Analysis and Interpretation of Findings Using Multiple Regression Techniques

William T. Hoyt
University of Wisconsin–Madison
Stephen Leierer
University of Memphis
Michael J. Millington
Abita Springs, LA

Multiple regression and correlation (MRC) methods form a flexible family of statistical techniques that can address a wide variety of different types of research questions of interest to rehabilitation professionals. In this article, we review basic concepts and terms, with an emphasis on interpretation of findings relevant to research questions of interest to rehabilitation researchers. To assist readers in using MRC effectively, we review common analytical models (e.g., mediator and moderator tests) and recent thinking on topics such as interpretation of effect sizes and power analysis.

In the nearly 40 years since the publication of Jacob Cohen's (1968) seminal article heralding multiple regression as a "general data-analytic system," multiple regression and correlation (MRC) techniques have become increasingly popular in both basic and applied research journals. This is also true of the journal *Rehabilitation Counseling Bulletin:* In our survey of five complete volumes (2000 through 2004), we found 29 articles in which some form of MRC analysis (e.g., simultaneous multiple regression, hierarchical regression, stepwise regression, logistic regression, or simple correlation) was used to test the research hypothesis. This represents 34% of the 83 articles published in these five volumes that reported some form of statistical analysis. A similar frequency was observed in a survey of recent issues of *Rehabilitation Psychology*. Clearly, researchers in rehabilitation counseling and rehabilitation psychology regard MRC techniques as an important research tool.

The purpose of this article is to review best practices for researchers using MRC. We assume that readers have some familiarity with MRC techniques. So, although we review basic terminology and procedures, we refer those interested in a more detailed treatment of fundamentals to other sources (e.g., Cohen, Cohen, West, & Aiken, 2003; Wampold & Freund, 1987). We focus on application of MRC techniques for testing hypotheses relevant to rehabilitation psychology and on conceptual and interpretational issues that have the potential to confound researchers making use of these techniques.

Regression Models

A major reason that MRC techniques are so attractive to researchers is their flexibility: MRC may be used to test hypotheses of linear or curvilinear associations among variables, to examine associations among pairs of variables controlling for potential confounds, and to test complex associations among multiple variables (such as mediator and moderator hypotheses). Predictor variables in multiple regression analyses may be correlated with one another, and they may be continuous, categorical, or a combination of the two. In fact, ANOVA and ANCOVA can be regarded as special cases of MRC in which categorical predictor variables are of primary interest, although continuous covariates may also be included (Cohen,

pro·ed

1968). Although bivariate (i.e., single-predictor) regression and correlation are frequently useful for assessing associations among pairs of variables, the analytical power of regression analyses is greatly enhanced when multiple predictor variables are studied. In this section, we describe three common models for multiple regression analyses. These models are distinguished by how predictors are entered into the regression equation: simultaneously, hierarchically (in an order predetermined by the investigator), or empirically (with the order of entry determined by which variables contribute most or least to prediction at a given step in the regression equation).

Simultaneous Regression

The basic application of multiple regression involves simultaneous use of a set of predictor variables to make the most accurate prediction possible of scores on the criterion variable (DV). This analysis provides information about variance in the DV accounted for by the predictors as a set and also the unique association of each predictor with the DV when all of the other predictors in the regression analysis are statistically controlled.

Example 1: Two Predictors of Worker Satisfactoriness. Millington, Leierer, and Abadie (2000) examined role-play employers' attitudes toward written descriptions of job applicants on the *Employment Expectations Questionnaire* (revised; EEQ-B) as predictors of these same employers' expectations about applicant job performance. For simplicity, we initially consider the first two EEQ-B dimensions as predictor variables (X_1 = Job Knowledge and Skill; X_2 = Socialization and Emotional Coping Skills), with the employers' predictions of the applicants' job satisfactoriness (Y) as the dependent variable. In this example, X_1 and X_2 are factor scores with a theoretical range of −5.0 to 5.0; in the present sample, M = 1.74 and −0.35; *SD* = 1.64 and 1.92, respectively. The Worker Satisfactoriness Scale (WSS) in this study had a range of 0 to 100, in 10-point intervals (0, 10, 20, etc.), with *M* = 49.85 and *SD* = 19.55 in the present sample.

TABLE 1. Predicting Worker Satisfactoriness From EEQ-B Subscales: Two-Predictor and Five-Predictor Models

Variable	*B*	*SE B*	β
Two-predictor model			
Constant	43.91	1.57	
Job Knowledge	3.98	0.66	.33*
Socialization	2.86	0.56	.28*
Five-predictor model			
Constant	39.96	1.75	
Job Knowledge	1.43	0.87	.12
Socialization	2.21	0.56	.22*
Trainability	−0.27	0.71	−.02
Dependability	2.73	0.73	.24*
Motivation	1.79	0.80	.16*

Note. EEQ-B = *Employment Expectations Questionnaire* (revised; Millington, Leierer, & Abadie, 2000). For the two-predictor model, $R^2 = .29$, $F(2, 316) = 63.11$, $p < .001$; for the five-predictor model, $R^2 = .34$, $F(5, 313) = 32.24$, * $p < .001$.

