rule is placed. Therefore, more false negatives are present regardless of where the decision rule is placed. Thus, finding statistical significance is not possible.

Implications for research on Type A behavior and arteriosclerosis. In the present section, some data from actual studies of coronary angiography is used to illustrate the influence that extremity of scores can have on statistical significance. The two graphs in Figure 11 illustrate the two published studies (Blumenthal et al., 1978; Krantz, Schaeffer, Davia, Dembroski, MacDougall, & Schaeffer, 1981) that

Figure 11(a)

Blumenthal et al. (1978)

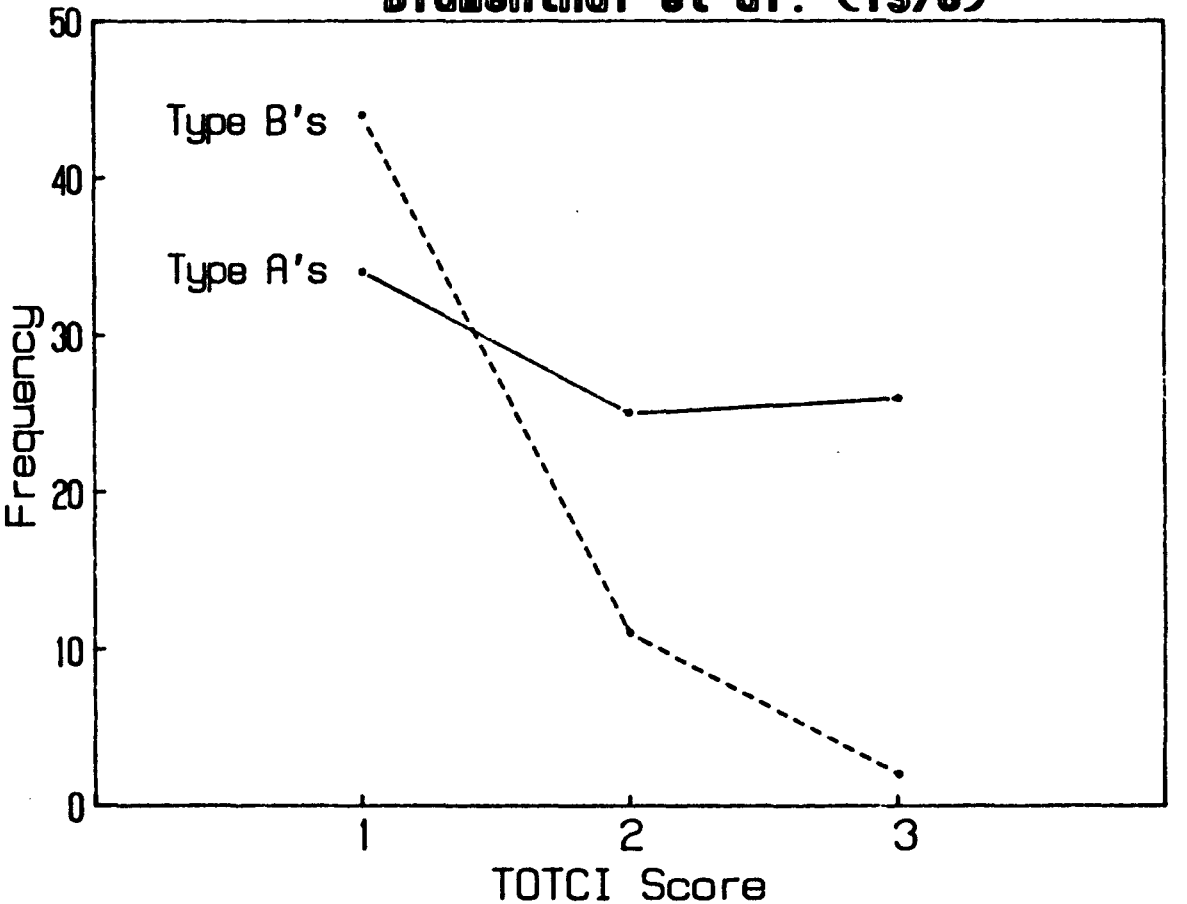
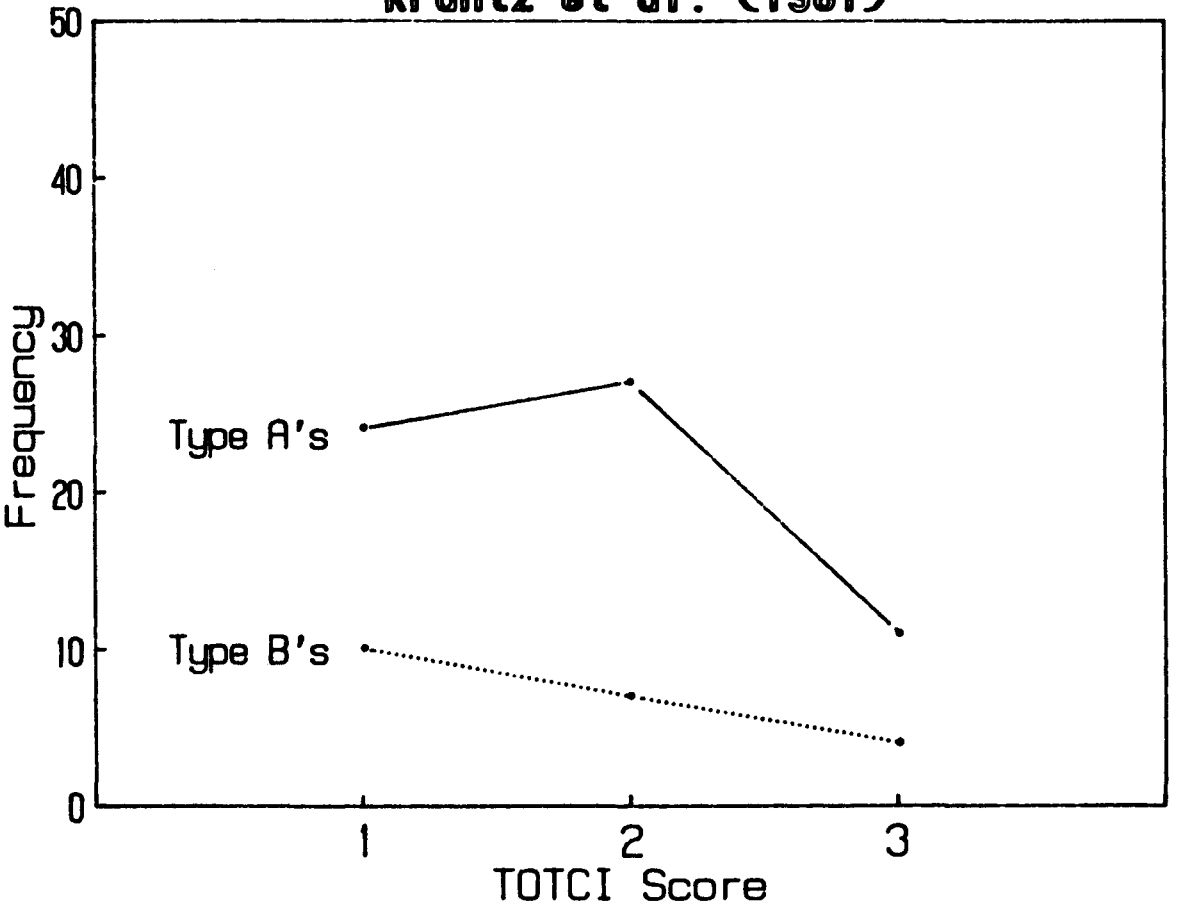


Figure 11(b)

Krantz et al. (1981)



reported sufficient data to reproduce frequency distributions using TOTCI as a disease criterion. The x-axis in each graph indicates the extent of disease by TOTCI score. As mentioned previously, the TOTCI score rates each major coronary artery vessel on a four point scale; 3 points for total occlusion, two points for a stenosis of 75%-99% decrease in luminal diameter, one point for a stenosis less than 75%, 0 points for non stenosis. The TOTCI is determined by taking the sum of the scores from all vessels that constitute the coronary artery system. Based on the TOTCI score, patients are grouped into three categories of mild < 3, moderate 3-6, and > 6 severe arteriosclerosis.

The y-axes in the graphs in Figure 11 indicate the number of patients that fall within each TOTCI disease category. At the bottom of each figure, the percentage of the study sample associated with each disease classification is given. The broken line labeled Type B's indicates the number of Type B's associated with each TOTCI score. The solid line indicates the frequency of Type A's for each TOTCI score. In the Blumenthal et al. (1978) study illustrated in Figure 11(b) more Type B's were present for those patients with a TOTCI score of 1 and more Type A's were present for TOTCI scores of 2 and 3. Thus, the cross-over pattern that occurs between a TOTCI score of 1 and 2 in the Blumenthal et al. (1978) study, indicates that a statistical association is present and that the location of the optimal cutpoint lies between a TOTCI score of 1 and 2. In the Blumenthal et al. study, 45 percent of the study

sample had a TOTCI of 1 and 55% had a score of 2 or 3. Thus, approximately equal numbers of actual positives and negatives were evenly distributed around the optimal cutpoint. For the Blumenthal et al. study, Table 1 in the column furthest to the right indicates that the results were statistically significant supporting the hypothesis that Type A behavior is associated with more severe disease.

The Krantz et al. (1981) study in Figure 11(b) presents a different picture. Across all levels of disease, a greater percentage of Type A's were present than Type B's. Thus, the statistical analysis of the Krantz et al. study indicates that there was no relationship between Type A behavior and arteriosclerosis. Comparing the two figures, 55 percent of the patients in the Blumenthal et al. (1978) study had a TOTCI score of 1 as compared with 41 percent in the Krantz et al. study. One reason why Blumenthal et al. may have found statistical significance and Krantz et al. didn't is that the optimal cutpoint is located towards the nondiseased end of the distribution of TOTCI scores. Blumenthal et al. may have found statistical significance because a larger percentage of Blumenthal et al.'s patients were located within the more nondiseased range. That is, there is a less extreme range of scores present in the Blumenthal et al. study.

