Regression coefficients cannot be interpreted as causal if the relationship can be attributed to an alternate mechanism. One may control for the alternate cause through an experiment (e.g., with random assignment to treatment and control) or by measuring a corresponding confounding variable and including it in the model. Unfortunately, there are some circumstances under which it is not possible to measure or control for the potentially confounding variable. Under these circumstances, it is helpful to assess the robustness of a statistical inference to the inclusion of a potentially confounding variable. In this article, an index is derived for quantifying the impact of a confounding variable on the inference of a regression coefficient. The index is developed for the bivariate case and then generalized to the multivariate case, and the distribution of the index is discussed. The index also is compared with existing indexes and procedures. An example is presented for the relationship between socioeconomic background and educational attainment, and a reference distribution for the index is obtained. The potential for the index to inform causal inferences is discussed, as are extensions.

Impact of a Confounding Variable on a Regression Coefficient

KENNETH A. FRANK
Michigan State University

INTRODUCTION: "BUT HAVE YOU CONTROLLED FOR . . . ?"

As is commonly noted, one must be cautious in making causal inferences from statistical analyses (e.g., Abbott 1998; Cook and Campbell 1979; Rubin 1974; Sobel 1995, 1996, 1998). Although there is extensive debate with regard to the conditions necessary and sufficient to infer causality, it is commonly accepted that the causality attributed to factor A is weakened if there is an alternative factor that causes factor A as well as the outcome (Blalock 1971; Einhorn and Hogarth 1986; Granger 1969; Holland 1986; Meehl 1978; Pratt and

AUTHOR'S NOTE: *Betsy Becker, Charles Bidwell, Kyle Fahrbach, Alka Indurkhya, Kyung-Seok Min, Aaron Pallas, Stephen Raudenbush, Mark Reckase, Wei Pan, Feng Sun, Alex von Eye, and the anonymous reviewers offered important suggestions. I am indebted to Feng Sun for helping me prepare figures, check formulas, and identify reference material. Kate Muth helped with the construct figures and the references.*

SOCIOLOGICAL METHODS & RESEARCH, Vol. 29 No. 2, November 2000 147-194

Schlaifer 1988; Reichenbach 1956; Rubin 1974; Salmon 1984; Sobel 1995, 1996; Zellner 1984). That is, we are more cautious about asserting a causal relationship between factor A and the outcome if there is another factor potentially confounded with factor A.

For example, in the famous debate on effectiveness of Catholic high schools, the positive effect of Catholic schools on achievement was questioned (see Bryk, Lee, and Holland 1993; Jencks 1985). Students entering Catholic schools were hypothesized to have higher achievement prior to entering high school than those attending public schools. Thus, prior achievement is related to both sector of school attended and later achievement. Because prior achievement is causally prior to the sector of the high school a student attends, prior achievement is a classic example of a confounding variable. No definitive interpretation with regard to the effect of Catholic schools on achievement can be made without first accounting for prior achievement as a confounding variable.

In the case of an evaluation of a treatment effect, one might control for prior differences by randomly assigning subjects to treatment groups. In the language of the counterfactual argument (Giere 1981; Lewis 1973; Mackie 1974; Rubin 1974), the difference in means between treatment and control groups then represents the effect a given subject *would have* experienced had he or she been exposed to one type of treatment instead of the other (see Kitcher 1989 for the potential of counterfactuals to inordinately dominate a theory of causality). If the causal effect is defined as the difference between two potential outcomes (only one of which can ever be observed), random assignment assures that the missing counterfactual outcome is missing completely at random. Therefore, under random assignment, and barring any other threats to internal validity (see Cook and Campbell 1979), the average difference between treatment and control groups within any subset of persons is an unbiased estimate of the average treatment effect for that subset of persons.

Random assignment to treatment and control is often impractical in the social sciences given logistical concerns, ethics, political contexts, and sometimes the very nature of the research question (e.g., Cook and Campbell 1979; Rubin 1974). For example, to randomly assign students to Catholic or public schools would be unethical. Therefore, we often turn to observational studies and use of statistical control of a confound (Cochran 1965; McKinlay 1975). In this approach, the potentially confounding variable is measured and included as a covariate in a quantitative model. Then, for each level of the covariate, and given that the assumptions of ANCOVA have been satisfied, individuals can be considered *conditionally* randomly assigned to treatment and control (Rosenbaum and Rubin 1983; Rubin 1974). (For a recent discussion of conditional random assignment and its impact on expected mean differences between treatment and control, see Sobel [1998].) For example, if one included a measure of prior achievement in a model assessing the effect of school sector, one could then consider students to be randomly assigned to Catholic or public school conditional on their prior achievement. One could also pursue such control through various matching schemes (Cochran 1953; McKinlay 1975; Rubin 1974).