To address the question of how well these two EEQ-B dimensions predict WSS scores, we can use a statistical application, such as SPSS, to regress the criterion variable (Y) onto the two predictor variables (X_1 and X_2). Using a least squares algorithm, which minimizes the sum of the squared errors of prediction (called *residuals*) across all cases in the sample, the software application outputs the optimal regression equation for predicting Y scores from scores on X_1 and X_2 for this sample. This equation takes the form

$$\hat{Y} = B_1X_1 + B_2X_2 + B_0$$

and can be used to compute a predicted score $\hat{Y}$ on the criterion variable for any person from the population whose scores on X_1 and X_2 are known. The *regression coefficients* B_1 and B_2 are the multipliers for X_1 and X_2, respectively, to be used in computing the predicted score. The third regression coefficient (B_0) is called the *constant* or the *intercept;* it denotes the predicted value of Y for a person with scores $X_1 = X_2 = 0$.

When X_1 and X_2 are theorized to be causally prior to Y (as here), coefficients B_1 and B_2 are interpreted in terms of the causal impact of the predictor on the criterion (see Note 1), or the predicted change in Y for a one-unit change in X_1 or X_2. When multiple predictors are included, B_1 and B_2 are *partial* regression coefficients, each indicating the causal effect of one predictor on Y, with the other predictor partialed out (i.e., statistically controlled). The interpretation of partial regression coefficients is discussed below, in the section on *Effect Sizes in Multiple Regression*.

In our example (see Table 1), we obtained the regression equation

$$\hat{Y} = (3.98)X_1 + (2.86)X_2 + 43.91$$

This tells us that a 1-point increase in perceived Job Knowledge (X_1) is expected to produce an increase of 3.98 points in employee satisfactoriness (Y) when perceived Socialization (X_2) is statistically controlled (i.e., held constant). By comparison, when Job Knowledge is held constant, a 1-point increase in Socialization (X_2) yields a predicted 2.86-point increase in satisfactoriness. The intercept in this equation (B_0 = 43.91) is the predicted Y score for a person scoring 0 on both X_1 and X_2.

SPSS also outputs a standard error, t statistic, and p value for each regression coefficient. If the p value is less than the designated alpha level for the study (e.g., $p < .05$), the regression coefficient differs significantly from zero, indicating a significant association between the designated predictor and the criterion variable, controlling for the remaining predictors. In this rather large sample ($N = 319$), both X_1 and X_2 are significant predictors of Y (both $p < .001$).

A final item of interest (especially when the goal of the study is to predict as much variance as possible in the criterion) is the multiple correlation ($R = .53$), representing the correlation between the predicted scores $\hat{Y}$ and the actual scores (Y) on the criterion variable; its square, $R^2 = .29$, $F(2, 316) = 63.11$, $p < .001$, is interpreted as the proportion of variance in Y that is accounted for by the predictor variables as a set. The significance test for R^2 is an F test (which is identically also a significance test for R); when the associated p value is less than the designated critical value (e.g., $p < .05$), the multiple correlation coefficient differs significantly from zero, indicating a significant association between the predictors *as a set* and Y.

Example 2: Five Predictors of Worker Satisfactoriness. In point of fact, the EEQ-B has five subscales denoting aspects of employability. (In addition to X_1 and X_2, described above, these are X_3 = Trainability/Flexibility; X_4 = Dependability; and X_5 = Motivation.) Table 1 (bottom section) shows how the regression of Y onto all five predictors compares with the simpler two-predictor model just discussed.

First, from the note to Table 1, we see that the three additional predictors increase the proportion of Y variance explained: $R^2 = .29$ for the two-predictor model, as compared with .34 for the five-predictor model. Second, notice that in the five-predictor model (unlike the simpler model), not all of the X variables contribute uniquely to prediction of scores on the WSS. In the five-predictor model, only Socialization, Dependability, and Motivation emerge as significant predictors of satisfactoriness.

This finding illustrates an important consideration in interpreting results of MRC analyses, namely that the association between a predictor and a given DV depends on the other predictors included in the regression equation. The Job Knowledge subscale (X_1) is a significant univariate predictor of Worker Satisfactoriness ratings ($r_{Y1} = .48$, $p < .05$), and it is also significant in the two-predictor model, controlling for X_2 ($\beta_{Y1 \cdot 2} = .33$, $p < .05$). In the five-predictor model, however, X_1 does not contribute significantly to predicting Y when the other four EEQ-B subscales are statistically controlled ($\beta_{Y1 \cdot 2345} = .12$, $p > .05$). This reduction in the standardized regression coefficient β (which corresponds to similar decreases in the unstandardized coefficient B) as additional predictor scales are added to the model reflects the intercorrelations among EEQ-B subscales, or, to put it another way, the *redundancy* of the information provided by X_1 ratings with that provided by ratings on the other four subscales. When all five EEQ-B dimensions are included, the unique (nonredundant) information provided by X_1 does not contribute significantly to our ability to predict WSS scores.

Summary. Simultaneous regression yields information about the joint association of a set of predictor variables with Y (multiple R^2 and associated significance test) and about the unique association of each predictor X_i with Y, when all other predictor variables are statistically controlled (B_i or β_i and the associated significance test). Because correlations among predictor variables are the rule in nonexperimental research, the interpretation of the regression coefficient (B or β) is relative to the other predictors included in the regression equation; variables that are significant predictors of the DV in one analysis may become nonsignificant in subsequent analyses if additional, overlapping predictor variables are added.

Hierarchical Regression

In hierarchical regression analysis (HRA), predictor variables are entered sequentially in two or more sets, with the groupings and order of entry predetermined by the investigator. Nielsen (2003) used HRA to determine whether social support added significantly to the prediction of posttraumatic stress disorder (PTSD) among 168 adults with spinal cord injury, over and above the variance predicted by demographic and injury-related variables (gender, age, education, marital status, time since injury, loss of consciousness as a result of injury, and neurological level).

