

Establishing Behavioral Correlates: The MMPI as a Case Study

Samuel B. Green
Auburn University

Two methods have been used to determine MMPI behavioral correlates: the whole-sample and split-sample techniques. The latter technique, which splits a sample into half samples, is currently in use because it ostensibly controls more adequately for Type I errors by requiring significance in both half samples. By simply adjusting the level of significance for the whole-sample approach, however, it can control for Type I errors as well as the split-sample technique. Furthermore, the whole-sample approach appears to control more adequately for Type II errors than the split-sample technique. Finally, data are presented suggesting that as an indicator of the average result obtained by repeated random splittings of a sample, results from the whole-sample approach are preferable to those from a single split of a sample. The whole-sample method was thus recommended; however, the probability of a Type I error for each correlate tested should be set at values smaller than those currently used.

Derivation of empirical correlates of Minnesota Multiphasic Personality Inventory (MMPI) codetypes has typically involved the comparison of individuals with similar MMPI profiles against the remainder of the sample on a large number of behaviors. However, some researchers have made these comparisons with split samples (e.g., Gynther, Altman, & Sletten, 1973; Lewandowski & Graham, 1972; Marks,

Seeman, & Haller, 1974), whereas others have made them within the total sample (e.g., Gilbertstadt & Duker, 1965; Marks & Seeman, 1963). The split-sample technique assumes a replication-like approach. It requires that a sample be split into two half samples and that significance be found in both half samples before a behavioral correlate is identified. In contrast to this approach, the whole-sample technique requires significance for only the entire sample. The split-sample technique is believed to be superior to the whole-sample method in controlling for Type I errors (the false rejection of a null hypothesis). The replication-like feature of the split-sample approach is perceived as a means to identify the true relationships between codetypes and behaviors among the many that are tested (Gynther et al., 1973; Marks et al., 1974). The purpose of the paper is to show that the superiority of the split-sample approach is illusory.

Comparison of Type I Error Rates

Researchers have split the sample into two half samples in two ways. First, subjects may be randomly assigned to one of the two subsamples (e.g., Marks et al., 1974). Within each of these subsamples, the probability of committing a Type I error (α) is set equal to .10 so that the probability of falsely rejecting the null hypothe-

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 6, No. 2, Spring 1982, pp. 219-224
© Copyright 1982 Applied Psychological Measurement Inc.
0146-6216/82/020219-06\$1.30

sis in both subsamples for any potential behavioral correlate is .01 ($.10 \times .10 = .01$). Alternatively, a sample may be nonrandomly divided into two subsamples on the basis of such variables as date when patient was admitted (Gynther et al., 1973). When α is set equal to .10 within each half sample, the probability of a Type I error occurring in both subsamples may be as small as .01, but as large as the α (.10) for each half sample. On the other hand, with the whole-sample method, the probability of a Type I error for any potential profile-behavior correlate is simply set for the total sample.

Consequently, a researcher who sets α equal to .01 using the whole-sample approach is being at least as conservative in terms of Type I errors as one who uses the split-sample method with α equal to .10 for each half sample. However, when the whole-sample approach has been applied, researchers (e.g., Gilberstadt & Duker, 1965; Marks & Seeman, 1963; Marks et al., 1974) have chosen higher levels of α , most typically .05. In contrast, when the split-sample approach has been chosen, researchers (Gynther et al., 1973; Lewandowski & Graham, 1972; Marks et al., 1974) have set α equal to .10 for each split sample, yielding an overall α as low as .01. Therefore, the perceived conservatism of the split-sample approach may be attributable to the α levels chosen by researchers, rather than to any inherent characteristic associated with these techniques.

Comparison of Type II Error Rates

A second explanation exists for the view that the split-sample technique controls more adequately for Type I errors. Because the cause of nonsignificance for any test can not be known, the fewer significant behavioral correlates found by this approach may have been erroneously attributed to only a greater control of Type I errors rather than, at least partially, a result of a poorer control of Type II errors (the nonrejection of a null hypothesis when it should be rejected). Since researchers have most frequently

used the chi-square test of association between two dichotomous variables to establish behavioral correlates, this test statistic was chosen to evaluate the relative power of the split-sample and whole-sample techniques.

