AN EPIDEMIOLOGICAL APPROACH TO MMPI-2 VALIDITY

Andrea T. Underhill

Certificate of Approval:

Steven G. LoBello, Ph. D.
Chairman
Associate Professor
Psychology

Glen E. Ray, Ph. D.
Associate Professor
Psychology

Peter Zachar, Ph. D.
Associate Professor
Psychology

Roger A. Ritvo, Ph. D.
Vice-Chancellor for
Academic Affairs

AN EPIDEMIOLOGICAL APPROACH TO MMPI-2 VALIDITY

Andrea T. Underhill

A Thesis

Submitted to

The Graduate Faculty of

Auburn University Montgomery

in Partial Fulfillment of the

Requirements for the

Degree of

Master of Science

Montgomery, Alabama

July 25, 2001

VITA

Andrea T. Underhill, daughter of Wayne Thames and Linda Taylor, was born May 1, 1973 in Columbia, South Carolina. She attended Georgia Southern University in Statesboro, Georgia and graduated with a Bachelor of Science in Chemistry in June, 1995. She attended the University of Georgia in Athens, Georgia and completed requirements for T-4 Certification in Secondary Science Education in March, 1997. Andrea taught high school in Washington, Georgia from 1997 until 1999 when she enrolled at Auburn University Montgomery. She entered Graduate School, Auburn University Montgomery, in March 2000.

THESIS ABSTRACT

AN EPIDEMIOLOGICAL APPROACH TO MMPI-2 VALIDITY

Andrea T. Underhill

Master of Science, August 3, 2001
(B.S., Georgia Southern University, 1995)

58 Typed Pages

Directed by Steven G. LoBello

The purpose of this research is to demonstrate the application of epidemiological methods to analysis of personality testing data and to show the relationship between MMPI-2 scale elevations (any clinical scale, AAS, APS, or FAM) and self-reported pathology or pathology in the family of origin. Results indicated that individuals with a personal history of drug or alcohol abuse are 11 times more likely to have an elevation on the MMPI-2 AAS, than individuals without such a history. Furthermore, this research demonstrated the utility of epidemiological methods for establishing personality test validity as well as showing that the AAS, APS, and FAM are more useful in identifying those without addiction, addiction potential, or family problems than in identifying individuals with these characteristics.

ACKNOWLEDGEMENTS

I would like to thank Dr. Steven G. LoBello for his endless support, advice, and especially for his unending patience through the conception, development, and completion of this project. Thanks also go to Drs. Glen E. Ray and Peter Zachar for their insights and suggestions. A special thank you is given to my family, and especially my husband, Russell, for their understanding of my infinite academic pursuits.

TABLE OF CONTENTS

LIST OF TABLES

An Epidemiological Approach to MMPI-2 Validity

Reliability and validity are crucial to establishing the usefulness of psychological tests and measures, as well as tests of physical health (e.g., blood tests). Psychologists have generally relied on correlational methods to determine the reliability and validity of tests. The frequently-used Pearson correlation technique is useful for determining the proportion of variance shared by two measures. However, this technique requires that both variables use a continuous scale of measurement. The quality of tests of physical health status is usually reported using measures such as sensitivity and specificity, which rely on categorical data. Epidemiological methods such as these have wide application in the field of psychometrics and could be used to enhance the utility of personality test data for clinicians. This paper describes those methods and reports the results of a study designed to illustrate their effectiveness for psychologists in clinical practice.

Validity

Validity refers to how well a test measures the criterion it purports to measure. Validity determines the suitability of the inferences drawn from the scores of the test (Cohen & Swerdlik, 1999). The association between test scores and a criterion measure acquired in the future is predictive validity. The relationship between the test scores and a criterion measure that is available at the time of testing is referred to as concurrent validity. Concurrent validity indicates the degree to which test scores may be used to assess an individual's current standing on a criterion (Cohen & Swerdlik, 1999). For example, a

researcher wishes to show the relationship between a high score on a test measuring reaction speed and success in a pilot training program. If the researcher administers this test to a group of people who have already been deemed successful or unsuccessful in the pilot training program, the data would contribute to establishing concurrent validity. If the researcher administers this test to a group of prospective pilots and later compares the test scores of those eventually deemed successful and unsuccessful in the pilot training program, the data would contribute to establishing predictive validity.

Odds Ratio

Odds ratios are widely used in medical research. The odds ratio provides an estimate, through the use of a confidence interval, of the relationship between two dichotomized variables and enables the examination of the effects of other variables on those variables (Bland & Altman, 2000). An odds ratio is a measure of association between an outside factor and a disease. It is the odds of development of disease for an individual exposed to a certain factor divided by the odds of development of disease for an individual not exposed to that factor (Gordis, 1996). If exposure to the factor is not related to development of disease, the odds ratio will be 1. An odds ratio can express both positive and negative relationships between exposure and disease development. If the relationship is positive, the odds ratio will be positive; if the relationship is negative, the odds ratio will be less than one (Gordis, 1996). One point to note with the use of an odds ratio is the relevance of the base rate of the condition in question. If exposure to some factor increases the likelihood of developing a disease 10

times, exposure to that factor poses more danger for an individual if the disease has a base rate of 1 in 20 than if the disease has a base rate of 1 in 1,000,000.

Sensitivity and Specificity

The sensitivity of a test is its ability to identify those in a population who have a disease. It is the number of positive tests divided by the number of tested individuals affected by the disease. Sensitivity is expressed as a percent, representing the likelihood that a positive test result actually identifies a person who has the disease in question. A test with high sensitivity produces few false negatives. The sensitivity of an instrument may be calculated by dividing the number of true positives by the number of true positives plus false negatives [(true positive)/(true positive + false negative)] (Turner, Herron, & Weiner, 1986). The true positives are those individuals with the disease who are correctly called positive by the test. The false negatives are those individuals with the disease who are called negative by the test (Gordis, 1996).

The specificity of a test is the ability to identify those in a population who are free of a disease. It is the number of negative tests divided by the number of tested individuals not affected by the disease. Specificity is expressed as a percent representing the likelihood that a negative test result actually identifies a person who does not have the disease in question. A test with a high specificity produces few false positives. The specificity of an instrument is calculated by dividing the number of true negatives by the number of true negatives plus false positives [(true negative)/(true negative + false positive)] (Turner et al., 1986). The true negatives are those individuals without the disease who are correctly

called negative by the test. The false positives are those individuals without the disease who are called positive by the test (Gordis, 1996).

Predictive Value

Positive predictive value (PPV) is the likelihood that a person with an elevated score actually has the condition in question (a true positive) (Malinchoc, Offord, Colligan, & Morse, 1994). The PPV of an instrument can be calculated by dividing the number of true positives by the total number of positives [(true positives) / (true positives + false positives)]. PPV is important clinically because by knowing the PPV of an instrument, the clinician can determine the likelihood that an individual actually has the disease if the diagnostic instrument used indicates the disease is present (Gordis, 1996).

Negative predictive value (NPV) is the probability that a person who tests negative for a disease actually does not have the condition in question (a true negative) (Malinchoc et al., 1994). The NPV of an instrument can be calculated by dividing the number of true negatives by the total number of negatives [(true negatives) / (true negatives + false negatives)]. As with PPV, NPV is important clinically. By knowing the NPV of a diagnostic instrument, the clinician can determine the likelihood that an individual is actually disease free if the instrument provides a negative result (Gordis, 1996).