Nielsen (2003) entered the demographic and injury-related variables to be statistically controlled as the first block in the HRA. The results for this first block are identical to those for a simultaneous regression of PTSD scores onto these seven predictor variables. As a set, these variables were significantly related to PTSD symptoms: $R^2 = .10$, $F(7, 160) = 2.6$, $p < .05$. Regression coefficients for two of the predictors in this set differed significantly from zero: marital status (married persons were less likely to report PTSD symptoms) and neurological level (higher neurological functioning predicted fewer PTSD symptoms).

Nielsen (2003) incorporated two social support scales, one measuring total quantity of support and one assessing satisfaction with support. She entered these two variables as a second block in the HRA. At this second step, the new predictors are added, and all of the original predictors remain in the predictor set. Thus, PTSD scores are simultaneously regressed onto demographic variables, injury-related variables, and the two social support scores;

however, R^2 for these nine variables as a set is not the focus of interpretation. Instead, Nielsen examined the *change* in R^2 (ΔR^2) from Block 1 to Block 2. In this case, the additional variance explained when the Block 2 variables were added to the predictor set was both substantial and statistically significant: $\Delta R^2 = .19$, $F(2, 158) = 21.0$, $p < .001$. This significant increment to the variance accounted for by the prediction model affirms that the social support measures, *as a set*, contribute significantly to the prediction of PTSD scores, over and above the demographic and injury-related predictor variables.

In addition, it is appropriate to examine the regression coefficients for the two social support scales to see their relative contributions to predicting PTSD. Nielsen found that the total support score was a significant unique predictor, but overall satisfaction was not significantly related to PTSD, controlling for the other eight predictors in the equation at Block 2.

Summary. In the social sciences, variables of interest are not always capable of experimental manipulation, for either ethical or practical reasons. As noted in the previous section, a central problem for disentangling causal relations among measured (as opposed to experimentally manipulated) variables is the issue of *correlated predictors*. When predictor variables are not statistically independent of one another, they will account for overlapping (common) variance as well as unique variance in the criterion variable. HRA allows the investigator to enter individual predictors or sets of predictors in a specified order (in accordance with causal or conceptual priority). In HRA, the initial predictor set gets credit for all of the criterion variance it can account for; the second predictor set gets credit for only the additional criterion variance it uniquely accounts for (beyond that accounted for in Block 1: i.e., ΔR^2). If there is a third predictor set, it gets credit for unique variance accounted for over and above that predicted by Blocks 1 and 2 combined, and so on. Thus, investigators can use HRA to examine the criterion variance uniquely accounted for by a predictor variable (or set) of theoretical interest (such as social support, for Nielsen, 2003), after controlling for potential confounding variables that have a causally prior association with the criterion variable. Other applications of HRA (some of which we discuss in more detail below) include analysis of nominal (i.e., categorical) variables using MRC, testing moderator relations (i.e., statistical interactions among two or more predictors), and assessment of incremental predictive validity from the addition of a new predictor variable to an existing predictor set.

Empirical (Stepwise) Regression

Another option for establishing an order of entry for variables in a hierarchical analysis is to use empirical rather than theoretical criteria. In a stepwise regression analysis, the bivariate association of each predictor variable with the criterion variable is examined, and the variable with the greatest predictive power is entered first. Then the remaining predictors are assessed for their incremental predictive validity, and the one that explains the most additional criterion variance (i.e., the one that results in the largest ΔR^2) is added second. This procedure is repeated until no further predictors would result in a significant ΔR^2, at which point the final predictor set (which normally contains only a subset of the possible predictor variables) is regarded as definitive. Most statistical software packages include several variants on this procedure (e.g., step-up, step-down), which automate the process of selecting variables for inclusion in the regression equation.

Empirical regression methods may be appealing because they relieve the researcher of having to make theory-based decisions about the order of entry of predictor variables (Cohen et al., 2003). Indeed, although only one study published in *Rehabilitation Counseling Bulletin* in the 5-year period we surveyed used stepwise regression analyses, this technique was much more common in *Rehabilitation Psychology:* In our survey of four consecutive issues (1 year) of this journal, we found that 5 of the 16 articles using MRC used empirical regression methods.

The consensus in the methodological community is that stepwise regression should be used very rarely (see Cohen et al., 2003, p. 162) or not at all (Thompson, 1995) in psychological research. The main critique of stepwise methods is that they yield a so-called optimal predictor set that is very unlikely to generalize to future samples. We recommend that rehabilitation researchers familiarize themselves with the weaknesses of stepwise methods and avoid using these procedures, substituting either simultaneous regression or HRA.

Effect Size in MRC

Following a decade of increasingly strident criticism of psychology's reliance on *p* values as summaries of study findings (see Kline, 2004, pp. 6–17, for a brief historical overview), the American Psychological Association convened a Task Force on Statistical Inference and charged it with making recommendations about research design and interpretation for the new century. Among many helpful recommendations in the report of this task force (Wilkinson & the Task Force on Statistical Inference, 1999) is the exhortation to "always present effect sizes for primary outcomes" (p. 599).

An *effect size* quantifies the magnitude of association between two (or more) variables. Effect sizes tell readers more than simply that "*X* is significantly related to *Y*"; they indicate both the strength and the direction of the relationship. Users of MRC are fortunate to have available a variety of effect size indices among which to choose. In this section, we highlight the most commonly presented effect size indices and discuss how each is interpreted.

Effect Sizes in Bivariate Regression

By definition, bivariate regression analysis involves one predictor variable and one criterion variable. There are two possible effect sizes that can be reported from such an analysis: unstandardized and standardized.