Figure 10 illustrates the hypothetical study sample distributions that would account for why Blumenthal et al. (1978) found statistical significance and Krantz et al. (1981) did not. The portion of the graph

in Figure 10 shaded with horizontal lines represents the sampling frame of the Krantz et al. study. The portion of the graph shaded by vertical lines represents the range of the sample distribution of the Blumenthal et al. study. Notice that the Krantz et al. distribution is located to the right of the Blumenthal et al. distribution. The Krantz et al. sample distribution includes greater numbers of actual and predicted positives. Thus, the Krantz et al. study should have a higher percentage of Type A's than the Blumenthal et al. study if Figure 10 is an accurate representation of why there are differences between the Blumenthal et al. and Krantz et al. studies. In fact, this is the case. Krantz et al.'s study had 75% Type A's while Blumenthal et al.'s study had only 60% Type A's. Similarly, the Krantz et al. study had a higher proportion of diseased patients (55%) than the Blumenthal et al. study (45%). Note in Table 1, that the two other studies (Blumenthal, 1985 and Dembroski, 1985a) that reported nonsignificant findings using a TOTCI disease scoring system had percentages of Type A's and B's and numbers of diseased patients similar to those obtained in the Krantz et al. study.

Another disease criterion that has been used is to count the number of arteries that are have greater than 50% occlusion. The disease criteria is not as stringent as the TOTCI scoring method. That is, the TOTCI method classifies fewer subjects as diseased than the 50% occlusion method does. Because the optimal cutpoint appears to be

located at the end of the TOTCI continuum, one would suspect that the 50% occlusion decision rule is even further from the optimal cutpoint than a TOTCI score of 1. Thus, finding statistically significant results should be more difficult using the 50% occlusion rule than using TOTCI.

Figure 12 illustrates the three studies that published their data that used 50% occlusion as a disease criterion. The basic setup of the graphs is the same as in Figure 11. The x-axis indicates the number of arteries that are occluded by more than 50%. The y-axis indicates the number of patients. The differences in frequencies of Type A's across the disease categories is indicated by the solid lines. The broken lines indicate frequency of Type B's.

Among various levels of disease across all three studies (13 levels in all) only one study found more Type B's than Type A's (Frank et al., 1978 in the nondiseased artery group). The results are consistent with the hypothesis that all three distributions have sampled subjects that are mostly far to the right, on the diseased side, of the optimal cutpoint. Only the Frank et al. study found statistically significant results. Note the cross-over pattern in the Frank et al. study suggests that the optimal cutpoint is located very close to the negative exclusion rule.

For the Dimsdale et al. (1979a) study only 16 percent of the subjects had no diseased arteries. The Frank et al. (1978) study

Figure 12(a)

Frank et al. (1978)

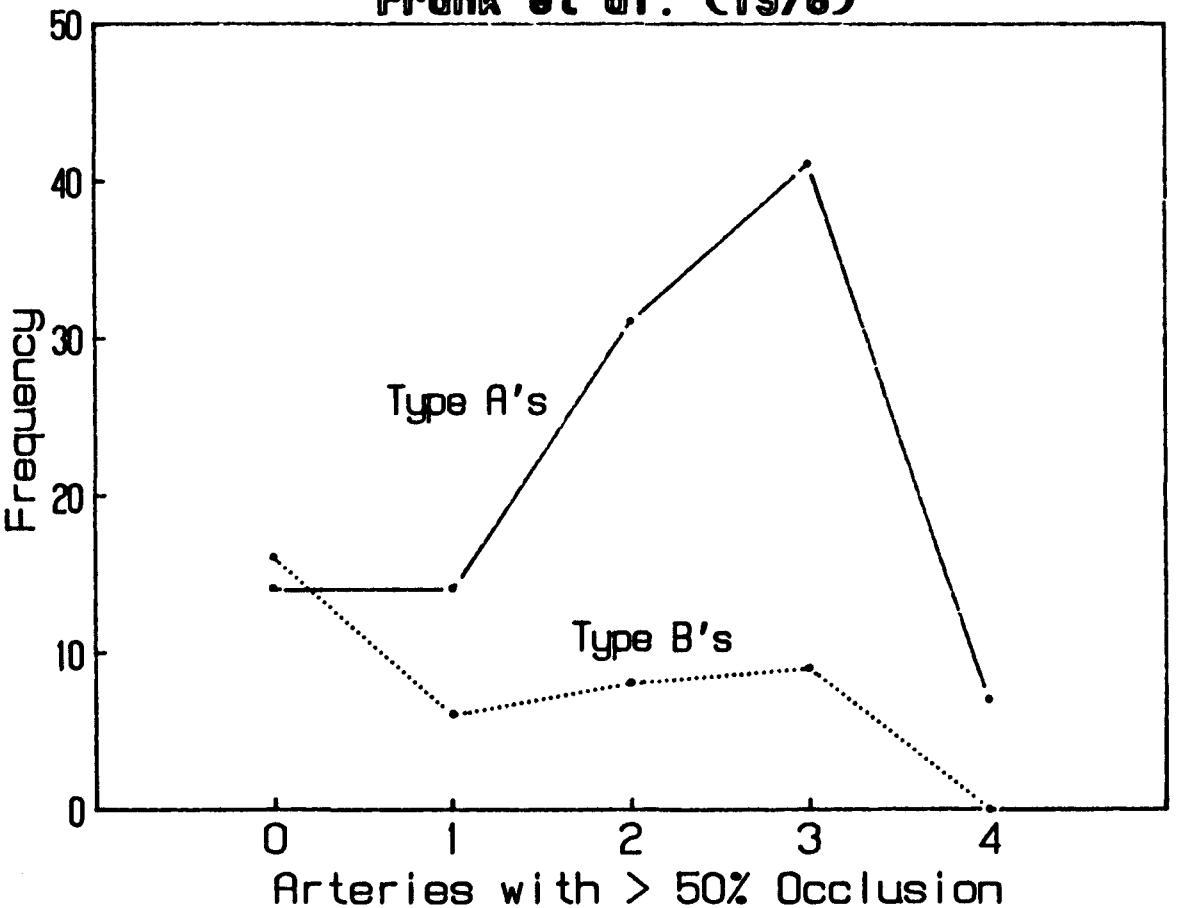


Figure 12(b)