The χ^2 for the test of association may be obtained by multiplying the sample size, N , by the degree of association between the two dichotomous variables, ϕ^2 . Because the ϕ^2 in a half sample is, on the average, approximately the same value as the ϕ^2 for the total sample, the χ^2 for half of a sample is on the average *half* the magnitude of the χ^2 for the whole sample. Therefore, if the alphas for the two techniques are the same, the split-sample technique tends to be less powerful than the whole-sample approach because the former not only requires significance in both half samples but also yields much smaller χ^2 values. Of course, this loss in power can be offset by setting the alphas associated with the split-sample approach disproportionately larger than the alpha for the whole-sample method. In the following examples, the split-sample technique clearly leads to a loss of power, even when the two methods are equated in terms of Type I error.

Based on Hedlund's (1977) findings, it is hypothesized that the relationship between a code-type and a behavior is weak, e.g., $\phi = .20$. In order that the two techniques control equally for Type I errors, α is set at .01 for the whole-sample technique and at .10 for each of the subsamples for the split-sample approach, assuming independence between the subsamples. Based on Cohen's (1977) power tables, if the total sample size is 400, the power, $1 - \beta$, for the whole-sample method is .92 and, consequently, β (the probability of a Type II error) is equal to .08. For half the sample and $\alpha = .10$, the power is .88. Since significance must be found in both half samples, the power for the split-sample technique is the product of the powers for each, $.88 \times .88 = .77$. Consequently, the probability of a Type II error for the split-sample technique is .23 ($1 - .77 = .23$), considerably larger than the β found for the whole-sample method, .08. If ϕ is

.10 rather than .20, the betas for the two methods are much larger: .83 and .72 for the split-sample and the whole-sample techniques, respectively. However, the advantage of the whole-sample approach in terms of control of Type II error remains.

Sampling Error Due to Sample Splitting

The split-sample technique has another disadvantage. With both the whole- and the split-sample approaches, sampling error results to the degree that the sample does not represent the population. However, with the latter technique, sampling error also occurs in the process of splitting the sample. Therefore, the researcher should not rest his/her conclusions on a single arbitrary split of the sample. If the split-sample technique is to be used, multiple splits of the sample should be taken. Ideally, to reduce this second type of error to zero, all possible splits should be made. However, it is likely that the whole-sample result reflects the average finding from all possible split samples. Consequently, only the whole-sample approach need be applied on a set of data.

In order to demonstrate this phenomenon, data collected by Kelley and King (1979) were reanalyzed. They used both the whole-sample and split-sample approaches to establish which of 175 dichotomous behaviors were related to the 2-7-8 MMPI profile. With the whole-sample technique, they set α equal to .01 for each behavior tested. For the split-sample approach, they randomly split the sample into two half samples with the restriction that each half sample have the same number (± 1) of 2-7-8 profiles and the same number (± 1) of other profiles. Given this procedure, the control over Type I error was comparable with the two approaches. Kelley and King found that the MMPI profile variable was significantly related to 22 behaviors with both methods, to 16 behaviors with only the whole-sample approach, and to 7 behaviors with only the split-sample approach. As expected, the split-sample technique obtained fewer significant results.

Rather than reanalyzing all 175 behaviors, four behaviors were randomly selected from each of the following four categories: (1) correlates found significant using both methods, (2) correlates found significant using only the whole-sample method, (3) correlates found significant using only the split-sample method, and (4) nonsignificant correlates using both methods. For each of the 16 behaviors, the following steps were performed.

1. A whole-sample chi-square was calculated on a 2×2 contingency table defined by the profile and behavior dimensions.
2. Each of the splits of this table that yielded a unique set of cell frequencies was identified. Following the method outlined by Kelley and King (1979), an equal number of individuals with 2-7-8 profiles and an equal number with other profiles were in each half sample. Given a 2×2 table such as the one shown in Figure 1, with the letters representing frequencies based on the whole sample, the number of unique tables produced by splitting the sample in this fashion is equal to $(A + 1)$ or $(B + 1)$, whichever is smaller, times $(C + 1)$ or $(D + 1)$, whichever is smaller.
3. The probability of each of these splits was calculated. The number of ways of splitting the total sample into two halves is the product of two combinatorials:

$$\binom{A + B}{\frac{A+B}{2}} \binom{C + D}{\frac{C+D}{2}} \quad [1]$$

Given that the same letters represent the cell frequencies in a split sample as in a whole sample except that they are noncapitalized, the number of ways of obtaining a particular split is equal to

$$\binom{A}{a} \binom{B}{\frac{A+B}{2} - a} \binom{C}{c} \binom{D}{\frac{C+D}{2} - c} \quad [2]$$

Figure 1
Tabular Presentation of Profile-Behavior Relationship
with Letters Representing Frequencies

| | Behavior Present | Behavior Absent |
|---------------|------------------|-----------------|
| 2-7-8 Profile | A | B |
| Other Profile | C | D |

The probability of each possible split is obtained by dividing Equation 2 by Equation 1.