Sensitivity/Specificity vs. PPV/NPV

While the sensitivity and specificity of an instrument are important, they are not of chief importance to the clinician or MMPI researchers. For these groups, the PPV & NPV are of primary interest. In a clinical setting, one knows

the individual's test scores and wishes to determine the presence or absence of a condition based on those scores, therefore, the accuracy of the instrument in determining presence or absence of a condition is of interest. This information is conveyed by the predictive power of the instrument, not its sensitivity or specificity (Butcher, Graham, & Ben-Porath, 1995).

One important point to note for the use of PPV & NPV is base rate. The PPV and NPV of an instrument vary with the base rate of the disorder among the population used in the research. It is important to use the PPV/NPV calculated for the base rate of the population of interest because as a variable's base rate decreases, so does the positive predictive power of the tests that are designed to identify the variable (Butcher et al., 1995).

Sensitivity and specificity are unique to each particular test, but predictive value is affected by the prevalence of the disease in the population examined and the specificity of the test used when the disease is infrequent (Gordis, 1996). The predictive value of an instrument is higher when the disease prevalence is higher in the tested population. The prevalence of a disease is the number of affected people divided by the total number of people in a particular population at a given time. As is demonstrated by the relationship between predictive value and disease prevalence, the results of any test must be interpreted from the standpoint of the prevalence of the disease in the participant's population. The same test can have dramatically different predictive values depending on whether it is administered to a high prevalence or low prevalence population (Gordis, 1996).

For a disease with a low prevalence, increasing the specificity of a testing instrument will increase the predictive value of that instrument within a given population. Because specificity is the ability of an instrument to detect the absence of disease in a population, when the disease prevalence is low, specificity is more important than sensitivity in determining predictive value. Because most of the population will be unaffected by a low prevalence disease, improvement in the ability of the instrument to detect the absence of the disease is more beneficial than improvement in its ability to detect the disease.

The preceding points are exemplified by the National Institute of Mental Health's (NIMH) data on depression in men and women in the United States. According to NIMH (2001), 12.4 million women and 6.4 million men in the United States suffer from a depressive disorder each year. Those numbers equal 12.0% of women in the U.S. and 6.6% of men in the U.S. As illustrated in Table 1, an instrument used to detect depression with a sensitivity of 99% and a specificity of 95% would have a PPV of 73.2% and a NPV of 99.9% for women and a PPV of 58% and a NPV of 99.9% for men because the prevalence of depression is different for women (12.0%) and men (6.6%) in the United States.

Table 1: Depression in Men and Women

	Sensitivity 99%	Specificity 95%
12% Prevalence	**Total Women:** 103.3 Million	**Depressed Women:** 12.4 Million
6.6% Prevalence	**Total Men:** 97 Million	**Depressed Men:** 6.4 Million

Prevalence	Test Results	Actually Depressed (in millions)	Actually Not Depressed (in millions)	PPV [true pos / (true pos + false pos)]	NPV [true neg / (true neg + false neg)]
12.0% (Women)	+	12.28	4.5	(12.28/16.78) x 100	(86.4/86.52) x 100
	-	0.12	86.4		
	Total	12.4	90.9	= 73.2%	= 99.9%
6.6% (Men)	+	6.34	4.53	(6.34/10.87) x 100	(86.07/86.13) x 100
	-	0.06	86.07		
	Total	6.4	90.6	= 58%	=99.9%

Dichotomized and Categorical Variables

Dichotomization of continuous variables in psychological research is a practice that is frowned upon (Cohen, 1983). The more accepted methodologies for research psychologists are group differences or correlation, which rely upon continuous dependent variables and statistical tests such as Analysis of Variance, t-test, and Pearson correlation.

One of the most frequently used methods of determining associations between variables in psychological research is through the product moment correlation, r (Farrington & Loeber, 2000). When correlation is used, certain assumptions are made about the measured variables. The three most salient of these assumptions are that the data are normally distributed, both variables are measured on an interval scale, and are linearly related to one another (Farrington & Loeber, 2000). One type of data commonly used as a variable in psychological research are test scores. Most test scores are measured on an ordinal scale, not an interval scale as is assumed by product moment correlation. Variables sometimes have skewed distributions, certain relationships between variables are severely affected by a few outliers, and variables may have interaction effects. Psychological practitioners are typically more interested in different types or different categories of individuals than in scale scores (Farrington & Loeber, 2000). This is the approach taken in psychiatric diagnosis using the Diagnostic and Statistical Manual of Mental Disorders, 4th edition (DSM-IV, American Psychiatric Association [APA], 1994).

There are statistical solutions to violations of assumptions with correlation, such as statistical transformations or forcing a variable into a normal distribution, but those solutions are complex. The complexity of those solutions makes communication with non-statisticians difficult (Farrington & Loeber, 2000). Non-statisticians need to be able to understand the results of research for themselves and not blindly accept the findings of the researchers. One way to simplify the presentation of research results, making them understandable to a diverse audience, is through the dichotomization of variables (Farrington & Loeber, 2000).

Probably the main objection to dichotomization of variables is decreased strength of association (Cohen, 1983). If one of two continuous, normally distributed variables is dichotomized, their product moment correlation falls to 80% of their product moment correlation before dichotomization. If both of the two continuous, normally distributed variables are dichotomized, their product moment correlation falls to 64% of their product moment correlation before dichotomization. These reductions in product moment correlations are the equivalent of discarding of 38% and 60% of the cases, respectively (Cohen, 1983). This loss of strength of association, however, is only an issue if the dichotomized variables are re-correlated after dichotomization. The correlation is not the only method for showing relationships between variables.

There are other measures of association that compensate for the effects of dichotomization. Techniques such as a biserial correlation (one dichotomous, one continuous variable) and tetrachoric correlation (two continuous variables)

result in a correlation between dichotomous variables that approximates the correlation before dichotomization (Farrington & Loeber, 2000). If one of two originally continuous variables is dichotomized, a biserial correlation amends the measured, or point-biserial, correlation between the continuous variable and the dichotomous variable to the true, or product moment, correlation. The biserial correlation assumes that both variables possess an underlying, normal distribution. If both originally continuous variables are dichotomized, a tetrachoric correlation amends the measured, or phi, correlation between the dichotomous variables to the true, or product moment, correlation. The tetrachoric correlation, like the biserial correlation, assumes that both variables are normally distributed (Farrington & Loeber, 2000). By using the biserial or tetrachoric correlation, there is no loss of strength of association when variables are dichotomized.

The loss of strength of association as measured by a correlation coefficient may not be of importance in some research. That loss of strength is only observed when a correlation is attempted with the dichotomous data. A better measure of strength of association between dichotomous variables is an odds ratio. A product moment correlation can give a misleading impression of weak relationships between dichotomized variables. Unlike correlation, an odds ratio remains constant through different prevalence rates (Farrington & Loeber, 2000). A more realistic impression of relationships between variables is provided by an odds ratio if the variables are dichotomized (Farrington & Loeber, 2000). An odds ratio is easily understood as the increase in the risk of an outcome associated with a risk factor.