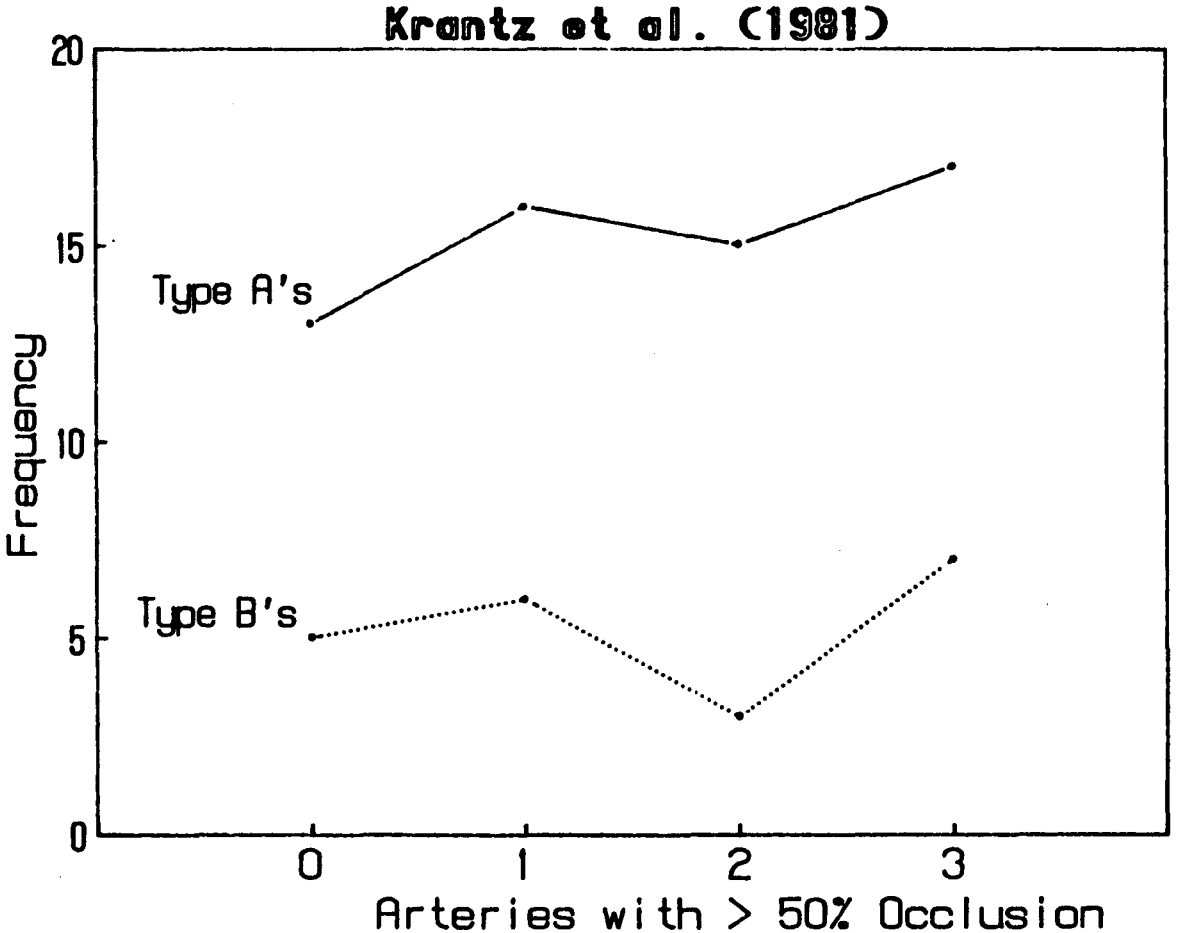
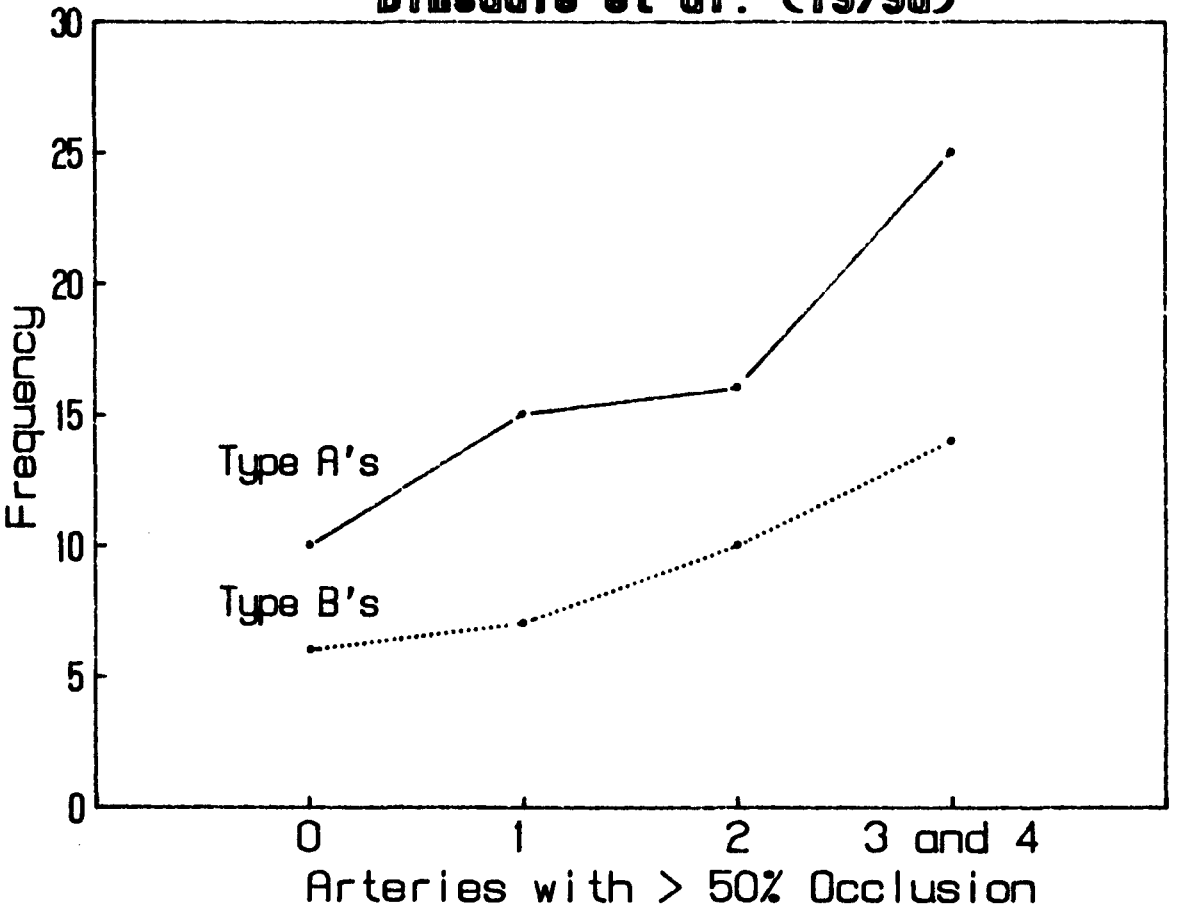

Krantz et al. (1981)
20
15
10
5
0
Frequency
Type A's
Type B's
0
1
2
3
Arteries with > 50% Occlusion

Figure 12(c)

Dimsdale et al. (1979a)



reported 20% nondiseased and the Krantz et al. (1981) study found 22% as nondiseased. There are higher percentages of diseased using the 50% occlusion method. Similarly, each study reported high percentages of Type A's (see Table 1).

Implications for medical research. Research in medicine can be viewed as developing diagnostic techniques in three stages. In the first stage, researchers experiment with the new technique only on patients they feel may receive substantial benefits at little risk. Research during this stage uses the case study approach and little effort is made to test whether the results support theories about the precursors of the disease that the diagnostic instrument attempts to measure.

In the second stage, the technique becomes more widely used because physicians become more confident using the techniques and begin to experiment. In this stage, researchers will use samples with wider ranges of disease because there is some uncertainty concerning when the technique should be used. Thus, researchers are more likely to include observations located around the optimal cutpoint so these studies are more likely to find statistically significant predictors of disease.

In the third stage, physicians become more experienced and can accurately determine who has the disease before confirming their results with the technique. Therefore, the diagnostic instrument is seldom used on nondiseased patients and so study samples becomes more diseased

because fewer nondiseased individuals are selected to be evaluated by the diagnostic test. Thus, more observations located near the optimal decision rule are excluded from the sample because the accuracy of the physician's judgments has increased. In the third stage, researchers may have trouble finding variables that significantly predict disease. This problem occurs because there will be too few actual negatives in the study sample.

The three stages of research as described above may have occurred in research on Type A behavior and arteriosclerosis. As physicians conducted more angiographies their experience developed and new research suggested ways for them to decrease the number of patients subjected to an angiography. In fact, physicians today are much less willing to do a coronary angiography than they were ten years ago (Pickering, 1985). Perhaps, as physicians developed more experience, fewer angiographies on nondiseased patients were performed resulting in research samples where almost the entire sample is diseased. Table 1 shows that recent studies have not found statistically significant results. Three of the four studies that found statistically significant results were published before 1981. The fourth study, published in 1982, was different in many respects from the other studies (e.g., Type A2's were excluded from analysis, only healthy normal individuals that were high on risk factors for disease were included in the sample). These differences may have produced the significant results.

A computer simulation of the effects of extremity of sample range and degree of range restriction on χ^2. It would seem that the most obvious effect of exclusion bias is restriction of range. As mentioned previously, range restriction has been the focus of most previous research on exclusion bias.

The final simulation of this paper compares the effects of range restriction with the effects that extremity of range has on statistical significance. Each study sample was operationalized as a section of a population as illustrated in Figure 3. A series of samples with different degrees of widths of ranges of values was simulated. The variance for each study sample was taken to be the standardized distance between the positive and negative exclusion rule. For example, the range in Figure 3 is 2 because that is the standardized distance between the exclusion rules. For this simulation, the variance was varied from 3 to 7 standard deviations between exclusion rules.

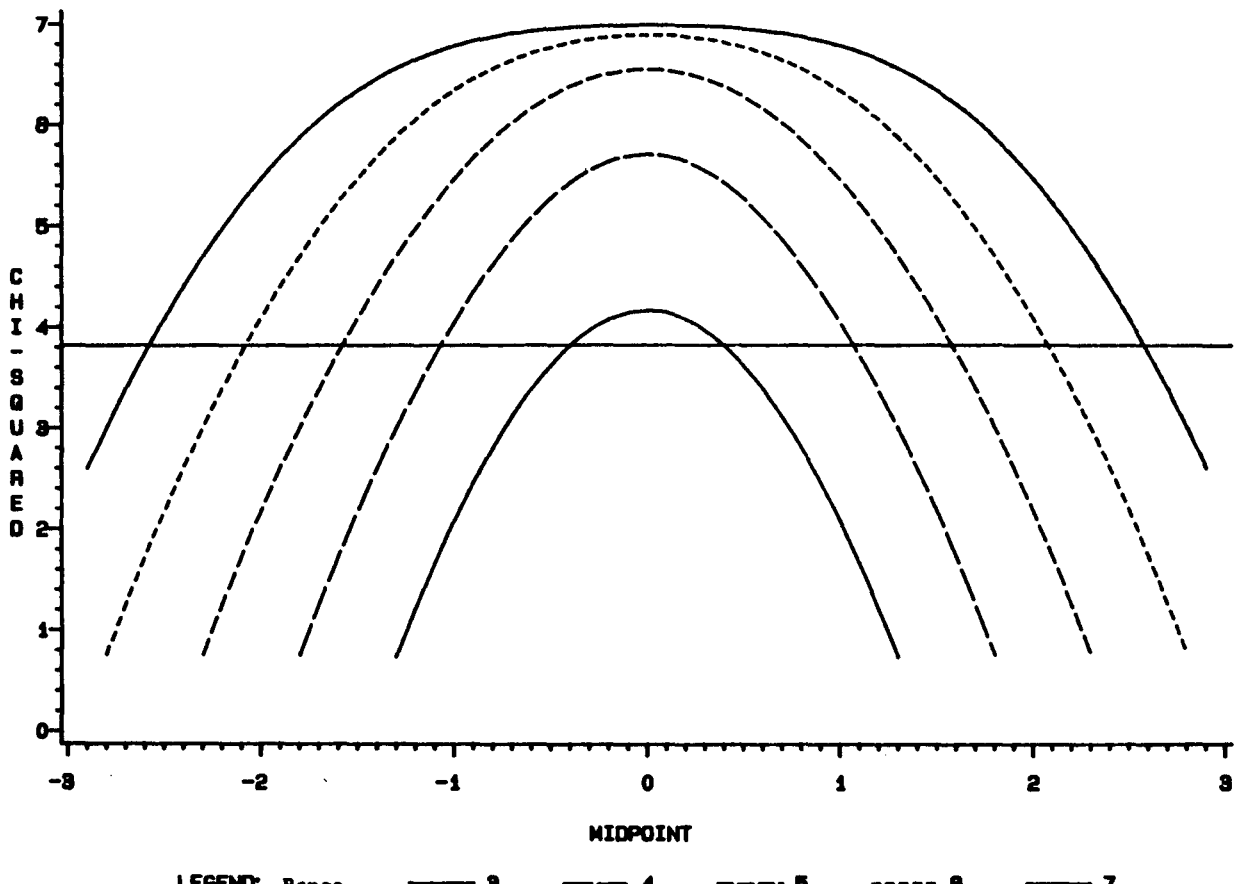
Extremity was operationalized as the distance between the median of the study sample and the optimal cutpoint. That is, the distance between the optimal cutpoint and the midpoint between the study sample's negative and positive exclusion rules. For example, the study sample midpoint in Figure 3 is at -1 because that is the midpoint between the exclusion rules located at -2 and 0. The optimal cutpoint in Figure 3 is located at 0. Thus, the distance between the study sample midpoint and the optimal cutpoint is -1 standard deviations. For the simulation, extremity of range was varied from -3 to 3.