Unfortunately, it is not always possible to measure all confounds and include each as a covariate in a quantitative model. This may be especially true for analyses of secondary data such as were used to assess the Catholic school effect. Furthermore, even if one confounding variable is measured and included as a control, no coefficient can be interpreted as causal until the list of possible confounds has been exhausted (Pratt and Schlaifer 1988; Sobel 1998). Paradoxically, the more remote an alternative explanation, the less likely is the researcher to have measured the relevant factor, the more prohibitive becomes the critique. The simple question, "Yes, but have you controlled for *xxx*?" puts social scientists forever in a quandary as to inferring causality.

Some have suggested social scientists should temper their claims of causality (Abbott 1998; Pratt and Schlaifer 1988; Sobel 1996, 1998). Alternatively, one can assess the sensitivity of results to inclusion of a confounding variable (Mauro 1990; Rosenbaum 1986). If a coefficient is determined to be insensitive to the impact of confounding variables, then it is more reasonable to interpret the coefficient as indicative of an effect.

But sensitivity analyses are often computationally intensive and tabular in their result, requiring extensive and not always definitive interpretations. Furthermore, they are difficult to extend to models including multiple covariates such as are typical in the social sciences. Not surprisingly, although the value of sensitivity analyses is often noted, sensitivity analyses are applied infrequently (e.g., Mauro's [1990] article has been cited by only eight other authors and Rosenbaum's [1986] article by only five according to the Social Science Citation Index [see also Cordray 1986]).

In this article, I extend existing techniques for sensitivity analyses by indexing the impact of a potentially confounding variable on the statistical inference with regard to a regression coefficient. The index is a function of the hypothetical correlations between the confound and outcome, and between the confound and independent variable of interest. The expression for the index allows one to calculate a single valued threshold at which the impact of the confound would be great enough to alter an inference with regard to a regression coefficient.

In the next section, I develop the index for bivariate regression and obtain the threshold at which the impact of the confounding variable would alter the inference of a regression coefficient. I then discuss the range and distribution of the index. In the following section, I relate the index to existing procedures and techniques. I then develop the index for the multivariate case. I apply the index to an example of the relationship between socioeconomic status and educational attainment and describe a reference distribution for the index. In the discussion, I comment on the use of the index and reference distribution relative to larger debates on statistical and causal inference, and explore extensions.

IMPACT OF A CONFOUNDING VARIABLE

A confounding variable is characterized as one that correlates with both a treatment (or predictor of interest) and outcome (Anderson et al. 1980, chap. 5). It is also considered to occur causally prior to the treatment (see Cook and Campbell 1979). In other contexts, the confounding variable might be referred to as a "disturber" (Steyer and Schmitt 1994), a "covariate" (Holland 1986, 1988), or, drawing from the language of R. A. Fischer, a "concomitant" variable (Pratt and Schlaifer 1988).

Consider the following two standard linear models for the dependent variable y_i of subject i:[1]

$$y_i = \beta_0 + \beta_1 x_i + e_i \tag{1a}$$

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 cv_i + e_i. \qquad (1b)$$

In each model, x_i is the value of the predictor of interest for subject *i*. The variable *x* might be a treatment factor, or a factor that the researcher focuses on in interpretation (Pratt and Schlaifer 1988). In the case of school research, *x* might indicate whether the student attends a Catholic school.

The scenario begins when the estimate of β_1, referred to as $\hat{\beta}_1$, is statistically significant in model (1a). That is, the *t* ratio defined by $[\hat{\beta}_1/se(\hat{\beta}_1)]$ is larger in absolute value than the critical value of the *t* distribution for a given level of significance, α, and degrees of freedom, $n - 2$ (where *n* represents the sample size). Therefore, we would reject the null hypothesis that $\beta_1 = 0$. (For the remainder of this article, this framework of null hypothesis statistical testing applies unless otherwise stated.) An inference with regard to β_1 is based on the assumption that the e_i are independent and identically normally distributed with common variance. This assumption is necessary for ordinary least squares estimates to be unbiased and asymptotically efficient, and for the ratio $\hat{\beta}_1/se(\hat{\beta}_1)$ to have a *t* distribution.