Another objection to dichotomization of variables is loss of information (Farrington & Loeber, 2000). Differences in the sensitivity of the measurement of explanatory variables can make the drawing of conclusions concerning the relative strength of those variables' relationships with an outcome variable quite difficult. The sensitivity of measurement of all variables is equated through dichotomization. This equating allows the comparison of the predictive strengths of explanatory variables. It has been argued that this equating causes a detrimental loss of information about the differences between individuals. Loss of information, however, is unavoidable in most psychiatric analyses because the amount of information amassed surpasses the researcher's ability to analyze and report that data. Furthermore, dichotomized variables do not necessarily contain less information than other types of variables. The amount of information conveyed by dichotomized variables depends on the relative number of each type of variable and the accuracy of the measurements (Farrington & Loeber, 2000).

Other benefits of using dichotomized variables include the ease with which multiple risk factors and potential interaction effects can be systematically studied. Moreover, dichotomization does not greatly affect the order of importance of explanatory variables and its results are similar to logistic and ordinary least squares (OLS) multiple regression (Farrington & Loeber, 2000).

Most of the criticism of dichotomization centers around the loss of statistical power and information attendant with changing a variable with considerable range into one with only two categories. It is interesting that

psychologists have embraced methods such as analysis of variance and other parametric techniques because they all involve dichotomization or categorization of the independent variable. Both the t -test and the biserial correlation involve one dichotomous and one continuous variable . Researchers readily use these methods of data analysis while still opposing dichotomization. Yet, the use of these methods of data analysis entails the acceptance of one dichotomized variable, making the justification for rejecting dichotomization difficult to maintain.

The information obtained from analysis of dichotomized variables is not inferior to the information obtained from the more traditional methods of data analysis (t-test, correlation, etc). These methods provide answers to a different set of questions. Those opposed to the use of dichotomized variables may be overlooking the types of questions answered by these methods. It is not suggested that t-tests, correlations, and other traditional methods of data analysis be abandoned. It is recommended that researchers become familiar with methods of analyzing dichotomized variables, or epidemiological methods of data analysis (odds ratio, sensitivity, specificity, predictive value) and incorporate those methods into their research.

Using both traditional correlational and epidemiological methods in data analysis will not add much additional work and will make that research more usable by clinicians by demonstrating not only the degree of relationship between variables but also the ability of those variables to predict behaviors of interest.

A risk factor approach to psychological research is desirable because it facilitates primary prevention efforts. If risk factors can be identified, then groups

exposed to these risk factors can be identified and steps taken to prevent the development of the condition in question. Also, primary prevention involves efforts to lower the incidence rate of disease. In the case of medical illness, primary prevention is almost always a more cost-efficient approach to treatment than finding cures. The same argument applies to severe and chronically debilitating psychiatric illnesses.

Description of the MMPI

The MMPI was developed by Starke Hathaway, Ph.D., and J. Charnley McKinley, M.D., while working at the University of Minnesota Hospitals. It was originally published in 1943 and its primary purpose was to assign appropriate psychodiagnostic labels to individuals. This type of assignment was seen as more efficient and reliable than the traditional method of individual interviews (Graham, 1993).

The 1943 MMPI consisted of ten clinical scales and four validity scales and remained unchanged until its revision in 1989. The revised version is the MMPI-2. The main goals of the revision were to create a contemporary normative sample and to update the language used in the statements. The MMPI-2 is similar to the MMPI in most ways. The MMPI-2 consists of 567 test items that make up ten clinical scales and four validity scales (Graham, 1993). Readers interested in the history of the MMPI are directed to Dahlstrom and Dahlstrom (1980).

Validity of the MMPI

The predictive validity (the association between test scores and a criterion measure acquired in the future) of the Minnesota Multiphasic Personality Inventory-2 (MMPI-2) is of primary importance. Clinicians use the MMPI-2 to predict a wide array of criteria including psychopathology, length of hospital stay, success of therapy, or suitability for a particular career. Because the MMPI and the MMPI-2 are among the most popular diagnostic personality tests, the research on the validity of the MMPI and MMPI-2 has been extensive. One MMPI handbook cited over six thousand studies of the MMPI. The continuity between the MMPI and the MMPI-2 means that the validity studies for the MMPI are relevant to the determination of the validity of the MMPI-2 (Graham, 1993).

When concurrent validity (the association between test scores and a criterion available simultaneously) has been established, it provides a faster, less expensive method of diagnosis or classification for individuals than diagnosis and classification through extensive interviews. It is easy to understand why it is important for the MMPI to have concurrent validity. The MMPI holds a potential for saving money and professional time. The MMPI has the ability to test for many different conditions in one relatively short period. It would take much longer for a diagnostician to conduct the proper interviews to determine the presence of the same conditions for which the MMPI tests.

MMPI Validity Research: Correlational Method

When researchers explore the MMPI-2's ability to correctly, or incorrectly, distinguish between groups of individuals, the information gathered

from that exploration is valuable to different factions of psychologists. The most obvious group for whom that information is helpful is other researchers. Other researchers use the procedures, methods of data analysis, and results of previous research to fuel current research. Clinicians also find that information helpful. Clinicians are primarily interested in the results of the research and how those results can aid practice.

A popular method in MMPI research is to find a correlation between an MMPI or MMPI-2 scale score and an outside criterion. For example, Faurie (1990) investigated the usefulness of the Psychopathic Deviate scale of the MMPI and demographic factors in predicting adolescents' length of hospitalization. The study explored adolescents admitted to a state hospital who obtained a T-score > 70 on the MMPI Psychopathic Deviate scale. The demographic information included age, gender, WISC-R scores (Verbal, Performance, and Full Scale IQ), parent situation, sibling information, and length of hospitalization. The MMPI did not have predictive value for determining length of hospital stay; however, the number of natural siblings was negatively correlated ($r = -0.24$, $p < 0.01$) with the Familial Discord. Familial Discord is a subscale of the MMPI Psychopathic Deviate scale. This correlation indicates that as the number of natural siblings increases, the amount of familial discord decreases.

Quereshi and Kleman (1996) sought to show concurrent validity between MMPI-2 components and first four of the Big Five factors through the use of the Mitchill Adjective Rating Scale (MARS). Through the use of Pearson correlations,

the researchers demonstrated that selected MMPI-2 components (Basic Depression, Content Anxiety, Content Depression, Basic Social Introversion, Content Social Discomfort, Content Anger, Content Type A, Content Work Interference, Supplementary Dominance) were significantly correlated with MARS components (Unhappiness, Extraversion, Self-assertiveness, Productive Persistence) in both men and women. The MARS components were found to be significantly correlated with the first four of the Big Five factors.

MMPI Validity Research: Correlational Method Critique

In both of the correlational studies described above, an odds ratio would have been a more straightforward manner of presenting the data. An odds ratio can easily deal with data that are not linearly related and it indicates the likelihood of one characteristic based on the presence of another, which is important to clinicians. The information conveyed by Faurie (1990) concerning familial discord, while interesting, is not particularly informative to the clinician. This finding does not indicate if there is an optimum number of siblings for decreased familial discord. It is not clear whether or not the relationship between number of siblings and decreased familial discord is linear. If the relationship is not linear, correlation is not the ideal method for determining strength of association. An odds ratio is better suited for dealing with variables that are not linearly related (Farrington & Loeber, 2000).