Thus, the simulation calculated χ^2, values based on the ranges, and extremity of values included in the sample. For this simulation, $\underline{d}'$ was held at a constant value of .5 and sample size was fixed at 200. Note that the value of the optimal cutpoint does not change as the aforementioned parameters are varied. For this simulation, the cutpoint was always located at its most optimal point at zero.

Note that this simulation assumes complete truncation on the criterion variable. That is, complete truncation occurs at values associated with an exclusion rule. As mentioned previously, the assumption of complete truncation does not usually hold; incomplete truncation at a single value on the criterion is more likely to occur. Other variables, usually some of which are unknown, also determine what observations are selected into the sample (Olson & Becker, 1983). This topic is discussed in more detail in the next section section of this paper.

Figure 13 illustrates the results of the simulation. The horizontal line that runs across Figure 13 indicates that χ^2 values above the line are statistically significant and χ^2 values below the line are nonsignificant. The y-axis indicates the value of χ^2. Extremity of the range of values included in the sample is indicated by the values on the x-axis. Each curved line corresponds to a different width or range of a study sample. The range of sample values in standard deviations that is associated with each curved line is

FIGURE 13
THE RELATIONSHIP BETWEEN CHI-SQUARED AND THE
RANGE AND EXTREMITY OF VALUES INCLUDED IN THE SAMPLE



indicated on the legend at the bottom of the figure. For example, when the distance between the study sample midpoint and the optimal cutpoint is -1 and the range of values included in a sample is equal to 4 standard deviations, the χ^2 value would be approximately equal to 4.0.

The steepness of the curves in Figure 13 indicates the effects of extremity of range on χ^2. Note that for each curve χ^2 decreases rapidly as the midpoint of the study sample moves away from the optimal cutpoint. The effects are dramatic because the ratios of positives to negatives (both actual and predicted) becomes more disproportionate as the midpoint of the study sample moves away from the optimal cutpoint. Note for each curve, values associated with midpoints further from the optimal cutpoint are less significant. In all, only about half of the values presented in Figure 13 are statistically significant.

The differences in the χ^2 values associated with each curve gives the reader an impression of the influence that range restriction has on study findings. The curve associated with the highest χ^2 values in Figure 13 represents a range of six standard deviations. The next curve is associated with a range of values that are five standard deviations in width and so forth. Note that for wider ranges more χ^2 values are statistically significant. Thus, the range of values included in the sample is related to statistical significance as range restriction formulas suggest.

Note the range of values included in the study sample must be fairly large--3 standard deviations--before statistically significant results are possible. Note that with wider ranges of values the range of statistically significant samples greatly increases. For a range of values of 7 standard deviations, the midpoint of the study sample can be located at -2.6 and still be statistically significant. Note that for statistical significance to be obtained the range of values must include the optimal cutpoint. Thus, statistical significance is not obtained when the midpoint of the study sample is 2 and the range of values is 3 because the optimal cutpoint would not exist within the study sample.

Implications for research on Type A behavior and arteriosclerosis. Figure 13 clearly indicates that with a greater range of values in the sample, the effects of exclusion bias will be less severe. Unfortunately, to give angiographies to a large representative sample of individuals would be impractical and unethical. Thus, other techniques need to be used to control or assess the effects of exclusion bias. The next section discusses techniques and methods for controlling for exclusion bias.

METHODS FOR CONTROLLING FOR EXCLUSION BIAS

Subject Selection

There are a number of factors that may influence the relationship between Type A behavior and arteriosclerosis. Figure 4 can be used to illustrate how risk factors other than Type A behavior can change the degree of statistical significance found between Type A behavior and arteriosclerosis.

Age as described in the previously discussed hypothetical example was a variable that influenced the statistical association between Type A behavior and arteriosclerosis. Sampling frames that include higher values (e.g., increasingly older populations) are represented on the graph located at the bottom of Figure 4 where most of the observations are located to the right of c. Figure 4(a) illustrates a situation where the median point of the sample would be 1.5 standard deviations from the optimal cutpoint and the range of sample values would be approximately equal to .4. Thus, with a sample size of 200, according to Figure 13, statistical significance would not be obtained. Therefore, the age of the study sample can have a strong influence on whether a study finds statistically significant results.

Some examples of this type of age bias are obvious. For example, in medical research one would not expect to find an association between smoking and cancer in a population composed of college students because

the relationship is only expressed after an individual has been smoking for a lifetime. That is, the college students are too young to have smoked enough years for differences between smokers and nonsmokers to become apparent. Therefore, the sampling frame of college students is inappropriate; an older population would be more appropriate.

For research on the relationship between smoking and cancer, whether the person smokes or not and the number of years they have smoked influences the strength of the relationship between smoking and cancer. The years and intensity with which the patient smoked has been used as an estimate of the cumulative effects of smoking. Perhaps, Type A research would benefit from measuring Type A behavior as smoking is measured. Presumably, individuals who have displayed more intense Type A behavior for longer periods of time would be more vulnerable to arteriosclerosis than Type A's who are younger and have displayed less extreme behaviors. Test-retest Type A scores from multi-stage prospective studies could be used to estimate the cumulative effects of Type A behavior. Presumably, such measures of Type A-ness would be better predictors of arteriosclerosis and would in part control for problems of exclusion bias.

Researchers need to recognize that in extremely diseased populations, there are a number of ways that arteriosclerosis may have been produced. For example, people with high cholesterol scores may not need to be extreme Type A's in order to develop arteriosclerosis.

Similarly, older individuals may score lower on other risk factors such as cholesterol and still develop arteriosclerosis. Thus, researchers need to plan their study samples so that they include sufficient numbers of individuals whose arteriosclerosis is unexplained by other risk factors.

If Type A behavior predicts coronary artery disease after controlling for traditional risk factors, the percentage of A's in the real population for different age groups should vary. Presumably, more heart attacks (unexplained by family history, cholesterol etc.) would occur with Type A's after a number of years (i.e., between the age of 40 and 60) than in populations that used sampling frames that included younger or older patients. Researchers have selected subjects that are too young to have expressed the disease because of Type A behavior. The age ranges of most studies has been from 20 to 70 (see Table 1, column 4). Because Type A behavior alone cannot cause heart failure at age twenty, including subjects that are only twenty years old reduces the power of the overall test. One study attempted to control for age using analysis of covariance but this is inappropriate because this procedure assumes that Type A's and B's have equal ages. From their own published data, Krantz et al. (1981) did not meet this assumption.

Similarly, including subjects that are too old can influence study findings. In the sample of sixty year olds illustrated in Figure 4(c), a substantial number of observations have been excluded from the sample

because of the positive exclusion rule. In Figure 4(c) more Type B's than A's would be dying of heart disease. This occurs because all the Type A's in the sixty year old sample have already died of artery disease.

Have problems associated with a positive exclusion rule actually occurred in research on Type A behavior and arteriosclerosis? Recently, Williams, Barefoot, Haney, et al. (1986) combined and reanalyzed all the studies conducted at Duke University and Massachusetts General Hospital, and found that the correlation between Type A behavior and arteriosclerosis was only significant in younger men. More Type B's had arteriosclerosis for older samples. The results are consistent with the prediction that an exclusion rule exists within the samples. As in Figure 4, older Type A's may have been excluded from the sample because of previous heart problems. Similarly, Haynes, Feinleib, and Kannel (1980) found in the Framingham Heart Study that for men between the ages of 65 and 74 a greater percentage of Type B's had heart attacks.

Type A researchers must begin to test more explicit hypotheses concerning the relationship between Type A behavior and arteriosclerosis. That is, problems of positive and negative exclusion rules and the effects of variables such as age need to be assessed for each sample. Negatives findings can be due to exclusion bias if such problems are not taken into account in the study design.

Using Appropriate Comparison Groups

The appropriate comparison group is not the case-control approach that recommends finding individuals similar in all respects except for presence of disease. The case-control approach only enhances the potential that individuals are different on some unknown third variable that is related to the disease. Instead, different disease groups within the same hospital should be studied. For example, researchers could begin to examine differences between patients at various disease stages within the same hospital. For example, patients in the same hospital who are being treated for different levels of disease (e.g., high blood pressure, angina, and MI) could be examined for what predicts differences between these different levels of disease.

Presumably, individuals that are in the more diseased comparison groups would be exposed to the same risk factors longer or more intensely than individuals with less disease. Similarly, one would hypothesize that all of these diseased groups should have been exposed to higher levels of the risk factor than other nondiseased individuals served by the hospital.