An implication of "identically distributed" is that e_i are independent of x_i, thus ensuring that the e_i have the same mean across all values of x_i (e.g., Hanushek 1977). This latter assumption is necessary for the ordinary least squares estimate of β_1 to be unbiased. The assumption is violated if there is a confounding variable, *cv*, that is correlated with both *e* and *x* and is considered causally prior to *x*.[2] For example, in the Catholic schools research, sector of school attended is not independent of the errors if the confounding variable, achievement prior to entering high school, is related to both sector of school attended and final achievement.

The concern then is that $\hat{\beta}_1$ is statistically significant in (1a) but not so in (1b), which includes a confounding variable, *cv*. In this case, $\hat{\beta}_1$ in (1a) cannot be interpreted as indicative of an effect of *x* on *y*. My focus is on the change in inference with regard to β_1 because social scientists use statistical significance as a basis for causal inference and for policy recommendations. Therefore, the threshold of statistical significance has great import in practice, even if cutoff values such as .05 are arbitrary (in the discussion, I will comment more on concerns with regard to the use of significance testing and causal inference).

To assess the impact of a confounding variable on the inference of a regression coefficient, we can ask the following question: Given $\hat{\beta}_1$ is statistically significant in (1a), how large must the correlations be between *cv* and *y* and between *cv* and *x* such that $\hat{\beta}_1$ is not statistically significant in (1b)? I will answer this question by expressing the components of a *t* test for a partial regression coefficient in terms of zero-order correlations. Throughout this presentation, I use *r* to represent correlations that are estimated in a given sample as well as correlations that would be observed were a measure of a confounding variable included in the data. In particular, define quantities used for estimation and inference in model (1a): $r_{y \cdot x}$ is the sample correlation between the outcome and the predictor of interest (e.g., *x* is a treatment dummy variable), R^2_y is the proportion variance explained in a regression with just the predictor of interest, *n* is the sample size, *q* is the number of parameters estimated in a model (not including the intercept, β_0), *sd*(*y*) is the standard deviation of *y*, and *sd*(*x*) is the standard deviation of *x*. Define additional quantities needed to estimate model (1b): $r_{y \cdot cv}$ is the correlation between the outcome and the confounding variable, and $r_{x \cdot cv}$ is the correlation between the predictor of interest and the confounding variable. Then $\hat{\beta}_1$, the standard error of $\hat{\beta}_1$, and their ratio defining *t* (on $n - q - 1$ degrees of freedom) for model (1b) are[3]

$$\hat{\beta}_1 = \frac{sd(y)}{sd(x)} \times \frac{r_{y \cdot x} - r_{y \cdot cv} r_{x \cdot cv}}{1 - r^2_{x \cdot cv}},$$

$$se(\hat{\beta}_1) = \frac{sd(y)}{sd(x)} \times \sqrt{\frac{1 - R^2_y}{n - q - 1} \times \frac{1}{1 - r^2_{x \cdot cv}}} = \tag{2}$$

$$\frac{sd(y)}{sd(x)} \times \sqrt{\frac{1 - \dfrac{(r^2_{y \cdot x} + r^2_{y \cdot cv} - 2 r_{y \cdot x} r_{y \cdot cv} r_{x \cdot cv})}{1 - r^2_{x \cdot cv}}}{n - q - 1} \times \frac{1}{1 - r^2_{x \cdot cv}}} \quad \text{and}$$

$$\Rightarrow t(\hat{\beta}_1) = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)} = \frac{r_{y \cdot x} - k}{\sqrt{\dfrac{1 - r^2_{x \cdot cv} - (r^2_{y \cdot x} + r^2_{y \cdot cv} - 2 r_{y \cdot x} k}{n - q - 1}}}.$$

Although alternative expressions can be developed to calculate the t ratio of a regression coefficient, equations such as (2), and similar equations used by Mauro (1990), are unique in expressing the t statistic in terms of components associated with a potentially confounding variable.