Strictly speaking, there is no problem with the concurrent validity study conducted by Quereshi and Kleman (1996). There is value in showing the amount of shared variance between the MMPI-2 scales and the MARS scales

through a correlation. From the standpoint of a clinician, however, the more salient issue would be determining the likelihood of an elevation on the MMPI-2 being indicative of an elevation on the MARS. The significant correlations between the MMPI-2 content scales and the MARS ranged from 0.26 to 0.67 for both males and females. The statistical significance of those numbers lies only in indicating that the relationship between the two variables is not zero in the population. As sample size increases, the strength of correlation needed to declare significance diminishes. A relatively small correlation may be statistically significant, but still lead to many misclassifications of individuals.

MMPI Validity Research: Group Differences Method

Much of the MMPI validity research compares groups with different characteristics by examining their respective average scores on the MMPI scales. The groups are determined by the independent variables of the study such as men and women or schizophrenic and non-schizophrenic, etc. Scores on the MMPI scales are commonly compared for the different groups by finding the mean and standard deviation for each group's scale scores. Analysis of variance or a t-test is then used to determine whether or not differences exist. This method of using group-differences can be used for establishing both predictive and concurrent validity (Walsh & Betz, 1995). For example, Craig and Olson (1992) compared the means and standard deviations of the MMPI scales between male and female PsyD students and found statistically significant differences on the Psychopathic Deviate, Psychasthenia, and Schizophrenia scales. The men

scored higher than women on these scales where significant differences were found.

Herkov and Myers (1996) compared MMPI scale scores of depressed adolescents with and without conduct disorder. For both groups, the mean and standard deviations of the clinical scales were calculated. The means for both groups on each scale were compared. The scales where significant differences were found included Frequency (F) scale, Depression scale, Hypomania scale, and Social Introversion scale. On these scales, the average scores were higher for individuals with manic depression and conduct disorder on the F Scale and the Hypomania scale. The average scores were higher for individuals with manic depression without conduct disorder on the Depression scale and the Social Introversion scale.

Hackney and Ribordy (1980) compare MMPI scale means among four groups of people: happily married, marriage counseling, divorcing, and divorced. The investigators found several significant differences among these groups. There are main effects for group found for the Depression, Psychasthenia, Hypochondriasis, Psychopathic Deviate, and Paranoia scales. Pairwise comparisons showed that the Happily Married group did not differ significantly from the Divorced group, and the Marriage Counseling group did not differ significantly from the Divorcing group in T scores for several scales. The Happily Married group and the Divorced group scored similarly on the Depression, Psychasthenia, Schizophrenia, and Hypochondriasis scales. The Happily Married

group and Divorced groups scored significantly lower on those scales than did the Marriage Counseling and Divorcing groups.

MMPI Validity Research: Group Differences Method Critique

The group differences studies cited above use scale score means as their method for comparing different groups. Because it is influenced by extreme scores and its value may not actually exist in the data, the mean is not always the best method for comparing groups (Howell, 1997). From a clinician's standpoint, a more useful method for this study would have been to present the numbers of participants who were elevated on a particular scale rather than presenting the means and standard deviations for participants. A significant t-test indicates there is a reliable difference between two groups. Clinicians are not interested in how groups are different, but in how well a test can predict extra-clinical behavior. This indication of ability to predict extra-clinical behavior is provided by epidemiological methods, not group differences. Non-elevated scores, those below 70 or 65 on the MMPI and MMPI-2 respectively, are not indicative of any particular characteristics whereas those scores that are elevated are indicative of particular characteristics.

In Craig and Olson's (1992) study investigating differences between male and female PsyD students, none of the scales where significant differences were found were elevated from a clinical perspective (e.g. greater than 70). Hackney and Ribordy's (1980) study involving the married, marriage counseling, divorcing, and divorced groups had no elevated average scale scores, either. Although Herkov and Meyers (1996) did find significant differences between groups'

average scale scores that involved some clinical elevations, not all of their significant findings involved an elevation. The significant difference found between the two groups on the Social Introversion scale involved an elevation of neither group. One must ask, how much information is presented in the data's current form? Even if there were no elevated average scale score, it is reasonable to assume that some of the individual scores were elevated. Had these data been dichotomized into elevated/non-elevated categories, the consumer would be able to see how many individuals in each category were elevated on a particular scale. These designs are better suited for finding differences in mean scores rather than for determining the predictive capability of the instrument.

A more straightforward manner of presenting the data from these studies reviewed would have been to indicate the numbers of individuals in each group who were elevated on each scale. By dichotomizing the data into elevated/non-elevated groupings, an odds ratio could have been calculated. An odds ratio is not only easily understood, but also makes the information more useful to practitioners by providing information that allows prediction of relevant behavior (Farrington & Loeber, 2000).

MMPI Validity Research: Categorical Methods

Not all MMPI-2 research is correlational or utilizes contrasting groups. Other researchers have employed categorical methods to investigate validity. Patalano (1998) categorized drug abusers as having a characterological disorder, a thinking disturbance, an emotional disturbance, or as being

asymptomatic through use of an MMPI diagnostic classification system with a cut-off score of T=70. The participants in this research were categorized by gender and subcategorized according to race (Caucasian or African American). Patalano presents the results of his study as frequencies of each disorder within each category and subcategory. Patalano presents the data in a table containing the number of individuals having each of four diagnostic classifications within each category and subcategory. Through the use of a chi square test, Patalano found no significant differences among the four groups with respect to their diagnostic classification.

Coleman and Frick (1994) compared elevations on MMPI-2 scales between adult children of alcoholics (ACOAs) and a control group. Although means of the MMPI-2 scales are compared between the ACOA and control groups, data are presented that indicate the proportions of participants in each group with a T score of greater than 70 for each particular scale. A chi-square test was utilized to determine if the proportions of participants in each group elevated on each scale were significantly different. Significance was determined for each MMPI-2 clinical scale except Masculinity-Femininity, Hypomania, and Social Introversion.

Trief and Yuan (1983) looked at the use of the MMPI in a chronic back pain rehabilitation program. They were interested in the ability of the MMPI to predict the outcome for individuals within the program in the areas of seeking more treatment, doing more, improved mobility, and working. The areas that were examined were dichotomized (either mobility was improved or not; either

the individual was working or not, etc). A frequency table was constructed showing the numbers of individuals within each category for "poor risk" types and "good risk" types as determined by the MMPI scale elevations. Although Trief and Yuan could have easily calculated odds ratios from the data they presented, they chose to use chi-square to determine significance of their results. The chi-square test showed that significantly more individuals with a "good risk" type were working following surgery than those with a "poor risk" classification.