An Information Synthesis Approach

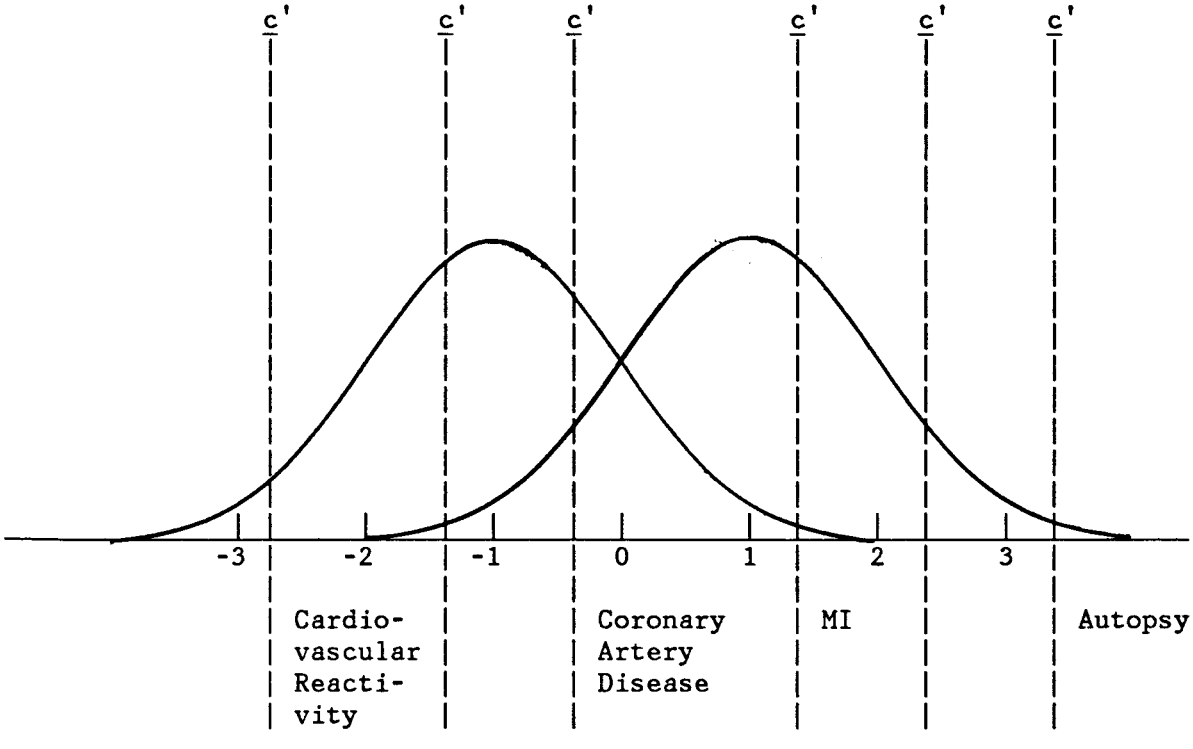
A similar approach could be applied to reviews of the previous literature. One can argue that the proportion of Type A's should be higher in more severely diseased populations (see Figure 13) if Type A behavior produces heart disease. The number of predicted positives and

negatives in the sample is influenced by the extremity of the range of values included in the study sample. There is a larger proportion of Type A's in samples that only include values from the extreme right end of the population (see Figure 4(c)). More Type B's are present in samples that come from observations on the left hand side of population distribution--as in Figure 4(a). Therefore, studies that sample more diseased populations should report higher percentages of Type A's. For example, studies of coronary occlusion at autopsy (e.g., see Friedman, Rosenman, Straus, Wurm, & Kositchek, 1968) should report higher percentages of Type A's than studies of heart attack recovery. Similarly, studies of coronary angiography should have higher percentages of Type A's than studies that sample healthy college students. A review of previous studies' percentages of Type A's at each decision point would provide a test of this hypothesis.

Figure 14 illustrates the continuum. The section furtherest to the left indicates a hypothetical distribution of disease for study samples of cardiovascular reactivity and Type A behavior. These studies usually consist of samples of healthy college students. Further to the right is the range of values included in studies of arteriosclerosis and Type A behavior. The study sample represents a more diseased group so the range of values is located further to the right of the cardiovascular reactivity studies. Even further to the right in Figure 14 are studies of myocardial infarcation. Still further to the right

Figure 14

Ranges of Sample Values Found in Type A Research.



are autopsy studies of the cardiovascular systems of subjects that die during prospective studies of heart disease (e.g., Friedman et al., 1968).

In addition, the percentages of Type A's should increase as disease severity increases. In more recent studies, the percentage of Type A's in the sample has increased (see Table 1) as disease severity in the sample increased, thereby supporting the hypothesis that Type A behavior is related to disease severity.

The Selector Variable Approach

Recent work by econometricians (see Heckman, 1980) has developed a method for modeling the selection process and obtaining an unbiased estimate of the correlation between two variables from a selected sample. The procedure requires the researcher to obtain an estimate of the λ^2 from a probit analysis between the predictor variable and a dichotomous "selector variable." The researcher must collect a representative sample of observations on the predictor variable. For example, a representative sample of the community that is served by the hospital where the coronary angiographies are conducted would be administered a measure of Type A behavior. For each observation, a selector variable is coded as 1 if the observation is included in the selected sample--patient has an angiography--and 0 if it is an excluded observation--a subject from the community sample that has not been administered an angiography. Thus, degree of disease would only be

measured by coronary angiography in the selected sample. Becker and Olson (1983) describe formulas that can be used for estimating the correlation coefficient between arteriosclerosis and Type A behavior in the community that is served by the hospital where angiographies are conducted.

Longitudinal Studies

Longitudinal studies that follow normal healthy individuals over time can be used to assess relationships between Type A behavior and arteriosclerosis. Krantz, Sanmarco, Selvester, and Matthews (1979) actually conducted such a study. In the Krantz et al. study, subjects were given two angiographies separated by an average of seventeen months. The study found that degree of Type A behavior predicted the extent of increase in arterial disease. Similarly, Corse, Manick, Cantwell, Giordani, & Matthews (1982) examined increases in coronary artery disease among survivors of an initial heart attack and found that coronary artery disease progressed more rapidly in Type A's.

However, the results of these studies may be interpreted in another way. Perhaps, the disease process progresses more rapidly, when more occlusion is present to begin with. If Type A's start with more disease then this could account for why Type A's become occluded at a more rapid rate than Type B's. Some research has suggested that Type A's delay seeking treatment longer so they are not scheduled for angiography until their disease becomes more severe (Matthews & Brunson,

1979; Matthews, Siegel, Kieller, Thompson, Varat, 1981; Weidner & Matthews, 1978). In the Krantz et al. (1981) study Type A's may have more disease because a longer time occurred between the time Type A's were scheduled for their second angiography. Krantz et al. should have controlled for initial disease severity and time between angiographies to rule out the possibility that the results were due to differences in Type A's initial levels of disease or length of time between angiographies.

The value of the Krantz et al. (1979) has been overlooked. The statistical power of this type of longitudinal study is much greater than the cross-sectional studies because exclusion bias is not as great a problem. The statistical power of longitudinal studies relies upon within subjects change so the degree of variability in change scores is what determines the power of the test. For this type of study, exclusion bias is the degree of sample attrition. Study attrition is not as much of a problem as unrepresentativeness in cross-sectional studies can be. Thus, statistical power is much greater in these type of studies. This may account for why Krantz et al. found statistically significant results with a small sample size.

Testing Study Power

The variance and range of scores of arteriosclerosis in the general population has been estimated from autopsy studies of accident victims (for a review see Pearson, 1983). Similarly, the variance and

range of Type A behavior scores in the general population can be estimated by surveys. These variance estimates of the population can be compared with the variances found in angiography studies. From these comparisons, the researcher can estimate the degree of exclusion bias present in his/her sample. By modifying the computer programs presented in Appendix A to conform to variance and range estimates researchers can estimate the degree of statistical power present in his/her sample.

Locating Optimal Cutpoints

Researchers that do find statistical significance should report their results in ways that permit other researchers to determine where the optimal cutpoints in their samples were located. This may include making an effort to measure the criterion variable as a continuous variable. Young et al's (1980) coronary occlusion index appears to be an appropriate way to measure coronary occlusion as a continuous variable.

Researchers that find nonsignificant results should report the range and variance of values in their studies so future researchers can determine if the negative findings were due to exclusion of the optimal cutpoint from the study sample.

SUMMARY

In this paper, an explicit quantitative definition of one type of bias (exclusion bias) was introduced. Computer simulations were used to illustrate how statistical power is influenced by exclusion bias. In addition, several suggestions were made for how problems of exclusion bias can be dealt with.

Suggestions for Future Research

This paper raises several questions concerned with the value judgments of applied versus theoretical researchers. For example, many medical researchers recommend only using diagnostic tests that are the "gold standard" and the most valid indicators of disease (e.g., Prorok, 1979). However, highly valid indicators are typically invasive, are accompanied by risk to the patient and thus can only be ethically used for individuals for whom there is a strong suspicion of disease. The value of less accurate procedures that can be used to sample the whole population is usually not considered. Thus, there is a trade-off between diagnostic accuracy and exclusion bias, in most applied research situations. Researchers must begin to recognize these tradeoffs and begin to consider whether the costs of using extremely accurate but more invasive diagnostic devices are really worthwhile.

In many areas of research, a laboratory study is conducted to test a hypothesis and then a follow-up study in a field setting is conducted to determine if the results of the laboratory study are "clinically

relevant." The two study samples may be at opposite tails of the distribution to which the researcher wants to generalize. Therefore, exclusion bias may bias both studies towards accepting the null hypothesis. Negative results should therefore should be considered in terms of the degree of exclusion bias present within the study sample and not only whether the range of values includes a clinically relevant sample.

For example, study samples obtained from more severely depressed populations (e.g., inpatient units as opposed to outpatient units) will encounter negative findings because exclusion bias may be more severe. The researchers may falsely conclude that although mild depression appears to be related to the predictor variable, the variable is not "clinically relevant" for severely depressed populations.

The quantitative approach taken in this paper could be applied to other types of biases. This may give researchers a more quantitative and systematic description of how biases influence the results of their research.