The product $r_{y\cdot cv} \times r_{x\cdot cv}$ appears in the numerator and denominator of the t ratio in (2) and includes the two relations defining confounding—the relationship between the outcome and the confounding variable ($r_{y\cdot cv}$) and the relationship between the predictor of interest and the confounding variable ($r_{x\cdot cv}$). Therefore, because the goal is to develop a concise expression for the impact of a confounding variable, I characterize the overall impact of the confound in terms of the product, $k = r_{y\cdot cv} \times r_{x\cdot cv}$. Replacing $r_{y\cdot cv} \times r_{x\cdot cv}$ in (2) with k,

$$t(\hat{\beta}_1) = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)} = \frac{r_{y\cdot x} - k}{\sqrt{\dfrac{1 - r_{x\cdot cv}^2 - (r_{y\cdot x}^2 + r_{y\cdot cv}^2 - 2r_{y\cdot x}k)}{n - q - 1}}} . \quad (3)$$

Next, the argument that inclusion of the confounding variable in the model will alter the interpretation of $\hat{\beta}_1$ is validated when $t(\hat{\beta}_1)$ for model (1b) becomes *less than* a critical value. Correspondingly, we seek to minimize $t(\hat{\beta}_1)$ with respect to $r_{y\cdot cv}$ and $r_{x\cdot cv}$ (thus affording the confounding variable the greatest impact on $t[\hat{\beta}_1]$).

Because n, q, and $r_{y\cdot x}$ can be obtained from the sample data, for a given value of k the right-hand side of (3) is minimized when the expression in the denominator is maximized, which occurs when $1 - r_{x\cdot cv}^2 - (r_{y\cdot x}^2 + r_{y\cdot cv}^2 - 2r_{y\cdot x}k)$ is maximized. Using a LaGrange multiplier based on the constraint $r_{y\cdot cv} \times r_{x\cdot cv} = k$, the maximum occurs when

$$\nabla([1 - r_{x\cdot cv}^2 - (r_{y\cdot x}^2 + r_{y\cdot cv}^2 - 2r_{y\cdot x}k)], r_{x\cdot cv}, r_{y\cdot cv}) = \lambda\nabla([r_{x\cdot cv}r_{y\cdot cv} - k], r_{x\cdot cv}, r_{y\cdot cv})$$

$$\Rightarrow -2r_{x\cdot cv}i - 2r_{y\cdot cv}j = \lambda\left(r_{y\cdot cv}i + r_{x\cdot cv}j\right) \quad (4)$$

$$\Rightarrow \frac{-2r_{x\cdot cv}}{r_{y\cdot cv}} = \frac{-2r_{y\cdot cv}}{r_{x\cdot cv}}$$

$$\Rightarrow r^2_{x \cdot cv} = r^2_{y \cdot cv} \ .$$

Thus, the constrained relative minimum of $t(\hat{\beta}_1)$ is at $r^2_{x \cdot cv} = r^2_{y \cdot cv} = k$. Replacing $r^2_{x \cdot cv}$ and $r^2_{y \cdot cv}$ with k in (3), we have

$$t(\hat{\beta}_1) = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)} = \frac{r_{y \cdot x} - k}{\sqrt{\dfrac{1 - k - (r^2_{y \cdot x} + k - 2r_{y \cdot x}k)}{n - q - 1}}} = \frac{r_{y \cdot x} - k}{\sqrt{\dfrac{(1 + r_{y \cdot x} - 2k)(1 - r_{y \cdot x})}{n - q - 1}}} \ . \qquad (5)$$

Thus, $t(\hat{\beta}_1)$ can be expressed as a function of $r_{y \cdot x}$, k, and the degrees of freedom $(n - q - 1)$.