Velasquez, Callahan, and Young (1993) present a table comparing the means and standard deviations of MMPI scale scores of Hispanics and Whites. However, they also performed stepwise discriminant analyses to ascertain how well the MMPI scales could determine ethnicity and diagnosis. These researchers calculated that by using only 6 particular scales, the MMPI could correctly classify 68.5% of Whites and 74.1% of Hispanics for an overall rate of 71.3%. The MMPI's ability to classify participants as having schizophrenia, major depression, or antisocial personality disorder was investigated using another combination of the test's scales. The discriminate analysis revealed the MMPI's ability to correctly classify 61.7% of all participants with schizophrenia, 57.1% of all participants with major depression, and 75.0% of all participants with antisocial personality disorder, for an overall correct classification rate of 63.0%.

Bartol (1991) examined the ability selected MMPI scales for predicting the success or failure of small-town police officers. Through the use of canonical discriminant analysis, he found that the use of the Immaturity Index (combined raw scores of the L, Psychopathic Deviate, and Hypomania scales of the MMPI)

along with the size of the department, the K scale, and the Hysteria scale were the most accurate predictors of officer success or failure. The model correctly classified 80.04% of the officers. The model missed 22.7% of the officers who eventually failed and rejected 17.2% of those officers who were successful.

Turner et al. (1986) placed their participants into different categories (good, fair, poor) based on their recovery from lumbar surgery. This categorization was then compared to the categorization predicted by the Pain Assessment Index (PAI) of the MMPI that was completed by the participants prior to surgery. The analysis of the data focuses on the sensitivity, specificity, and correct classification rates of MMPI predictors of surgical outcomes, and on the sensitivity, specificity, and correct classification rates of MMPI predictors of return to work after surgery. The MMPI predictors include the PAI, the individual scales of Hypochondriasis and Hysteria, and the Hypochondriasis and Hysteria scales combined. Of the four different MMPI predictors of surgical outcome, the Hypochondriasis scale alone was best overall with a sensitivity of 63%, a specificity of 90%, and a correct classification of 83%. For prediction of returning to work after surgery, none of the MMPI predictors stood out as markedly better than the others.

Malinchoc et al. (1994) developed the MMPI-2's Common Alcohol Logistic-Revised scale (CAL-R) and determined its sensitivity and specificity by testing alcoholic inpatients and nonalcoholic medical outpatients. The CAL-R's sensitivity is 92% for females and 91% for males. The specificity of the CAL-R is 95% for females and 88% for males. In other words, when using a cutting score

of T=61 for females, the CAL-R can correctly classify 92% of the alcoholic females as alcoholic and 95% of the nonalcoholic females as nonalcoholic. For males, the cutting score is T=58 and the CAL-R can correctly classify 91% of the alcoholic males as alcoholic and 88% of the nonalcoholic males as nonalcoholic. Furthermore, they determined the CAL-R's positive and negative predictive values for a variety of prevalence rates. The NPV and PPV for differing prevalence rates was presented because the prevalence of alcoholism depends on the population being examined and predictive values differ as prevalence rate changes.

MMPI Validity Research: Categorical Methods Critique

Each of these studies is similar in that the researchers made some use of categorical data. All had the raw data necessary to calculate odds ratios, PPV, NPV, sensitivity, and specificity. Some researchers chose to include that information in their publication and some did not, but none of the researchers took the analysis of their categorical data as far as they could have taken it. None of these researchers made use of all of the epidemiological methods, such as odds ratio, predictive value, sensitivity, and specificity, that are available.

Three of the studies (Coleman & Frick, 1994, Patalano, 1998, Trief & Yuan, 1983) provide enough information (e.g. raw numbers or proportions) to allow for the calculation of odds ratios, sensitivity, specificity, and predictive value. For example, the data included in the article by Coleman and Frick (1994) can be analyzed to reveal that ACOAs are 28.1 times more likely to have an elevation on the Paranoia scale than non-ACOAs. Trief and Yuan (1983) report a

significant chi-square for data on post-rehabilitation return to work by MMPI profile type ("good risk"/"poor risk"). The odds ratio, not provided in the article, indicates that patients with "good risk" MMPI profiles were 29.5 times more likely to return to work than people with "poor risk" profiles.

Multivariate techniques such as the stepwise discriminant analysis employed by Velasquez et al. (1993) and the canonical discriminant analysis of Bartol (1991) are helpful in demonstrating that MMPI scores may be used to classify cases or predict outcomes. However, these methods typically involve multiple scales to predict an outcome, and are poorly understood by those who do not have an adequate background in statistical analysis. Because of this, the results of these studies are not easily applied in a clinical setting.

Turner et al. (1986) and Malinchoc et al. (1994) provide information about the sensitivity and specificity of the MMPI scales. Predictive values were included in the Malinchoc, et al. (1994) study. Odds ratios could have been presented in each study. These studies come closest to providing information that would be useful to clinicians in practice settings.

Purpose

There seems to be a point of miscommunication between psychometric research psychologists and clinical psychologists. Researchers use continuous data and report correlation coefficients or group mean differences. Clinicians could better use validity data based on dichotomized variables since they usually view people categorically, as diseased or normal. These conflicting points of view make the work of the research psychologists of less practical value to the clinical

psychologists. The clinical world could benefit greatly from the research done by the experimental world if that information was written in the same language. By using odds ratios, PPV, NPV, etc., the experimental psychologist can still support his/her theory scientifically through mathematics while allowing the clinical psychologist to use the data in his/her practice.

The purpose of this research is to demonstrate the application of epidemiological methods to analysis of personality testing data and to show the relationship between MMPI-2 scale elevations and self-reported pathology or pathology in the family of origin. This research will investigate the value of gross inspection of an MMPI-2 profile for determining the likelihood that the test-taker has a history of drug or alcohol abuse or a history of psychopathology, either personally or within his/her family. Furthermore, this research will determine if comparing certain special scale elevations (e.g. Addiction Acknowledgement Scale (AAS), Addiction Potential Scale (APS), Family Problems Scale (FAM)) to determine which is a better indicator of personal or familial drug or alcohol problems or personal or familial psychopathology than gross inspection.

Hypotheses

The following hypotheses will be investigated in this study: (1) participants with a personal or familial history of drug or alcohol abuse or psychopathology will be more likely to have any elevation on the clinical scales of the MMPI-2 (excluding the Masculinity/Femininity and Social Introversion scales) than individuals with no personal or familial history of drug or alcohol abuse or psychopathology, (2) individuals with a personal history of drug or alcohol abuse

will be more likely to have an elevated score on the MMPI-2 Addiction Acknowledgement Scale (AAS) than participants with no personal history of drug or alcohol abuse, (3) individuals with a family history of drug or alcohol abuse will be more likely to have an elevation on the MMPI-2 Addiction Potential Scale (APS) than those who have no familial drug or alcohol abuse, (4) individuals with a personal or familial history of psychopathology will be more likely to have an elevation of the MMPI-2 Family Problems Scale (FAM).

Method

Participants

Participants were individuals who had volunteered for practice MMPI-2 administrations given by graduate students enrolled in an objective personality appraisal course at Auburn University Montgomery. The MMPI-2 was the first personality test administered by all students enrolled in this class. All students administered ten MMPI-2's, scored those tests, and plotted the corresponding profiles. The students also wrote detailed psychological reports on five of the ten tests administered.

There are 68 participants. Of those 68 participants, 51 were Caucasian, 15 were African American, 1 was Native American, and 1 profile failed to record a race. There were 23 male and 45 female participants. The age of the participants ranged from 18 to 60 years. Information for this sample was obtained from Data Summary sheets completed at the time of the student's interview with the participant for the psychological report.