Problems of exclusion bias may lead to nonsignificant findings in research on the relationship between job performance and selection tests. For example, as a personnel selection test becomes more accurate (or more highly correlated with the selection practices) fewer actual negatives will be included in the study sample. Note that the goal of personnel researchers is to design a test where the exclusion rule (who

is hired) is located at the same optimal cutpoint (the point where the test maximally discriminates between those hired and not hired). Note under these circumstances the test is guaranteed not to be significantly related to any measures of on the job performance. Thus, as test designers begin to design their tests to optimally discriminate at the number of positions available nonsignificant findings should occur between the selection device and job performance. In other words, a valid hiring device has a much different optimal cutpoint than a valid measure of job performance for those applicants that are hired (see Alexander, Barrett, & Doverspike, 1983). Note that range restriction formulas do not correct for this problem.

REFERENCES

Alexander, R. A., Barrett, G. V., & Doverspike, D. (1983). An explication of the selection ratio and its relationship to the hiring rate. Journal of Applied Psychology, 68(2), 342-344.

American Psychological Association Division of Industrial-Organizational Psychology. (1980). Principles for the use of validation and use of personnel selection procedures: Second Edition. Dayton, OH: Author

Brand, R. J. (1977). An examination of the association between A-B behavior and coronary heart incidence. In T. M. Dembroski (Ed.), Proceedings of the forum on coronary-prone behavior. Washington, D. C.; Department of Health Education, and Welfare, 29-45.

Brand, R. J., Rosenman, R. H., Sholtz, R. I., & Friedman, M. (1976). Multivariate prediction of coronary heart disease in the Western Collaborative Group Study compared to the findings of the Framingham Study. Circulation, 53, 348-355.

Brewer, J. K. & Hills, J. R. (1969). Univariate selection: the effects of size of correlation, degree of skew, and degree of restriction. Psychometrika, 26, 347-372.

Bristow, J. D., Buchell, H. B., Campbell, R. W., et al. (1977). Report of the ad hoc committee on the indications for coronary angiography. Circulation, 55, 969A-74A.

Christensen-Szlankski, J. J. J., Diehr, P. H., Bushyhead, J. B., & Wood, R. W. (1978). Two studies of good clinical judgment. Journal of Medical Decision Making, 2(3), 275-283.

Cohen, J. (1977). Statistical power analysis for the behavioral sciences. New York: Academic.

Conti, C. R. (1977). Coronary arteriography. Circulation, 55, 227-237.

Corse, C. D., Manick, S. B., Cantwell, J. D., Giordani, B., & Matthews, K. A. (1982, Nov.). Coronary-prone behavior pattern and cardiovascular response in persons with and without coronary heart disease. Psychosomatic Medicine, 44(5), 449-459.

Costa, P. T., Zonderman, A. B., Engel, B. T., Baile, W. F., Brimlow, D. L., & Brinker, J. (1985). The relation of chest pain symptoms to coronary angiographic findings of coronary artery disease stenosis and neuroticism. Psychosomatic Medicine, 47(3), 285-293.

Cureton, E. E. (1957). Recipe for a cookbook. Psychological Bulletin, 52, 494-497.

Davidoff, M. D., & Goheen, H. W. (1953). A table for the rapid determination of the tetrachoric correlation coefficient, Psychometrika, 18, 115-121.

Friedman, M., & Rosenman, R. H. (1974). Type A behavior and your heart. New York: Knopf.

Ghiselli, J. P., Campbell, J. P., & Zedeck, S. (1981). Measurement theory for the behavioral sciences. San Francisco: Freeman.

Glass, D. C. (1977). Behavior patterns, stress, and coronary disease. Hillsdale, NJ: Erlbaum.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. Educational Researcher, 5, 3-8.

Greener, J. M. & Osburn, H. G. (1979). An empirical study of the accuracy of correlations for restriction in range due to explicit selection. Applied Psychological Measurement, 3, 31-41.

Greener, J. M. & Osburn, H. G. (1980). Accuracy of corrections for for restriction in range due to explicit selection in heteroscedastic and nonlinear distributions. Educational and Psychological Measurement, 40, 337-348.

Gulliksen, H. (1950). Theory of mental tests. New York: Wiley.

Haynes, S. G., Feinleib, N., & Kannel, W. B. (1980). The relationship of psychosocial factors to coronary heart disease in the Framingham study, American Journal of Epidemiology, 111(1), 37-58.

Heckman, J. J. (1980). Sample selection bias as a specification error. In E. Stromsdorfer and G. Farkas (Eds.), Evaluation studies review annual (Vol. 5). Beverely Hills, Calif.: Sage Publications.

Hlatky, M. A., Lee, K. L., Botvinick, E. H., & Brundage, B. H. (1983, Oct.). Diagnostic test use in different practice settings: A controlled comparison. Archives of Internal Medicine, 143, 1886-1889.

Jenkins, C. P. (1976). Recent evidence supporting psychological and social risk factors for coronary disease, New England Journal of Medicine, 294, 987-994, 1033-1038.

Jenkins, C. P., Zyzanski, S. J., & Rosenman, (1971). Association or coronary-prone behavior scores with reoccurrence of coronary heart disease. Journal of Chronic Disease, 24, 601-611.

Kleinbaum, D. G., Morgenstern, H., & Kupper, L. L. (1984). Selection bias in epidemiological studies. American Journal of Epidemiology, 113(4), 452-463.

Lee, R. & Foley, P. P. (1986). Is the validity of a test constant throughout the test score range? Journal of Applied Psychology, 71(4), 641-644.

Linn, R. L., & Dunbar, S. B. (1982). Predictive validity of admissions measures: Corrections for selection on several variables. Journal of College Student Personnel, 23, 222-226.

Lord, F. M., & Novick, M. R. (1968). Statistical theories of mental test scores. Reading, Mass.: Addision-Wesley.

Matthews, K. A. (1982). Psychological perspectives on the Type A Behavior pattern, Psychological Bulletin, 91(2), 293-323.

Matthews, K. A., & Brunson B. I. (1979). Allocation of attention and the Type A coronary-prone behavior pattern. Journal of Personality and Social Psychology, 37, 2081-2090.

Matthews, K. A., & Haynes, S. G. (1986, June). Type A behavior pattern and coronary disease risk: Update and critical evaluation. American Journal of Epidemiology, 123(6), 923-960.

Matthews, K. A., Siegel, J. M., Kieller, L. H., Thompson, M., Varat, M. (1981). Determinants of decisions to seek medical treatment by myocardial infarcation victims. Paper presented at the American Psychological Association, Los Angeles.

Metz, C. E. (1978, Oct.). Basic principles of ROC analysis. Seminars in Nuclear Medicine, 8(4), 283-298.

Olson, C. A., & Becker, B. E. (1983). A proposed technique for the treatment of restriction of range in selection validation. Psychological Bulletin, 93(1), 137-148.

Pearson, K. (1903). Mathematical contributions to the theory of evolution-XI. On the influence of natural selection on the variability and correlation of organs. Royal Society Philosophical Transaction (Series A, 200), 1-66.

Pickering, T. G. (1985). Should studies of patients undergoing coronary angiography be used to evaluate the role of behavioral risk factors for coronary heart disease? Journal of Behavioral Medicine, 8(3), 203-213.

Prorok, P. (1979). Evaluation of the impact of screening on the natural history of disease. In D. R. Brodie (Ed.) Screening and early detection of cancer: Consensus development conference proceedings. U. S. Department of Health, Education, and Welfare (pp. 25-34).

Raiffa, H. (1968). Decision Analysis. Reading, Mass.: Addision-Wesley.

The Review Panel on Coronary-prone Behavior and Coronary Heart Disease. 1981, June). Coronary-prone behavior and coronary heart disease: A critical review. Circulation, 63(6), 1199-1215.

Rissanen, V. T. (1975). Coronary atherosclerosis in cases of coronary death as compared with that occurring in the population. Annuals of Clinical Research, 7, 412-25.

Rorer, L. G., Hoffman, P. J., LaForge, G. E., Hsieh, K. (1966a). Optimum cutting scores to discriminate groups of unequal size and variance, Journal of Applied Psychology, 50(2), 153-164.

Rorer, L. G., Hoffman, P. J., Hsieh, K. (1966b). Utilities as base-rate multipliers in the determination of optimum cutting scores for the discrimination of groups of unequal size and variance. Journal of Applied Psychology, 50(5), 364-368.

Rosenman, R. H. (1978). The interview method of assessment of the coronary-prone behavior pattern. In T. M. Dembroski, S. M. Weiss, J. L. Shields, S. G. Haynes & M. Feinleib (Eds.) Coronary-prone behavior. New York: Springer-Verlag.

Rosenman, R. H., Brand, R. J., Jenkins, C. D., Friedman, M., Straus, R., & Wurm, M. (1975). Coronary heart disease in the Western Collaborative Group Study: Final follow-up experience of 8.5 years. Journal of the American Medical Association, 233, 872-877.

Rosenman, R. H., Friedman, M., Straus, R., Wurm, M., Jenkins, D. C., & Messinger, H. B. (1966, Jan). Coronary heart disease in the Western Collaborative Group Study: A follow-up experience of two years. JAMA, 195(2), 130-136.

Rosenman, R. H., Friedman, M., Straus, R., Jenkins, C. D., Zyzanski, S. J., & Wurm, M. (1970). Coronary heart disease in the Western Collaborative Group Study: A follow-up experience of 4.5 years. Journal of Chronic Disease, 23, 173-190.

Rosenman, R. H., Friedman, M., Straus, R., Wurm, M., Kositchek, R., Hahn, W., & Werthessen, N. T. (1964). A predictive study of coronary disease: The Western Collaborative Group Study, JAMA, 189(1), 103-110.

Rydberg, S. (1963). Bias in prediction: on correction methods. Stockholm: Almquist and Wiksells.

SAS Institute Inc. (1985). SAS/GRAPH user's guide version 5 edition. Cary, NC: SAS Institute Inc.

Schmidt, F. L., Hunter, J. E., McKenzie, R. C., & Muldrow, T. W. (1979). Impacts of valid selection procedures on work-force productivity, Journal of Applied Psychology, 64(6), 609-626.