Unstandardized Regression Coefficient. The brief summary of MRC given previously focused on what is properly called the *unstandardized* regression coefficient, or B_1. When Y is regressed onto a single predictor variable (X_1), B_1 tells the predicted change in Y (or, equivalently, the change in Y for a one-unit change in X_1. Geometrically, it represents the slope of the Y-on-X_1 regression line, when X_1 and Y are scaled in their original (raw score) units. In causal terms, we can think of B_1 as reflecting the causal impact (or *effect*) on Y of a one-unit increase in X_1.

Consider the Millington et al. (2000) study described in Examples 1 and 2 in the Regression Models section. If Worker Satisfactoriness (Y) is regressed onto Job Knowledge (X_1) as the sole predictor variable, we obtain a regression coefficient of $B_{Y1} = 5.69$. This represents a predicted increase of 5.69 satisfactoriness points for every 1 point gained in ratings of Job Knowledge. So, if a training program were created to enhance job-specific knowledge and skills, and that program led to average gains of 2 points in X_1 scores, we would expect an indirect effect on WSS ratings of 2(5.69), or about 11 points.

Standardized Effect Sizes. Note that in this example, if readers are unfamiliar with either or both measures (EEQ-B or WSS), the unstandardized regression coefficient B_{Y1} may not be very meaningful. It is difficult at a glance to tell whether a 5.69-unit increase on Y for each unit increase on X_1 is a large or important effect. When the units of measurement on one or both variables are not readily interpretable, Wilkinson and the Task Force on Statistical Inference (1999) recommend reporting *standardized* coefficients. A *standardized regression coefficient* is the regression coefficient that would be obtained if we first transformed Y and X_1 into their respective z scores (z_Y and z_1; see Note 2) and then regressed z_Y onto z_1. We can convert the unstandardized regression coefficient B_{Y1} into the equivalent standardized regression coefficient (denoted as β_{Y1}) by multiplying it by the ratio sd_1/sd_Y. Thus,

$$\beta_{Y1} = B_{Y1}\frac{sd_1}{sd_Y} = 5.69\left(\frac{1.64}{19.55}\right) = .48$$

Standardized regression coefficients are scaled identically to the Pearson r (i.e., $-1 \le \beta_{Y1} \le 1$), with large positive (or large negative) coefficients indicative of a strong relation between X_1 and Y. In fact, for bivariate regression, β_{Y1} is identical to the Pearson product-moment correlation (r_{Y1}), and each can be interpreted as the predicted change in Y, in standard deviation (*SD*) units, for a 1-*SD* change in X_1. That is, if two people differ by 1 *SD* (1.64 points) in their Job Knowledge scores, we expect a difference of about 0.48 *SD*s (or about 9.4 points) in their satisfactoriness ratings. Because the coefficient is positive, we expect that the person scoring higher on the EEQ-B will be rated higher on the WSS.

A second use of the Pearson r as an aid to interpreting magnitude of association is to square the correlation coefficient and interpret it as an index of variance accounted for. In this example, $r^2 = (.48)^2 = .23$. Thus, job knowledge ratings account for 23% of the variance in WSS scores.

When the units of measurement are meaningful on a practical level (e.g., number of cigarettes smoked per day), it is usually preferable to report an unstandardized measure (regression coefficient or mean difference) rather than a standardized measure (r or d; Wilkinson & the Task Force on Statistical Inference, 1999). When reporting an effect size, it is also helpful to add brief comments (e.g., comparisons with effect sizes obtained in related investigations) to assist readers in gauging the practical and theoretical importance of this association.

Effect Sizes in Multiple Regression

Multiple regression is often used to examine the (presumed causal) effects of correlated predictors on a DV. When two or more predictors are examined simultaneously, the coefficients in the regression equation are termed *partial* regression coefficients, to reflect the fact that in determining the predicted effect of each variable on the DV, the effect of each of the other predictor variables is held constant, or *partialed out*.

Meaning of Partial Coefficients. To illustrate, consider Example 1 in the Regression Models section, in which we simultaneously regressed WSS scores onto X_1 and a second EEQ-B subscale (X_2 = Socialization and Emotional Coping Skills). In this example, we are interested in the joint effects of Job Knowledge (X_1) and Socialization (X_2) on Worker Satisfactoriness (Y). Because X_1 and X_2 are correlated ($r_{12} = .50$), the squared bivariate correlation of each of these predictors with satisfactoriness (i.e., r^2_{Y1}or r^2_{Y2}) reflects a combination of unique variance shared with Y and common variance shared with both the other predictor and Y. To assess the unique contribution of each variable to predicting Y, we used multiple regression, constructing a least squares regression equation for predicting Y from scores on both X_1 and X_2 (see Table 1).

Because X_1 and X_2 are correlated, the partial regression coefficient in Equation 1 is not equal to the bivariate regression coefficient described in the preceding section. This difference is reflected by a change in the notation of the regression coefficient: The bivariate coefficient is denoted as B_{Y1}, whereas the partial coefficient is denoted as $B_{Y1\cdot2}$ (i.e., the regression of Y on X_1 from which X_2 has been partialed). In general, it is expected that $B_{Y1\cdot2}$will be smaller in absolute value (i.e., closer to zero) than B_{Y1} (see Note 3).