The definition of k and the expression in (5) can be used to characterize the distribution of k. First, when $k = 0$ the expression in (5) is equivalent to the standard test for a regression coefficient, save for a one-unit change in degrees of freedom (but this is merely the same issue as including a nonpredictive variable in a model). That is, when $k = 0$ the test of $\hat{\beta}_1$ in model (1b) is comparable to the test of $\hat{\beta}_1$ in (1a) that does not include the confounding variable. Second, k is based on the product of two correlation coefficients. Therefore, $-1 \leq k \leq 1$. Furthermore, when $2k \geq 1 + r_{y \cdot x}$ the expression in (5) becomes undefined or complex, indicating that $t(\hat{\beta}_1)$ cannot be interpreted. This restriction corresponds to the mathematically defined lower limit of $r_{y \cdot x}$ as given by $r_{y \cdot x} > r_{y \cdot cv} \times r_{x \cdot cv} - [(1 - r^2_{x \cdot cv})(1 - r^2_{y \cdot cv})]^{1/2}$ (see Cohen and Cohen 1983:280). When the limit is exceeded, the matrix of correlations containing $r_{y \cdot x}$, $r_{y \cdot cv}$, and $r_{x \cdot cv}$ cannot represent a set of product moment correlations from a single sample and is ill conditioned. Therefore, the restriction in (5) is equivalent to the restriction associated with estimation of the general linear model.

For the moment, assume $k > 0$ (the case for $k < 0$ will be addressed through symmetry). To find the value of k that would alter the inference with regard to $\hat{\beta}_1$, solve (5) for k. Define $t = t(\hat{\beta}_1)$. Then,[4]

$$k = \frac{b \pm \sqrt{b^2 - ac}}{a},$$

where

$$a = -(n - q - 1),$$

$$b = t^2(1 - r_{y \cdot x}) - (n - q - 1)r_{y \cdot x},$$

and

$$c = t^2(1 - r_{y \cdot x}^2) - (n - q - 1)r_{y \cdot x}^2.$$

$$\Rightarrow k = \frac{t^2(1 - r_{y \cdot x}) - (n - q - 1)r_{y \cdot x} \pm \sqrt{\begin{array}{l}(t^2(1 - r_{y \cdot x}) - (n - q - 1)r_{y \cdot x})^2 \\ +(n - q - 1)(t^2(1 - r_{y \cdot x}^2) - (n - q - 1)r_{y \cdot x}^2)\end{array}}}{-(n - q - 1)}. \quad (6)$$

Define $d = t^2 + (n - q - 1)$, then

$$k = ITCV = \frac{t^2 + t\sqrt{d}}{-(n - q - 1)} + \left[\frac{-d - t\sqrt{d}}{-(n - q - 1)}\right] r_{y \cdot x}, \text{ or}$$

$$k = ITSV = \frac{t^2 - t\sqrt{d}}{-(n - q - 1)} + \left[\frac{-d + t\sqrt{d}}{-(n - q - 1)}\right] r_{y \cdot x}.$$

In the last equivalences of (6), I refer to the expressions for k as impact thresholds (ITs).[5] Assuming that $t = t_{\text{critical}}$ (with degrees of freedom $n - q - 1$) and other assumptions of inference have been satisfied, the impact threshold for a confounding variable (ITCV) indicates the impact necessary to make a $\hat{\beta}_1$ (and $r_{y \cdot x}$) that is positive and statistically significant become positive and just statistically significant.

The impact threshold for a suppressor variable (ITSV) would make an $r_{y \cdot x}$ (and $\hat{\beta}_1$) that is positive become just statistically significantly and negative. This modifies typical definitions of suppression that are stated in terms of a change in magnitude or sign of $\hat{\beta}_1$. But is the change enough to alter the inference with regard to β_1? Typically, definitions of suppression do not address this issue. Thus, suppression is defined here as the complement to confounding, making $\hat{\beta}_1$ statistically significant when it is not so in the original model.

The relationship between the IT (ITCV or ITSV) and $r_{y \cdot x}$ for IT > 0 is shown in the upper half of Figure 1 for $n = 29$, $t = 2.05$, and $\alpha = .05$. The lines for the ITCV and the ITSV each pass through the point (1, 1) as each equation in (6) approaches $-n/(-n)$ as $r_{y \cdot x}$ approaches unity. Figure 1 also includes lines for the ITCV and the ITSV when the IT <