- For each split with a unique set of cell frequencies, chi-squares were calculated for both half samples. It was noted whether both of these chi-squares were significant with α equal to .10.
- The denominators for some of the chi-squares obtained in Step 4 were equal to zero. The splits that produced these results were eliminated and the probabilities of the other splits were rescaled, so that the sum of the probabilities of the reduced set of splits equaled one.
- The probability of obtaining a significant split-sample correlate was found by summing the probabilities associated with the splits that yielded significance in both half samples.

The results confirmed that the split-sample technique tends to be less powerful. For five of the eight behaviors that were found related to the profile type using the whole-sample approach, the probability of rejecting the null hypothesis with the split-sample technique was less than .40.

The zero-order correlation between the chi-square for the whole sample and the probability of obtaining a significant result for the split-sample technique over the 16 behaviors was .96. The correlation did not increase by adding a squared chi-square term to the nonsquared term in predicting the probability of significance with the split-sample approach; however, a cubic term increased the multiple correlation coefficient to .99. An inspection of the bivariate scatterplot suggested the relationship between the chi-square values and the probabilities may best be described by an ogive curve. For small chi-square values ($\chi^2 < 5.0$), the probabilities were all close to zero; for large chi-square values (none of the χ^2 reached these values) the probabilities would approach one; and, as the whole-sample chi-squares increased in magnitude from small to large values ($5 < \chi^2 < 22$), the probability of a significant split-sample result increased in magnitude.

The practical implication of this relationship may be better understood by comparing the mean probabilities of rejecting the null hypothesis with the split-sample method for the four categories of behaviors. The mean probabilities were the same (.03) for the four behaviors that

were found significant by Kelley and King (1979) using only the split-sample approach and for the four behaviors that were found nonsignificant using both approaches. In contrast, the mean probabilities for those behaviors that were significant using only the whole-sample technique and using both techniques were .38 and .62, respectively. Clearly, if given the choice as to which behaviors to choose as related to the 2-7-8 profile, the ones that are identified as significant with only the whole-sample approach should be chosen over those that are significant using only the split-sample technique.

Three of the four behaviors that were significantly related to the profile variable with only the split-sample method were substantially similar. The three dealt with emotional difficulties of family members including specifically (1) aunts, (2) siblings, and (3) at least one relative. Although the data had been recorded in such a fashion that the relationship between these may not be determined, it is highly likely that these measures are at least moderately related. Any particular split of the sample that produced significance in both half samples for a particular behavior when the probability of significance in both subsamples is small is likely to produce comparable results for related behaviors. In other words, some splits of a sample may produce a relatively large number of identified correlates even if the likelihood for significance in both half samples for each of these behaviors is low because the behaviors are correlated.

Conclusions and Recommendations

The conclusion from comparing the two techniques appears evident. If the two are equated in terms of committing a Type I error, the whole-sample approach is more powerful and a better indicator of results from the split-sample method than a single application of the method itself. This conclusion does not imply rejecting the methods of replication and/or cross-validation, but rather is in line with these methods. Rosenthal (1978) has reviewed a number of ways

to combine results of replicate studies. These methods could be applied to the results from the split-sample technique in order to combine the information obtained from the half samples; they would offer an alternative to the standard combining rule of identifying a behavioral correlate only if the null hypothesis is rejected in both half samples. From this perspective, the author has argued that the χ^2 value from the whole sample is a better indicator of the results from the half samples than this standard rule used with the split-sample technique.

A comprehensive recommendation against the split-sample method is inappropriate. For example, if stepwise regression analysis is used to derive an equation containing only a subset of potential predictor variables, no formulas are available to estimate the cross-validated multiple correlation coefficient. Consequently, a large sample might be randomly split in order to estimate the regression weights in one-half of the sample and to apply the weights to determine the cross-validated correlation coefficient in the other half of the sample (Cattin, 1980). Application of the split-sample technique may be used in general when attempting to estimate cross-validated strength-of-relationship indices when no formulas are appropriate to estimate their magnitude.