Smith, T. W., & Anderson, N. B. (1986). Models of personality and disease: An interactional approach to Type A Behavior and cardiovascular risk, Journal of Personality and Social Psychology, 50(6), 1166-1173.

Steel, R. G. D., & Torrie, J. H. (1960). Principles and procedures of statistics. 2nd ed. New York: McGraw-Hill.

Swets, J. A., Pickett, R. M., Whitehead, S. F., Getty, D. J., Schnur, J. A., Swets, J. B., & Freeman, B. A. (1979, Aug). Assessment of diagnostic technologies. Science, 205 No. 4407(24), 753-759.

Thorndike, R. L. (1947). Research problems and techniques (Report No. 3). Washington DC: AAF aviation Psychology Program Research Reports, U. S. Government Printing Office.

Thorndike, R. L. (1949). Personnel selection: Test and measurement techniques. New York: Wiley.

Weidner, G., Matthews, K. A. (1978). Reported physical symptoms elicited by unpredictable events and the Type A coronary-prone behavior pattern. Journal of Personality and Social Psychology, 36, 1213-1220.

Weinstein, M. C., & Fineberg, H. V. (1980). Clinical decision analysis. Philadelphia, PA: Saunders.

Williams, Jr., R. B. (1979). Physiological mechanisms underlying the association between psychosocial factors and coronary disease. In W. D. Gentry & R. B. Williams (Eds.) Psychological aspects of myocardial infarcation and coronary care, (Chp. 3). St. Louis: C. V. Mosley Co.

APPENDIX A

Appendix A

SAS/GRAPH Computer Programs to Generate Figures

Figure 5

```
GOPTIONS;
DATA FIVE;
N = 200;
 DO d = .3 TO .7 BY .1;
 DO C = -2.5 TO 2.5 BY .1;
  LABEL C = 'LOCATION OF CUTPOINT';
  LABEL d = EFFECT SIZE;
  LABEL N = SAMPLE SIZE;
ZTN = C + d/2;
TN = PROBNORM(ZTN);
TN= TN * N/2;
 LABEL TN = FREQ OF TRUE NEGATIVES;
FP = N/2 - TN;
 LABEL FP = FREQ OF FALSE POSITIVES;
ZFN = ZTN - d;
 LABEL ZFN = Z OF FALSE NEGATIVES;
FN = PROBNORM(ZFN);
FN = FN * N/2;
 LABEL FN = FREQ OF FALSE NEGATIVES;
TP = N/2 - FN;
ONE = (TN * TP) - (FP * FN);
TWO = (TN + FP) * (FN + TP) * (TN + FN) * (FP + TP);
 LABEL TP = FREQ OF TRUE POSITIVES;
 X = N *((ABS(ONE) - N/2)**2)/TWO;
 LABEL X = CHI-SQUARED;
OUTPUT;END;END;
PROC GPLOT;
TITLE1 .F= NONE .H=2 FIGURE 5;
TITLE2 .F= NONE .H=2 THE RELATIONSHIP BETWEEN CHI-SQUARED;
TITLE3 .F=NONE .H=2 AND THE CUTPOINT FOR VARIOUS EFFECT SIZES;
PLOT X*C=D/VREF=3.816 CTEXT=BLACK HAXIS= -3 TO 3 BY 1;
SYMBOL1 I=SPLINE C=BLACK L=1;
SYMBOL2 I=SPLINE C=BLACK L=4;
SYMBOL3 I=SPLINE C=BLACK L=3;
SYMBOL4 I=SPLINE C=BLACK L=2;
SYMBOL5 I=SPLINE C=BLACK L=1;
ENDSAS;
```

Figure 6

```
GOPTIONS;
DATA SIX;
D = 0.5;
 DO N = 50 TO 250 BY 50;
 DO C = -2.0 TO 2.0 BY .1;
  LABEL C = CUTPOINT;
  LABEL D = d'';
  LABEL N = SAMPLE SIZE;
ZTN = C + D/2;
TN = PROBNORM(ZTN);
TN = TN * N/2;
 LABEL TN = FREQ OF TRUE NEGATIVES;
FP = N/2 - TN;
 LABEL FP = FREQ OF FALSE POSITIVES;
ZFN = ZTN - D;
 LABEL ZFN = Z OF FALSE NEGATIVES;
FN = PROBNORM(ZFN);
FN = FN * N/2;
 LABEL FN = FREQ OF FALSE NEGATIVES;
TP = N/2 - FN;
ONE = (TN * TP) - (FP * FN);
TWO = (TN + FP) * (FN + TP) * (TN + FN) * (FP + TP);
 LABEL TP = FREQ OF TRUE POSITIVES;
 X = N *((ABS(ONE) - N/2)**2)/TWO;
 LABEL X = CHI-SQUARED;
OUTPUT;END;END;
PROC GPLOT;
TITLE1 .F=NONE .H=2 FIGURE 6;
TITLE2 .F=NONE .H=2 CHI-SQUARED BY CUTPOINT;
TITLE3 .H=2 FOR SEVERAL DIFFERENT SAMPLE SIZES;
PLOT X*C=N/ CTEXT= BLACK ;
SYMBOL1 I=SPLINE C=BLACK L=1;
SYMBOL2 I=SPLINE C=BLACK L=4;
SYMBOL3 I=SPLINE C=BLACK L=3;
SYMBOL4 I=SPLINE C=BLACK L=2;
SYMBOL5 I=SPLINE C=BLACK L=1;
ENDSAS;
```

Figure 8

```
GOPTIONS ;
DATA EIGHT;
 DO  D = .3 to .7 BY .1;
 DO RATIO = 1 TO 9;
   n = 200 / (1 + RATIO);
   C = (1 / RATIO) - 1;
  LABEL C = CUTPOINT;
  LABEL D = d'';
  LABEL N = SAMPLE SIZE;
  LABEL RATIO = RATIO OF POSITIVES TO NEGATIVES;
ZTN = C + D/2;
TN = PROBNORM(ZTN);
TN= TN * N;
 LABEL TN = FREQ OF TRUE NEGATIVES;
FP = N - TN;
 LABEL FP = FREQ OF FALSE POSITIVES;
ZFN = ZTN - D;
 LABEL ZFN = Z OF FALSE NEGATIVES;
FN = PROBNORM(ZFN);
FN = FN * N * RATIO;
 LABEL FN = FREQ OF FALSE NEGATIVES;
TP = (N * RATIO) - FN;
ONE = (TN * TP) - (FP * FN);
TWO = (TN + FP) * (FN + TP) * (TN + FN) * (FP + TP);
 LABEL TP = FREQ OF TRUE POSITIVES;
 X = 200 *((ABS(ONE) - 200/2)**2)/TWO;
 LABEL X = CHI-SQUARED;
OUTPUT;END;end;
PROC gpLOT;
TITLE1 .F= NONE .H=2 FIGURE 8;
TITLE2 .F= NONE .H=2 CHI-SQUARED BY RATIO OF PREDICTED POSITIVES TO;
TITLE3 .H=2 NEGATIVES FOR VARIOUS EFFECT SIZES ;
PLOT X*RATIO=D/
  CTEXT=BLACK VREF=3.816 HAXIS=1 to 9 BY 1;
SYMBOL1 I=SPLINE C=BLACK L=1;
SYMBOL2 I=SPLINE C=BLACK L=4;
SYMBOL3 I=SPLINE C=BLACK L=3;
SYMBOL4 I=SPLINE C=BLACK L=2;
SYMBOL5 I=SPLINE C=BLACK L=1;
ENDSAS;
```

Figure 9

```
DATA NINE;
  D = .5;
 DO RATIO = 1 TO 9;
 DO N = 50 TO 250 BY 50;
   NSIZE = N / (1 + RATIO);
   C = (1 / RATIO) - 1;
  LABEL C = CUTPOINT;
  LABEL D = d'';
  LABEL N = SAMPLE SIZE;
  LABEL RATIO = RATIO OF POSITIVES TO NEGATIVES;
ZTN = C + D/2;
TN = PROBNORM(ZTN);
TN= TN * NSIZE;
 LABEL TN = FREQ OF TRUE NEGATIVES;
FP = NSIZE - TN;
 LABEL FP = FREQ OF FALSE POSITIVES;
ZFN = ZTN - D;
 LABEL ZFN = Z OF FALSE NEGATIVES;
FN = PROBNORM(ZFN);
FN = FN * NSIZE * RATIO;
 LABEL FN = FREQ OF FALSE NEGATIVES;
TP = (NSIZE * RATIO) - FN;
ONE = (TN * TP) - (FP * FN);
TWO = (TN + FP) * (FN + TP) * (TN + FN) * (FP + TP);
 LABEL TP = FREQ OF TRUE POSITIVES;
 X = n *((ABS(ONE) - N/2)**2)/TWO;
 LABEL X = CHI-SQUARED;
OUTPUT;END;END;
PROC GPLOT;
TITLE1 .F= NONE .H=2 FIGURE 9;
TITLE2 .F= NONE .H=2 CHI-SQUARED BY RATIO OF PREDICTED POSITIVES TO;
TITLE3 .H=2 NEGATIVES FOR VARIOUS SAMPLE SIZES ;
PLOT X*RATIO=N/
  CTEXT=BLACK VREF=3.816 HAXIS=1 to 9 by 1;
SYMBOL1 I=SPLINE C=BLACK L=1;
SYMBOL2 I=SPLINE C=BLACK L=4;
SYMBOL3 I=SPLINE C=BLACK L=3;
SYMBOL4 I=SPLINE C=BLACK L=2;
SYMBOL5 I=SPLINE C=BLACK L=1;
ENDSAS;
```