Instruments

Instruments used in this research include the MMPI-2 and a Data Summary sheet. The Data Summary sheet (see Appendix A) includes questions concerning the participants age, gender, and race. It also inquires about the marital status (parents married, parents divorced, single parent, etc.) of the person or persons who raised the participant. The Data Summary sheet asks the participant to identify the frequency and degree of conflict within his or her family

of origin and asks for any history, either personal or familial, of drug or alcohol abuse or of psychological disorders.

Procedure

Information about personal or familial drug or alcohol abuse and personal or familial history of psychological disorder was gathered from Data Summary sheets. These Data Summary sheets were given to all students in Advanced Objective Testing to ensure the gathering of routine information during the interview. The information from the Data Summary sheets was used to classify participants according to their self-reporting status with regard to personal/family drug or alcohol use and personal/family psychological history.

Gross inspection of the MMPI-2 profile refers to examination only of the clinical scales. The Masculinity/Femininity and Social Introversion scales were not considered for elevation because, strictly speaking, they do not measure psychopathology. The sample was divided into two groups based on whether or not any of the remaining clinical scales were elevated, having scores of 65T or above. The AAS consists of thirteen MMPI-2 items, each of which has obvious content related to substance abuse. Persons scoring above 60T on the AAS are acknowledging substance abuse problems (Graham, 1993). The APS consists of thirty-nine MMPI-2 items that do not have content obviously related to substance abuse. These thirty-nine items deal with extroversion, excitement seeking, risk taking, self-doubt, self-alienation, and cynical attitudes. Optimal classification using this scale is achieved by using a cut-off score of 60T (Graham, 1993). The FAM consists of twenty items. Scores above 65T are indicative of individuals who

describe considerable discord within their families, who describe their families as lacking love, who resent the demands and advice of family members, who experience anger and hostility toward family members, and who view marital relationships as lacking affection and involving unhappiness (Graham, 1993).

For each hypothesis, the PPV, NPV, sensitivity, specificity, and odds ratio, was calculated (see Appendix B). A Pearson correlation and a t-test were also calculated for each hypothesis. This allowed comparisons to be made between the number of elevations on the clinical scales and the scales specifically constructed to detect particular psychological problems, and allowed for comparison between results from epidemiological methods and the more traditional methods.

Results

For each hypothesis, an odds ratio, sensitivity, specificity, positive predictive value, and negative predictive value were calculated. To determine the statistical significance of the odds ratio, a two-tailed Fisher's Exact Test was utilized. Chi-square is usually appropriate for determining the significance of an odds ratio; however, because of relatively small sample size, the contingency tables created to test the hypotheses had at least one cell frequency less than five. Because of this asymmetry in the cell blocks, the two-tailed Fisher's Exact Test is more appropriate than chi-square. To compare the epidemiological results to the more traditional results, a Pearson correlation and a t-test were calculated for the hypotheses involving the Addiction Acknowledgement Scale, Additction Potential Scale, and Family Problems Scale. The Pearson correlation was between one categorical (presence or absence of characteristic) and one continuous (scale score) variable. For the hypotheses involving gross inspection, a Pearson correlation was calculated between the number of elevated scales and the reported personal or family histories Relevant statistics programs in SAS were used for the calculation of the correlations, t-tests, and odds ratios. Table 2 at the end of this section contains a summary of the results.

Gross Inspection of Clinical Scales

The Pearson correlation between the number of clinical scale elevations and a personal history of drug or alcohol abuse was significant ($r = 0.27$, $p = 0.03$). Using gross inspection, having any elevation of the MMPI-2 clinical scales suggests that an individual will be 3 times (95% confidence interval 0.342 -

26.278) more likely to have a personal history of alcohol or drug abuse than someone without an elevation. This odds ratio was not statistically significant (Fisher's Exact Test (2) $\underline{p}$ = 0.642). The sensitivity of gross inspection as an indicator of personal alcohol or drug abuse was 80%, specificity was 43%, positive predictive value was 10%, and negative predictive value was 96%. As was stated earlier, elevated scores are those T scores of 65 or above and inspection of the clinical scales excludes the Masculinity/Femininity scale and the Social Introversion scale.

The Pearson correlation between the number of clinical scale elevations and a family history of drug or alcohol abuse was not significant (r = 0.07, $\underline{p}$ = 0.55). Having any clinical scale elevation suggests that an individual will be 0.8 times (95% confidence interval 0.277 – 2.313) more likely to have a family history of alcohol or drug abuse than someone without an elevation. This odds ratio was not statistically significant (Fisher's Exact Test (2) $\underline{p}$ = 0.789). The sensitivity of gross inspection as an indicator of family alcohol or drug abuse was 55%, specificity was 40%, positive predictive value was 28%, and negative predictive value was 68%.

The Pearson correlation between the number of clinical scale elevations and a personal history of psychopathology was significant (r = 0.29, $\underline{p}$ = 0.02). Having any clinical scale elevation suggests that an individual will be 8 times (95% confidence interval 1.203 – 51.086) more likely to have a personal history of psychopathology than someone without an elevation. This odds ratio was statistically significant (Fisher's Exact Test (2) $\underline{p}$ = 0.039). The sensitivity of gross

inspection for determining a personal history of psychopathology was 90%, specificity was 47%, positive predictive value was 23%, and negative predictive value was 96%.

The Pearson correlation between the number of clinical scale elevations and family history of psychopathology abuse was not significant ($r = -0.01$, $p = 0.94$). Having any clinical scale elevation suggests that an individual will be 0.8 times (95% confidence interval 0.277 – 2.313) more likely to have a family history of psychopathology than someone without an elevation. This odds ratio was not statistically significant (Fisher's Exact Test (2) $p = 0.789$). The sensitivity of gross inspection as an indicator of family psychopathology was 55%, specificity was 40%, positive predictive value was 28%, and negative predictive value was 68%.

Addiction Acknowledgement Scale

The Pearson correlation between an elevated MMPI Addiction Acknowledgement Scale (AAS) score (T score of 60 or greater) and a personal history of drug or alcohol abuse is statistically significant ($r = 0.460$, $p = 0.0001$) indicating a moderate positive relationship between scores on the AAS and reported history of drug or alcohol abuse. A t-test was conducted on the mean AAS score for those who reported a personal history of drug or alcohol abuse and those who did not. The t-test between means was significant ($t(66) = -4.21$, $p = 0.0001$). Having an elevated AAS score suggests an individual will be 11 times (95% confidence interval 1.612 – 72.665) more likely to have a personal history of alcohol or drug abuse than someone without an elevated score. This odds ratio was statistically significant (Fisher's Exact Test (2) $p = 0.029$). The

sensitivity of AAS for determining a personal history of alcohol or drug abuse was 80%, specificity was 73%, positive predictive value was 19%, and negative predictive value is 98%.