The intent of this presentation was to compare the two methods. Within this context, the original purpose for introducing the split-sample approach by researchers has likely been obscured: to minimize the number of spurious behavioral correlates obtained from multiple significance tests. Larzelere and Mulaik (1977) have considered a number of methods to control for Type I errors due to multiple testing. However, if an MMPI researcher wishes to make a statement about particular profile-behavior correlates, only two approaches are relevant: the Bonferroni and the modified Bonferroni methods. Because the latter is a variation of the former and offers nothing conceptually to the discussion, only the former method is considered.

With the Bonferroni approach, a family of significance tests is defined a priori. The familywise error rate (the probability of falsely rejecting one or more null hypotheses associated with a family of tests) is equal to, or greater than, the sum of the alphas of the separate tests. Therefore, if the desired familywise error rate is .05, the α for each test within a family should be set equal to .05 divided by the number of tests.

An MMPI researcher might define a family to include only those tests that evaluate the relationships between a particular profile and those behaviors that should relate to it on the basis of the scale content associated with that profile. Assuming a familywise error rate of .05, and that five behaviors have been identified a priori to be associated with a profile, the α for each of the tests should be set equal to .01. The restriction of a family to include only five tests would be seen by most MMPI researchers as inappropriate in that the resulting profile interpretation would offer the clinician a very limited descriptive picture of his/her client. However, to control for familywise error rate with a larger number of tests would require the α for the separate tests to be quite small, certainly less than .01. Under these conditions, large sample sizes would be required in order to assure the researchers sufficient power to establish valid correlates. None of the researchers who have published MMPI codetype systems have set the alphas to be less than .01 for the separate behaviors, probably because they believed their sample sizes to be too small to identify behaviors that are truly related to the codetypes of interest.

In summary, the researcher who wishes to establish behavioral correlates should obtain larger samples and make significance tests using the whole sample at reduced α values. Although the split-sample approach may initially be seen as a reasonable method to reduce the number of spurious behavioral correlates, as it has been applied with raised alphas for the half samples, the whole-sample procedure with reduced alphas

actually should generate fewer of them. Consequently, no justification exists for the use of the split-sample technique, given its other undesired properties.

References

- Cattin, P. Estimation of the predictive power of a regression model. *Journal of Applied Psychology*, 1980, 65, 407-414.
- Cohen, J. *Statistical power analysis for the behavioral sciences*. New York: Academic Press, 1977.
- Gilberstadt, H., & Duker, J. A. *A handbook for clinical and actuarial MMPI interpretation*. Philadelphia: W. B. Saunders, 1965.
- Gynther, M. D., Altman, H., & Sletten, I. W. Replicated correlates of MMPI two-point code types: The Missouri actuarial system. *Journal of Clinical Psychology Monograph*, 1973, 29, 263-289.
- Hedlund, J. L. MMPI clinical scale correlates. *Journal of Consulting and Clinical Psychology*, 1977, 45, 739-750.
- Kelley, C. K., & King, G. D. Behavioral correlates of the 2-7-8 MMPI profile type in students at a university mental health center. *Journal of Consulting and Clinical Psychology*, 1979, 47, 679-685.
- Larzelere, R. E., & Mulaik, S. A. Single-sample tests for many correlations. *Psychological Bulletin*, 1977, 84, 557-569.
- Lewandowski, D., & Graham, J. R. Empirical correlates of frequently occurring two-point MMPI code types: A replicated study. *Journal of Consulting and Clinical Psychology*, 1972, 39, 467-472.
- Marks, P. A., & Seeman, W. *The actuarial description of abnormal personality: An atlas for use with the MMPI*. Baltimore: Williams & Wilkins, 1963.
- Marks, P. S., Seeman, W., & Haller, D. L. *The actuarial use of the MMPI with adolescents and adults*. Baltimore: Williams & Wilkins, 1974.
- Rosenthal, R. Combining results of independent studies. *Psychological Bulletin*, 1978, 85, 185-193.

Acknowledgments

I thank Malcolm Gynther for his helpful suggestions concerning the manuscript and Glen King and Crystal Kelley for the use of their data presented in this article. Send requests for reprints or further information to Samuel B. Green, Department of Psychology, Auburn University, Auburn AL 36849.