Figure 13

```
DATA THIRTEEN;
N = 200 ;
D = .5;
 DO CSIZE = 3 to 7 BY 1;
 DO C = .2 TO 7 BY .2;
  LABEL C = UPPER LIMIT;
  C2 = X - CSIZE;
  LABEL C2 = LOWER LIMIT;
  LABEL D = D'';
  LABEL N = SAMPLE SIZE;
CUTPOINT = 0;
ZTN = CUTPOINT + D/2;
ZC = C + D/2;
ZT = C2 +D/2;
TN = PROBNORM(ZTN) - PROBNORM(ZT);
 LABEL TN = FREQ OF TRUE NEGATIVES;
FP = 1 - PROBNORM(ZTN) - (1 - PROBNORM(ZC));
ZF = ZT - D;
ZFT = ZF + CSIZE;
ZFN = ZTN - D;
TP = 1 - PROBNORM(ZFN) - (1 - PROBNORM(ZFT));
 LABEL FP = FREQ OF FALSE POSITIVES;
 LABEL ZFN = Z OF FALSE NEGATIVES;
FN = PROBNORM(ZFN) - PROBNORM(ZF);
 LABEL FN = FREQ OF FALSE NEGATIVES;
TOTAL = FP + TP + TN + FN;
FP = N * FP / TOTAL;
FN = N * FN / TOTAL;
TP = N * TP / TOTAL;
TN = N * TN / TOTAL;
ONE = (TN * TP) - (FP * FN);
TWO = (TN + FP) * (FN + TP) * (TN + FN) * (FP + TP);
 LABEL TP = FREQ OF TRUE POSITIVES;
 X = N *((ABS(ONE) - N/2)**2)/TWO;
 LABEL X = CHI-SQUARED;
IF X GE 50 THEN X = 0;
IF C2 GE 0 THEN X = 0;
IF TWO LT .5 THEN X = .;
MIDPOINT = (C + C2)/2;
OUTPUT;END;end;
PROC GPLOT ;
TITLE1 .F= NONE .H=2 FIGURE 13;
TITLE2 .F= NONE .H=2 THE RELATIONSHIP BETWEEN CHI-SQUARED AND THE ;
TITLE3 .H=2 RANGE AND EXTREMITY OF VALUES INCLUDED IN THE SAMPLE;
PLOT X*MIDPOINT=CSIZE/VREF=3.816  HAXIS= -3 TO 3 BY 1;
```

```
SYMBOL1 I=SPLINE C=RED L=1;
SYMBOL2 I=SPLINE C=RED L=4;
SYMBOL3 I=SPLINE C=RED L=3;
SYMBOL4 I=SPLINE C=RED L=2;
SYMBOL5 I=SPLINE C=RED L=1;
ENDSAS;
```

APPENDIX B

Appendix B

Arrowood, M. E., Uhrich, Jr., Gomillion, C., Popio, K. A., & Raft, D., (1982, Mar). New markers of coronary prone behavior in a rural population. Psychosomatic Medicine, 44(1), 119.

Bass, C. (1984). Type A behavior in patients with chest pain: test-retest reliability and psychometric correlates of bortner scale. Journal of Psychosomatic Research, 28(4), 289-300.

Bass, C. & Wade, C. (1982). Type A behavior: not specifically pathogenic? The Lancet, 1147-1150.

Blumenthal, J. A., Williams, Jr., R. B., Kong, Y., Schanburg, S. M., & Thompson, L. W., (1978, Oct). Type A Behavior Pattern and coronary atherosclerosis. Circulation, 58(4), 634-639.

Blumenthal, J. A., Herman, S., O'Toole, L. C., Haney, T. L., Williams, R. B., Barefoot, J. C. (1985). Development of a brief self-report measure of the Type A (coronary prone) Behavior Pattern, Journal of Psychosomatic Research, 29(3), 265-274.

Cottier, C., Adler, R., Vorchauf, H., Gerber, R., Heller, T., & Hurney, C. (1983, June). Pressured pattern or Type A behavior in patients peripheral arteriovascular disease: controlled retrospective exploratory study. Psychosomatic Medicine, 45(3), 187-193.

Dembroski, T. M., MacDougall, Williams, R. B., Haney, T. L., Blumenthal, J. A. (1985). Components of Type A, Hostility, and Anger-In: Relationship to Angiographic Findings. Psychosomatic Medicine, 47(3), 219-233.

Dimsdale, J. E., Hackett, T. P., Hutter, A. M., Block, P. C. (1980). The risk of Type A mediated coronary artery disease in different populations. Psychosomatic Medicine, 42(1), 55-62.

Dimsdale, J. E., Hackett, T. P., Hutter, A. M., Block, P. C., Catanzano, D. (1978). Type A personality and extent of coronary atherosclerosis. The American Journal of Cardiology, 42, 583-586.

Dimsdale, J. E., Hackett, T. P., Hutter, A. M., Block, D. L., Catanzano, D., & White, P. J. (1979a). Type A behavior and angiographic findings. Journal of Psychosomatic Research, 23, 273-276.

Dimsdale, J. E., Hackett, T. P., Hutter, A. M., Block, D. L., Catanzano, D., & White, P. J. (1979b). The relationship between diverse measures for type A personality and coronary angiographic findings. Journal of Psychosomatic Research, 23, 289-293.

Frank, K. A., Heller, S. S., Kornfeld, D. S., Sporn, A. A., Weiss, M. B. (1978). Type A Behavior Pattern and coronary angiographic findings, JAMA, 240(8), 761-763.

Friedman, M., Rosenman, R. H., Straus, R., Wurm, M., & Kositchek, R. (1968). The relationship of Behavior Pattern A to the state of the coronary vasculature: A study of fifty-one autopsy subjects. American Journal of Medicine, 44, 525-537.

Kahn, J. P., Kornfeld, D. S., Blood, D. K., Lynn, R. B., Heller, S. S., & Frank, K. A. (1982, Nov.). Type A Behavior and the Thallium Stress Test. Psychosomatic Medicine, 44(5), 431-436.

Kornitzer, M., Magotteau, V., Pegre, C. et al. (1982). Angiographic findings and the Type A pattern assessed by means of the Bortner scale. Journal of Psychosomatic Medicine, 5, 313-320.

Krantz, D. S., Sanmarco, M. I., Selvester, R. H., & Matthews, K. A. (1979). Psychological correlated of progression of atherosclerosis in men. Psychosomatic Medicine, 41(6), 469-475.

Krantz, D. S., Schaeffer, M. A., Davia, J. E., Dembroski, T. M., MacDougall, J. M., & Schaeffer, R. T. (1981). Extent of coronary atherosclerosis, Type A Behavior, and cardiovascular response in social interaction. Psychophysiology, 18(6), 654-664.

MacDougall, J. M., Dembroski, T. M., Dimsdale, J. E., & Hackett, T. P. (1985). Components of type A, hostility, and anger-in: Further relationships to angiographic findings. Health Psychology, 4(2), 137-152.

Pearson, T. A. (1983). Risk factors for arteriographically defined coronary artery disease. Doctoral Dissertation from John Hopkins University. #83-169-89. Ann Arbor, MI: University Microfilms International.

Pearson, T. A., Achuff, S. L., Kwiterovid, P. O., & Gordis, L. (1985). Type A personality and coronary atherosclerosis: An indirect association?, Circulation, 66(Supp II), 114.

Schwertner, H. A., Troxler, R. G., Uhl, G. S., & Jackson, W. G. (1983). Relationship between cortisol and cholesterol in men with coronary artery disease and Type A Behavior. Arteriosclerosis, 4(1), 59-64.

Scherwitz, L., McKelvian, R., Laman, C., Patterson, J., Dutton, L., Yusim, S., Lester, J., Kraft, I., Rochelle, D., & Leachman, R. (1983, Mar). Type A behavior self-involvement, and coronary atherosclerosis. Psychosomatic Medicine, 45(1), 47-57.

Silver, L., Jenkins, C. D., Ryan, T. J., & Melidossian, C. (1980). Sex differences in the psychological correlates of cardiovascular diagnosis and coronary angiographic findings. Journal of Psychosomatic Research, 24, 322-334.

Tennant, C. C., & Langeluddecke, P. M. (1985). Psychological correlates of coronary heart disease. Psychological Medicine, 15, 581-588.

Williams, Jr., R. B., Haney, T. L., Lee, K. L., Kong, Y., Blumenthal, J. A., & Whalen, R. E. (1980). Type A Behavior, hostility, and coronary atherosclerosis. Psychosomatic Medicine, 42(6), 539-549.

Williams, Jr., R. B., Barefoot, J. C., Haney, T. L., et al. (1986). Type A behavior and angiographically documented coronary atherosclerosis in a sample of 2,289 patients. Paper presented at the annual meeting of the American Psychosomatic Society; Baltimore, Md.

Young, L. D., Barboriak, J. J., Anderson, A. A., Hoffman, R. G. (1980). Attitudinal and behavioral correlates of coronary heart disease. Journal of Psychosomatic Research, 24, 311-318.

Young, L. D., Barboriak, J. J., Hoffman, R. G., et al. (1984). Coronary-prone behavior attitudes in moderate to severe coronary artery occlusion. Journal of Behavioral Medicine, 7, 205-215.

Zyzanski, S. J., Jenkins, C. D., Ryan, T. J., Flessas, A., & Everist, M. (1976). Psychological correlates of coronary angiographic findings, Archives of Internal Medicine, 136, 1234-1237.

APPROVAL SHEET

The thesis submitted by Todd Q Miller has been read and approved by the following committee:

Dr. R. Scott Tindale, Director
Assistant Professor, Psychology, Loyola

Dr. Emil J. Posavac
Professor, Psychology
Chairman, Department of Psychology, Loyola

The final copies have been examined by the director of the thesis and the signature which appears below verifies the fact that any necessary changes have been incorporated and that the thesis is now given final approval by the Committee with reference to content and form.

The thesis is therefore accepted in partial fulfillment of the requirements for the degree of Master of Arts.

R. Scott Tindale 4/9/87
Director's Signature Date