The partial regression coefficient $B_{Y1\cdot2}$ is the predicted change in Y for a given change in X_1, when X_2 scores are statistically controlled (i.e., held constant). In reality, if a person's Job Knowledge score (X_1) increases, we expect a corresponding (although somewhat smaller) increase in Socialization (X_2) (because $r_{12} > 0$), with both changes (in X_1 and X_2) producing corresponding changes in Y. To examine the unique effect of X_1 on Y, independent of the common variance shared with X_2, we need to hold the value of X_2 constant and see what happens to Y with a given change in X_1. Although this is not possible in reality, we can accomplish this feat mathematically; this is what is meant when we say that $B_{Y1\cdot2}$ is the (partial) regression coefficient for Y on X_1, *statistically controlling* for X_2.

The partial regression coefficient $B_{Y1\cdot2}$ can also be thought of as the slope of the *partial regression line* when Y is regressed onto X_1 for a sample of individuals who share the same score on X_2. This formulation reminds us of an important assumption underlying this discussion of partial coefficients: These interpretations hold as long as the slope of the partial regression of Y on X_1 is identical for all values of X_2. This assumption can be tested by examining the significance of the X_1-by-X_2 interaction. When there is no significant interaction between X_1 and X_2 in predicting Y, the partial regression of Y on X_1 is independent of X_2, or, equivalently, the slope of the partial regression line when Y is regressed onto X_1 does not depend on the value of X_2. When the X_1-by-X_2 interaction is significant, however, the effect of X_1 on Y differs at different levels of X_2, which is to say that different Y-on-X_1 partial regression lines (i.e., regression lines for different constant values of X_2) have different slopes (see Note 4).

Unstandardized Partial Coefficients. As just noted, the unstandardized partial regression coefficient $B_{Y1\cdot2}$ reflects the predicted change in Y for a one-unit change in X_1 when X_2 is held constant. This quantifies the unique (presumed causal) effect of X_1 on Y, from which its joint effect with X_2 has been partialed. From Table 1, the partial regression coefficient $B_{Y1\cdot2}$= 3.98, about 30% smaller than the corresponding bivariate coefficient (B_{Y1} = 5.69) given earlier. Thus, almost one third of the bivariate association between X_1 and Y is attributable to the overlap between X_1 and X_2. The interpretation of this discrepancy between bivariate and partial coefficients (and the interpretation of partial coefficients more generally) depends on the hypothesized causal model for associations among these variables (see Cohen et al., 2003, pp. 75–79).

Standardized Partial Coefficients. As in the case of bivariate regression, when the units of predictor or criterion variables are not intuitively meaningful, it is usually preferable to report standardized effect sizes. Two standardized effect size measures, β and sr^2, are commonly used to reflect the unique contribution of a single predictor variable in standardized units; a third measure, R^2, reflects the variance accounted for by a set of predictors in MRC.

Standardized partial regression coefficient (β). As in the bivariate case, β_{Y12} can be computed from B_{Y12} by multiplying the latter by the ratio sd_1/sd_Y. Thus,

$$\beta_{Y1\cdot2} = B_{Y3\cdot2}\frac{sd_1}{sd_Y} = 3.98 \cdot \frac{1.64}{19.55} = .33$$

Thus, when Socialization scores are statistically controlled, a 1-*SD* (i.e., 1.64-point) increase in Job Knowledge is predicted to result in a corresponding increase in Y of 0.33 *SD* units (or about 6.5 points). Note that when the *SD*s of the predictor variables differ, their unstandardized regression coefficients are not comparable. The two beta weights, however, are both standardized and directly reflect the relative strength of association. For this example, $\beta_{Y2\cdot1}$ = .28, which is somewhat smaller than $\beta_{Y1\cdot2}$ = .33. This implies that socialization ratings have a slightly smaller unique association with predicted satisfactoriness than job knowledge ratings; however, a statistical significance test (e.g., Azen & Budescu, 2003) should be conducted before strong inferences are made about the relative importance of predictors in MRC.

Squared semipartial correlation (sr^2). When MRC is used for predictive purposes, rather than for analysis of presumed causal associations, investigators may wish to report on the proportion of variance in Y uniquely accounted for by one predictor variable. The semipartial correlation (which some statistical packages refer to by the older name *part correlation*), when squared, is the relevant effect size. Because these are standardized effect sizes, comparison of the squared semipartials gives an indication of the relative unique contributions of different predictors (although again, strong conclusions about differences in predictive validity should be based on significance tests rather than on numerical differences alone). For Example 1, sr_1^2 = .08 and sr_2^2 = .06, again reflecting the slightly larger unique contribution of Job Knowledge, relative to Socialization, in predicting Worker Satisfactoriness.

It is helpful to recall that significance tests for B, β, and sr^2 are identical—if one is statistically significant, the others will be as well. The choice of which effect size to report is based on the nature of the research question (i.e., causal analysis vs. predictive validity) and on whether the units of measurement are inherently meaningful.

Squared Multiple Correlation (R^2). A final effect size commonly reported in multiple regression analyses reflects the proportion of variance in Y accounted for by all of the predictors together, as a set. The multiple correlation coefficient (R) is the correlation between the predicted $\hat{Y}$ scores (computed for each participant in the study using Equation 1) and the actual measured Y scores. For our example study, $R^2 = .29$, which indicates that job knowledge and socialization jointly account for 29% of the variance in Worker Satisfactoriness ratings.