Addiction Potential Scale

The Pearson correlation between an elevated score (T score of 60 or greater) on the MMPI-2 Addiction Potential Scale (APS) and a personal history of drug or alcohol abuse is not statistically significant ($r = 0.207$, $p = 0.091$). A t-test was conducted on the mean APS score for those who reported a personal history of drug or alcohol abuse and those who did not. The t-test between means was not significant ($t(66) = -1.72$, $p = 0.09$). Having an elevated score on the APS suggests that an individual will be 12 times (95% confidence interval 2.301 – 62.575) more likely to have a personal history of alcohol or drug abuse than someone without an elevated APS score. This odds ratio was statistically significant (Fisher's Exact Test (2) $p = 0.020$). The sensitivity of APS for determining a personal history of alcohol or drug abuse was 60%, specificity was 89%, positive predictive value was 30%, and negative predictive value was 97%.

The Pearson correlation between an elevated APS score and a family history of drug or alcohol abuse was not statistically significant ($r = 0.168$, $p = 0.170$). A t-test was conducted on the mean APS score for those who reported a family history of drug or alcohol abuse and those who did not. The t-test between means was not significant ($t(66) = -1.39$, $p = 0.17$). Having an elevated score on the APS suggests that an individual will be 1 time (95% confidence interval 0.236 – 4.526) as likely to have a family history of alcohol or drug abuse as someone

without such an elevation. This odds ratio was not statistically significant (Fisher's Exact Test (2) p = 1.000). The sensitivity of APS for determining a family history of alcohol or drug abuse was 15%, specificity was 86%, positive predictive value was 30%, and negative predictive value was 71%.

Family Problems Scale

The Pearson correlation between an elevated score (T score of 65 or greater) on the MMPI Family Problems Scale (FAM) and a personal history of drug or alcohol abuse was statistically significant (r = 0.309, p = 0.010). A t-test was conducted on the mean FAM score for those who reported a personal history of drug or alcohol abuse and those who did not. The t-test between means was significant (t (66) = -2.64, p = 0.01). Having an elevated score on the FAM suggests that an individual will be 4 times (95% confidence interval 0.647 – 24.718) more likely to have a personal history of alcohol or drug abuse than someone without such an elevation. This odds ratio was not statistically significant (Fisher's Exact Test (2) p = 0.181). The sensitivity of FAM for determining a personal history of alcohol or drug abuse was 40%, specificity was 86%, positive predictive value was 18%, and negative predictive value was 94%.

The Pearson correlation between an elevated FAM score and a family history of drug or alcohol abuse was not statistically significant (r = 0.051, p = 0.682). A t-test was conducted on the mean FAM score for those who reported a family history of drug or alcohol abuse and those who did not. The t-test between means was not significant (t (66)= -0.41, p = 0.68). Having an elevated score on the FAM suggests that an individual will be 1.5 times (95% confidence interval

0.375 – 5.719) more likely to have a family history of alcohol or drug abuse thans someone without such an elevation. This odds ratio was not statistically significant (Fisher's Exact Test (2) p = 0.719). The sensitivity of FAM for determining a family history of alcohol or drug abuse was 36%, specificity was 72%, positive predictive value was 20%, and negative predictive value was 85%.

The Pearson correlation between an elevated FAM score and a personal history of psychopathology was not statistically significant (r = 0.102, p = 0.406). A t-test was conducted on the mean FAM score for those who reported a personal history of psychopathology and those who did not. The t-test between means was not significant (t (66) = -0.84, p = 0.41). Having an elevated score FAM scale suggests that an individual will be 1.4 times (95% confidence interval 0.246 – 7.542) more likely to have a personal history of psychopathology than someone without such an elevation. This odds ratio was not statistically significant (Fisher's Exact Test (2) p = 0.660). The sensitivity of FAM for determining a personal history of psychopathology was 20%, specificity was 85%, positive predictive value was 18%, and negative predictive value was 86%.

The Pearson correlation between an elevated FAM score and a family history of psychopathology was not statistically significant (r = -0.125, p = 0.311). A t-test was conducted on the mean FAM score for those who reported a family history of psychopathology and those who did not. The t-test between means was not significant (t (66)= 1.02, p = 0.31). Having an elevated score on the FAM suggests that an individual will be 0.5 times (95% confidence interval 0.096 – 2.424) more likely to have a familial history of psychopathology than someone

without such an elevation. This odds ratio was not statistically significant (Fisher's Exact Test (2) $\underline{p} = 0.487$). The sensitivity of FAM for determining a familial history of psychopathology was 18%, specificity was 68%, positive predictive value was 10%, and negative predictive value was 81%.

Table 2: Summary of Results

Variable	r	t	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Odds Ratio	Confidence Interval
Number of Elevations and								
PERALC	0.27*		80	43	10	96	3	0.34- 26.28
FAMALC	0.07		55	40	28	68	0.8	0.28 - 2.31
PERPSYCH	0.29*		90	47	23	96	8 *	1.20 - 51.09
FAMPSYCH	-0.01		55	40	28	68	0.8	0.28 - 2.31
AAS and								
PERALC	0.46 **	4.5 **	80	73	19	98	11 *	1.61 - 72.67
APS and								
PERALC	0.21	1	60	89	30	97	12 *	2.30 - 62.58
FAMALC	0.17	1.7	15	85	30	71	1	0.24 - 4.53
FAM and								
PERALC	0.31 **	2.1**	40	86	18	94	4	0.65 - 24.72
FAMALC	0.05	1.5	36	72	20	85	1.5	0.38 - 5.72
PERPSYCH	0.1	1.1	20	85	18	86	1.4	0.25 - 7.54
FAMPSYCH	-0.13	1.1	18	68	10	81	0.5	0.10 - 2.42

Note: PERALC = personal history of drug or alcohol abuse, FAMALC = family history of drug or alcohol abuse, PERPSYCH = personal history of psychopathology, FAMPSYCH = family history of psychopathology.

* $p < 0.05$

** $p < 0.01$

Discussion

Results of this research indicate that, as predicted, participants with a personal history of psychopathology are more likely to display any elevation on the clinical scales of the MMPI-2 than individuals without such a history. Furthermore, the results indicate, as predicted, that individuals with a personal history of drug or alcohol abuse are more likely to have an elevated score of the MMPI-2 AAS than individuals without such a history. The other hypotheses posed were not supported by the data.

Both the odds ratio and the correlation of the relationship between personal history of psychopathology and any clinical scale elevation were statistically significant. Similarly, the odds ratio, correlation, and t-test of the relationship between personal history of drug or alcohol abuse and an elevated score on the AAS were statistically significant. There were other instances, however, when the different methodologies resulted in different conclusions.

One of the purposes of this study was to demonstrate that exclusive reliance on correlational and group-differences methods of establishing validity might lead to different conclusions than those reached by epidemiological methods. To illustrate this, the relationship between personal history of drug or alcohol abuse and an elevated APS score and the relationship between personal history of drug or alcohol abuse and an elevated FAM score were investigated. The odds ratio between personal history of drug or alcohol abuse and elevated APS score was statistically significant, but the correlation and t-test of that relationship resulted in non-significant findings. Conversely, the correlation

between personal history of drug or alcohol abuse and an elevated FAM score was statistically significant, while the odds ratio and t-test were not significant.