Specialized Applications of MRC

The shift from a simultaneous approach to multiple regression (in which the dependent variable is regressed onto all predictors simultaneously) to a hierarchical approach (in which sets of predictors are entered sequentially in an order predetermined by the investigator) greatly enhances the flexibility of MRC analyses to address a variety of research hypotheses of interest to researchers in the social sciences. We have already noted one important application of HRA, in which the first predictor set serves as a statistical control of potential confounding variables, and the predictors of theoretical import are entered as a second block. The ΔR^2 for Block 2 represents the unique (and hence unconfounded) association of this second predictor set with the criterion variable. In the next sections, we consider applications of MRC to research questions involving categorical rather than continuous measures, to mediator hypotheses, to analysis of change, and to moderator hypotheses (i.e., statistical interactions).

Categorical Variables and MRC

Categorical Variables as Predictors

Coding Dichotomous Variables. Regression and correlation methods were developed to quantify relations among continuous variables. However, categorical (nominal) variables may also be analyzed using MRC. This process is straightforward for *dichotomous* variables (i.e., nominal variables with exactly two categories). For example, Nielsen (2003) included gender and marital status among her control variables. Each of these was a dichotomous variable, which could be included in the analysis by assigning a numerical code to each of the categories (e.g., 0 = male; 1 = female). Although any two different numerical values will work, the zero–one coding (called *dummy coding*) is a good approach, as it makes the regression coefficient for this coded variable easy to interpret. Specifically, if gender (coded 0 or 1, as above) is the only predictor (X_1), then the intercept (B_0) represents the predicted Y score for males (best estimate of the criterion when $X_1 = 0$). By extension, B_1 represents the difference between means for females and males (i.e., $B_1 = M_f - M_m$). This follows from the definition of the regression coefficient: the predicted change in Y for a one-unit change in X_1 (i.e., the change in means from the group coded as 0 to the group coded as 1).

Interpretation of the (partial) regression coefficient for gender when multiple predictors are included in the regression analysis follows this same principle, except that in this case B_0 and B_1 will be functions of *adjusted means* (controlling for the other predictor variables in the equation; see Cohen et al., 2003, pp. 342–350). Because regression coefficients for dichotomous variables are interpretable only if the numerical codes for the categories are known, it is crucial that investigators state how these variables were coded, either in the methods section or in the results section (and also in the table note, when regression results are tabulated).

Coding Polychotomous Variables. When a nominal variable has more than two categories (groups), the information from that variable cannot be completely represented by a single code variable. In general, if a nominal variable encompasses g groups, a set of $(g - 1)$ code variables will be needed to represent membership in these groups as a predictor in MRC. These code variables can then be entered as a set in HRA to assess the association between the categorical predictor variable and the continuous dependent variable. More details on creating dummy coded variables, on other coding schemes, and on interpreting regression output involving sets of variables coded to represent nominal scales can be found in most graduate level textbooks on MRC (e.g., Cohen et al., 2003, chap. 8; Pedhazur, 1982, chap. 9). Using these coding strategies, and entering coded variables as sets in HRA, allows researchers to combine categorical and continuous predictors within the MRC framework.

Categorical Variable as Criterion

Rehabilitation researchers are often interested in outcome variables that are dichotomous rather than continuous. For example, Taylor et al. (2003) followed children for 4 years following a traumatic brain injury (TBI) to identify predictors of long-term education outcomes. The dependent variable for this study was placement in special education (vs. no special education). When the dependent variable is naturally categorical, as here, traditional MRC techniques cannot be used. However, a related analytic technique called *logistic regression* is designed for categorical and continuous predictors of categorical criterion variables. Logistic regression can be conducted with either simultaneous or sequential (hierarchical) entry of predictors, and it provides estimates of both the unique contribution of individual predictors and the joint contribution of sets of predictors to the prediction of outcome status (Cohen et al., 2003, chap. 13). Thus, categorical dependent variables, like categorical predictor variables, can be accommodated within the broad family of MRC techniques.

Tests of Mediator and Moderator Hypotheses in MRC

When previous research has demonstrated an association between a predictor variable (designated as the *independent variable*, or IV, to reflect its presumed causal priority) and a dependent variable (DV), investigators may wish to examine proposed mediators or moderators of this association. Although the terms *mediator* and *moderator* are sometimes used interchangeably, they have distinct meanings in the context of hypothesis formulation and data analysis (Baron & Kenny, 1986). A *mediator* is an intervening variable that is caused by the IV, and that in turn causes the DV, so that at least part of the causal effect of the IV on the DV is explained by its indirect effect via the mediator. A *moderator* is a third variable that affects the strength or the direction of the association between the IV and the DV.

Baron and Kenny (1986) provided an informative discussion of these terms, with details about how to test each of these types of hypotheses using MRC. Another very helpful resource for testing moderator hypotheses using MRC is Aiken and West (1991, especially chap. 2). More recently, Frazier, Tix, and Barron (2004) published a useful, step-by-step guide to testing mediator and moderator hypotheses and reporting findings. The astute reader will note that all of these sources recommend MRC as the optimal analysis for testing mediator and moderator hypotheses involving naturally continuous variables. In particular, researchers should avoid the common practice of dichotomizing continuous predictor variables (using a "median split," for example) so that ANOVA may be used to test moderator hypotheses (statistical interactions). As demonstrated by MacCallum, Zhang, Preacher, and Rucker (2002), dichotomization of continuous variables for any type of analysis (not just moderator analyses) compromises statistical power and can yield misleading results. Any of the references just cited can assist readers in analyzing and interpreting interactions between continuous predictors using MRC.

Analysis of Change in MRC

When data are collected at two or more time points, investigators may test hypotheses involving prediction of change in the dependent variable over time. Such analyses always (either explicitly or implicitly) involve computation of *change scores* reflecting the change in status for each individual on the dependent variable between the two time points of interest. Appropriate analysis of hypotheses involving change is the subject of a rich literature (see the classic paper by Cronbach & Furby, 1970, for a thorough introduction); here we mention only one small controversy, which relates to the use of MRC.

The most natural, and perhaps the most common method of testing hypotheses of change between two time points uses *difference scores*, which are derived by subtracting each person's score at Time 1 from his or her score at the later Time 2, as an index of change. Difference scores are an intuitive means to quantify change over time, and are used implicitly in such common analyses as the *t* test for dependent samples and the group × time (repeated measures) ANOVA for assessing treatment effects (Huck & MacLean, 1975). Difference scores have been criticized as an index of change, however, because they are necessarily (negatively) correlated with Time 1 scores (Cohen et al., 2003, pp. 59–60). When investigators wish to create change scores that are statistically independent of initial status, they should use *residualized change scores* (also called *partialed change scores;* Cohen et al., 2003, pp. 570–571), regressing Time 2 scores (DV) onto Time 1 scores (predictor) and saving the residuals as an index of change. By definition, these residuals are independent of (i.e., uncorrelated with) Time 1 status. Residualized change scores are implicitly analyzed when ANCOVA is used to test for group differences with initial scores on the DV as a covariate (Kenny, 1979, chap. 11). Likewise, in any regression analysis with Time 2 scores as the dependent variable, where Time 1 scores on this DV are entered as one of the predictor variables, the partial regression coefficients for the remaining predictor variables reflect their association with *change* on the DV (i.e., with participants' residualized change scores on this variable, described above).

Miscellaneous Topics

Power Analysis in MRC

An important question for the design of research studies using MRC concerns the sample size necessary to have adequate statistical power. Methodologists commonly recommend that researchers aim for power of at least .80, which corresponds to a Type II error rate of 20%. Procedures for estimating statistical power differ for two types of hypothesis testing common to studies using MRC: (a) tests that determine whether the multiple correlation (R) between a set of predictors and the criterion variable is different from zero and (b) tests to determine whether the association of a single predictor variable (or a set of variables, when HRA is used) with the DV is nonzero, when other predictors are also in the regression equation (and therefore are statistically controlled)—that is, tests of the significance of B, β, or sr^2. A challenge for power analysis for each of these analyses, but especially for the second type of research question, is the determination of the expected effect size. Because of space limitations (and because excellent resources on power analysis are readily available), we provide only a brief introduction to power analysis in MRC.

Although several rules of thumb have been proposed for determining sample size in MRC (e.g., $N = 10k$, where k is the number of predictor variables), none of these is above reproach, and most bear little or no relation to the actual (complex) relation between power and sample size (Maxwell, 2000). An important reason for this failure is that these rules of thumb completely ignore the most important determinant of power other than sample size, which is the effect size. The best approach is to conduct a power analysis using a predicted effect size to determine the sample size (N) necessary to attain a specified power level (Cohen, 1988). See Cohen et al. (2003) for concise instructions on conducting a power analysis for testing the hypothesis that $R^2 = 0$ (p. 92) or for instructions on testing that incremental variance explained by a new set of variables (denoted as Set B; variance explained uniquely by this set of predictors is denoted sR^2_B) is zero (pp. 176–177). The latter test works for sets containing a single variable (i.e., $k_B = 1$) or multiple variables (i.e., $k_B > 1$). When $k_B = 1$, the significance test for sR^2_B is identical to that for B or β (even if all predictors are entered simultaneously rather than using HRA).

In summary, rules of thumb for determining sample size based only on the number of predictor variables are misleading, and they bear little relation to actual statistical power. Power calculations should be based on the expected effect size (either R^2 or sR^2_B).

Factors Affecting Effect Size Estimates

We noted that a convenience of MRC is the plethora of effect size measures available from regression analyses. This boon is mitigated somewhat by the caveat that, for a variety of reasons, observed correlation and regression coefficients typically distort the true magnitude of association between constructs in the population under study. Fortunately, the direction of bias in the effect size estimates is predictable and can (and should) be taken into account in interpreting study findings.

Measurement Error and Attenuation

A factor that universally affects research using measured (rather than experimentally manipulated) variables is measurement error. Scores on psychological assessments are always less than perfectly reliable, so that variance in scores reflects a composite of error variance and true score variance. Because the error components of two sets of scores are, by definition, uncorrelated, correlations between measured variables are always attenuated (i.e., reduced) relative to the correlations between their respective true scores. As score reliability decreases, the degree of attenuation of the bivariate correlation increases (Schmidt & Hunter, 1999). In other words, the poorer the reliability of its measures, the greater the degree to which a study's observed correlation is expected to underestimate the true (population) correlation between the constructs of interest.

Distribution, Range Restriction, and Dichotomization

Other factors can attenuate effect sizes in correlational research (see Cohen et al., 2003, pp. 51–62). When scores on one variable are significantly skewed, correlations with other measures will be attenuated. When the range of scores in the sample is restricted relative to the range in the population, correlations with scores on another variable will be attenuated. When researchers convert continuous measures to dichotomous measures (usually so that they can use ANOVA rather than MRC), they discard valid variance and further attenuate correlations between this variable and others in their study. In bivariate analyses, all of these factors act to attenuate the observed effect size, producing observed effect sizes (r or B) that underestimate the corresponding population effect size.