Generally speaking, the sensitivities of gross inspection for any elevation and AAS were higher than the specificities for each of those measures. For the FAM and APS scales, the sensitivities were lower than the specificities. Neither sensitivity nor specificity was particularly high for any measure. Overall, the NPV was much stronger for all the measures than was the PPV. The higher specificities and NPV's suggest that these scales are better at detecting people who do not have the disorder than at detecting those who do have the disorder. The high NPV and high specificities could have been a result of a problem with the scales' ability to detect personal or familial history of drug or alcohol abuse or psychopathology or could have been a result of a low base rate for each of those characteristics in the sample. Had this research utilized a larger sample size, the cause of the high NPV, high specificities, low PPV, and low sensitivities could have been better determined. Additionally, the MMPI-2 profiles in this research were not evaluated for profile validity. Some of the profiles used may have been invalid, thus affecting the results of the study.

Regardless of the potential problems, this research is valuable in demonstrating the use of alternative methods for establishing personality test validity. The use of epidemiological methods to establish test validity provides researchers with a procedure that is better matched to how tests are used in clinical practice. As was stated earlier, epidemiological methods provide results that are neither inferior nor superior to the results obtained from the more

traditional methods of data analysis such as correlation or t-tests; the results provided by epidemiological methods supply answers to a different set of questions than traditional methods. Traditional methods indicate the amount of shared variance between variables or determine whether or not the correlation between variables is zero in the population. That type of information is certainly valuable to the field of psychology, but not of primary importance to the clinician. Epidemiological methods are better suited to indicate the likelihood of the one characteristic based on the existence of another characteristic. It is this type of information that is most valuable to the clinician.

Researchers in the area of test validity should incorporate epidemiological methods into their research. Using both the traditional and epidemiological methods in data analysis adds little work and makes the research more practical to a wider audience by demonstrating the degree of relationship between variables and those variables' ability to predict behavior.

References

Bartol, C. R. (1991). Predictive validation of the MMPI for small-town police officers who fail. Professional Psychology: Research and Practice, 22, 127-132.

Bland, M. & Altman, D. G. The odds ratio. British Medical Journal, 320, 1468-1469.

Butcher, J. N., Graham, J. R., & Ben-Porath, Y. S. (1995). Methodological problems and issues in MMPI, MMPI-2, and MMPI-A research. Psychological Assessment, 7, 320-329.

Coleman, F. L., & Frick, P. J. (1994). MMPI-2 profiles of adult children of alcoholics. Journal of Clinical Psychology, 50, 446-454.

Cohen, J. (1983). The Cost of Dichotomization. Applied Psychological Measurement, 7, 249-253.

Craig, R. J., & Olson, R. (1992). Personality factors in Psy.D. students: Empirically derived comparisons between men and women. Psychological Reports, 70, 763-770.

Dahlstrom, W. G., & Dahlstrom, L. (Eds.). (1980). Basic Readings on the MMPI: A New Selection on Personality Measurement. Minneapolis, MN: University of Minnesota Press.

Farrington, D. P., & Loeber, R. (2000). Some benefits of dichotomization in psychiatric and criminological research. Criminal Behaviour and Mental Health, 10, 100-122.

Faurie, W. C. (1990). Prediction of length of hospitalization of adolescent psychiatric inpatients utilizing the Pd scale of the MMPI and demographic data. Adolescence, 25, 305-310.

Gordis, L. (1996). Epidemiology. W.B. Saunders Company: Philadelphia.

Herkov, M. J., & Myers, W. C. (1996). MMPI profiles of depressed adolescents with and without conduct disorder. Journal of Clinical Psychology, 52, 705-710.

Howell, D. C. (1997). Statistical methods for psychology (4th ed.). Boston: Duxbury Press.

Hackney, G. R., & Ribordy, S. C. (1980). An empirical investigation of emotional reactions to divorce. Journal of Clinical Psychology, 36, 105-110.

Malinchoc, M., Offord, K. P., Colligan, R. C., & Morse, R. M. (1994). The common alcohol logistic-revised scale (CAL-R): A revised alcoholism scale of the MMPI and MMPI-2. Journal of Clinical Psychology, 50, 436-445.

National Institute of Mental Health (2001). The numbers count: Mental disorders in America. (NIMH Publication NO. 01-4584).

Patalano, F. (1998). MMPI diagnostic classification of substance abusers of the 1970s. Psychological Reports, 82, 505-506.

Trief, P. M., & Yuan, H. A. (1983). The use of the MMPI in a chronic back pain rehabilitation program. Journal of Clinical Psychology, 39, 46-53.

Turner, J. A., Herron, L., & Weiner, P. (1986). Utility of the MMPI pain assessment index in predicting outcome after lumbar surgery. Journal of Clinical Psychology, 42, 764-769.

Velasquez, R. J., Callahan, W. J., & Young, R. (1993). Hispanic-white MMPI comparisons: Does psychiatric diagnosis make a difference? Journal of Clinical Psychology, 49, 528-534.

Walsh, W. B. & Betz, N. E. (1995). Tests and Assessment (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

APPENDIX A

Data Summary Sheet

Sex: M F **Age_____** **Race_____**

1. Which of the following best describes your family of origin (from birth to age 18)?

___two parent home, parents married and living together

___ two parent home, one parent and one stepparent

___single parent home, one parent deceased

___single parent home, parents never married

___single parent home, parents divorced

___reared by family members other than parents

___other (please describe below)

2. Please rate the frequency and degree of conflict with parents or other family members on the scale below. Circle the number that is most descriptive.

Frequency		Degree	
1 –	no conflict	1 –	no conflict
2 –	rare/infrequent conflicts	2 –	mild conflicts
3 –	moderate amount of conflicts	3 –	moderate conflicts
4 –	frequent conflicts	4 –	intense conflicts
5 –	constant conflicts	5 –	abusive conflicts

3. Is there a history of alcohol or drug abuse, or treatment for alcohol or drug abuse, among members of your family of origin?

YES____ NO_____

If yes, please describe_____

4. Do you have a history of alcohol or drug abuse, or treatment for alcohol or drug abuse?

YES_____ NO_____

If yes, please
describe_____

5. Is there a history of psychological disorder among any members of your family of origin?

YES_____ NO_____

If yes, please
describe_____

6. Do you have a history of psychological disorder?

YES_____ NO_____

If yes, please
describe_____

APPENDIX B

Formulae for Calculation of Validity Measures

2x2 Contingency Table

Test Results	Condition: positive	Condition: negative
positive	A True Positives (TP)	B False Positives (FP)
negative	C False Negatives (FN)	D True Negatives (TN)

Formulae

Odds Ratio

$$\frac{TP \times TN}{FP \times FN} \quad \text{or} \quad \frac{A \times D}{C \times B}$$

Sensitivity

$$\frac{TP}{TP + FN} \times 100 \quad \text{or} \quad \frac{A}{A + C} \times 100$$

Specificity

$$\frac{TN}{TN + FP} \times 100 \quad \text{or} \quad \frac{D}{D + B} \times 100$$

Positive Predictive Value

$$\frac{TP}{TP + FP} \times 100 \quad \text{or} \quad \frac{A}{A + B} \times 100$$

Negative Predictive Value

$$\frac{TN}{TN + FN} \times 100 \quad \text{or} \quad \frac{D}{D + C} \times 100$$