Summary

Researchers studying measured (rather than experimentally manipulated) variables should be aware that observed effect sizes are generally attenuated relative to the corresponding population effect size and that the degree of attenuation can be minimized by (a) using reliable measures of the constructs of interest; (b) transforming highly skewed variables prior to analysis; (c) sampling broadly to reduce the risk of range restriction; and (d) using MRC methods to analyze continuous variables (rather than dichotomizing these scores to analyze them using

ANOVA). When multiple predictor variables are analyzed, the effects of measurement error, deviations from normality, and range restriction on partial coefficients are more complex. For example, low reliability in a variable that is statistically controlled in a given analysis can have the effect of inflating, rather than attenuating, the partial regression coefficient between another predictor and the criterion (Cohen et al., 2003, pp. 122–124). Also, the statistical power of moderator analyses is particularly strongly reduced by unreliability of measurement, because the reliability of the product term (which carries the interaction variance) is primarily a function of the product of the reliabilities of the IV and the moderator variable (Aiken & West, 1991, pp. 144–145), especially when the IV and moderator are only weakly correlated with one another.

Conclusion

MRC techniques give researchers the flexibility to address a wide variety of research questions of interest to rehabilitation professionals. Good data analysis begins with careful conceptualization (selecting constructs of interest and creating theory-derived hypotheses about the relations among them) and thoughtful choice of measures. Power analysis, relying on estimates from past research or estimates about the likely magnitude of hypothesis-relevant effect sizes, is an essential component of good research design. It is critical that the analysis chosen conform to the hypothesis to be tested, and that observed effect sizes, as well as significance tests, be presented and interpreted as substantive findings concerning the magnitude of hypothesized associations.

NOTES

1. Although we follow the linguistic conventions that treat the predictor variable as the putative cause of the criterion variable, it is important to remember that regression is a correlational analysis and does not by itself provide empirical evidence of a cause-and-effect relation between two variables.
2. The z score is a deviation score that is expressed in *SD* units:

$$z_i = \frac{X_i - M_X}{sd_X}$$

where z_i is the z score for person i, X_i is the raw score for that person, M_X is the mean of all the X scores in the sample, and SD_X is their standard deviation. If $z_i = 1.0$, this means person i scored 1 *SD* above the sample mean.
3. The exception to this general rule, in which the partial coefficient $B_{Y1\cdot2}$ is larger in absolute value than the corresponding bivariate coefficient B_{Y1} (or, equivalently, the standardized partial coefficient $\beta_{Y1\cdot2}$ is larger than the bivariate correlation r_{Y1}), is known as *suppression*. Although bona fide cases of suppression appear to be fairly rare in the social science literature, they do exist, and this pattern of relations can have theoretical significance. For a detailed discussion of suppression, with substantive examples, see Tzelgov and Henik (1991).
4. By symmetry, everything said in the last three paragraphs about $B_{Y1\cdot2}$ also applies to $B_{Y2\cdot1}$, if X_1 and X_2 are interchanged.

REFERENCES

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

Azen, R., & Budescu, D. V. (2003). The dominance analysis approach for comparing predictors in multiple regression. *Psychological Methods, 8*, 129–148.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173–1182.

Cohen, J. (1968). Multiple regression as a general data-analytic strategy. *Psychological Bulletin, 70*, 426–443.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/ correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Cronbach, L. J., & Furby, L. (1970). How should we measure "change"—or should we? *Psychological Bulletin, 74*, 68–80.

Frazier, P. A., Tix, A. P., & Barron, K. E. (2004). Testing moderator and mediator effects in counseling psychology research. *Journal of Counseling Psychology, 51*, 115–134.

Huck, S. W., & MacLean, R. A. (1975). Using a repeated measures ANOVA to analyze the data from a pretest–posttest design: A potentially confusing task. *Psychological Bulletin, 82*, 511–518.

Kenny, D. A. (1979). *Correlation and causality*. New York: Wiley.

Kline, R. B. (2004). *Beyond significance testing: Reforming data analysis in behavioral research*. Washington, DC: American Psychological Association.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*, 19–40.

Maxwell, S. E. (2000). Sample size and multiple regression analysis. *Psychological Methods, 5*, 434–458.

Millington, M. J., Leierer, S. J., & Abadie, M. (2000). Validity and the Employment Expectation Questionnaire–beta version: Do attitudes mediate outcomes? *Rehabilitation Counseling Bulletin, 44*, 39–47.

Nielsen, M. S. (2003). Prevalence of posttraumatic stress disorder in persons with spinal cord injuries: The mediating effect of social support. *Rehabilitation Psychology, 48*, 289–295.

Pedhazur, E. J. (1982). *Multiple regression in behavioral research: Explanation and prediction* (2nd ed.). Fort Worth, TX: Harcourt Brace.

Schmidt, F. L., & Hunter, J. E. (1999). Theory testing and measurement error. *Intelligence, 27*, 183–198.

Taylor, H. G., Wade, S. L., Stancin, T., Yeates, K. O., Drotar, D., & Montpetite, M. (2003). Long-term educational interventions after traumatic brain injury in children. *Rehabilitation Psychology, 48*, 227–236.

Thompson, B. (1995). Stepwise regression and stepwise discriminant analysis need not apply here: A guidelines editorial. *Educational and Psychological Measurement*, *55*, 525–534.

Tzelgov, J., & Henik, A. (1991). Suppression situations in psychological research: Definitions, implications, and applications. *Psychological Bulletin*, *109*, 524–536.

Wampold, B. E., & Freund, R. D. (1987). Use of multiple regression in counseling psychology research: A flexible data-analytic strategy. *Journal of Counseling Psychology*, *34*, 372–382.

Wilkinson, L., & Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *The American Psychologist*, *54*, 594–604.