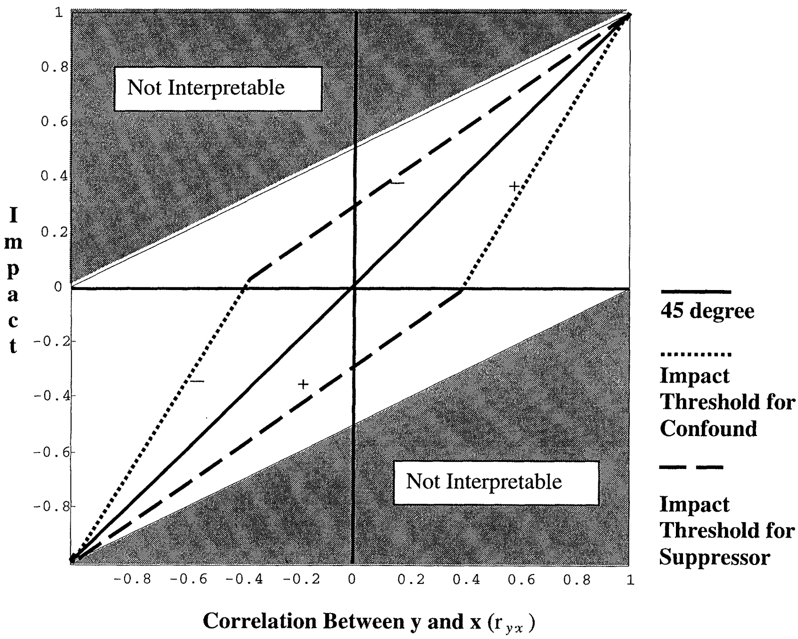

Figure 1: Impact Threshold and Correlation Between *y* and *x* ($n = 29$, $\alpha = .05$)
NOTE: + Impact makes $r_{y \cdot x}$ positive and statistically significant; – Impact makes $r_{y \cdot x}$ negative and statistically significant.

0, which can be obtained through arguments symmetric to those applied in the development of (6) and by interchanging the intercepts and slopes for the ITCV and the ITSV.

The graph in Figure 1 is symmetric because magnitudes of the impact thresholds for $r_{y \cdot x} < 0$ are the same as those for $r_{y \cdot x} > 0$. Thus, the roles of the ITCV and the ITSV can be stated in more general terms: when $r_{y \cdot x}$ is statistically significant, an impact of the ITSV will make $r_{y \cdot x}$ just statistically significant with opposite sign, and an impact of the ITCV will make $r_{y \cdot x}$ just statistically significant while retaining its sign. When $r_{y \cdot x}$ is not statistically significant, the ITCV is not defined, and the ITSVs represent the impacts necessary for $r_{y \cdot x}$ to become statistically significant in either direction. Graphically, inside the parallelogram, $\hat{\beta}_1$ would not be statistically significant were the confounding variable included in the model. Outside the boundary, $\hat{\beta}_1$ would be statistically significant, and on the border $\hat{\beta}_1$ would become just statistically significant. Values of impact in the shaded regions correspond to a mathematical inconsistency among $r_{y \cdot x}$, $r_{y \cdot cv}$, and $r_{x \cdot cv}$ and, therefore, are not interpretable.

Together, the functions ITCV > 0 and ITSV < 0 define the impact necessary to make $r_{y \cdot x}$ positive and just statistically significant. But note a given change in $r_{y \cdot x}$ is associated with larger changes in the ITCV than in the ITSV. That is, the impact of suppression is relatively stronger than the impact of confounding. For example, $r_{y \cdot x}$ is just statistically significant at .36, at which point ITCV = ITSV = 0. Moving .3 units to the left of .36, at $r_{y \cdot x}$ = .06, the ITSV is about –.2, a change of .2 units. On the other hand, .3 units to the right of .36, at $r_{y \cdot x}$ = .66, the ITCV is about .52, a change of more than .5 units.

The strength of the suppression impact relative to the confounding impact is consistent with the principle of analysis of covariance: Including covariates in a model tends to reduce residual variation and thus improve our certainty and inferences. Thus, in suppression, the algebraic effect of $r_{y \cdot cv} \times r_{x \cdot cv}$ in the numerator of the t statistic is complemented by the effect in the denominator, whereas in confounding, the two effects compete. In terms of the statistical mechanics of the linear model, it is easier to induce suppression than to establish confounding.

An alternate way to regard Figure 1 is to note the value of $r_{y \cdot x}$ defines a natural delimiter for the impact of a confounding variable. An impact equivalent to $r_{y \cdot x}$ can occur when cv is perfectly correlated with x ($r_{x \cdot cv}$ = 1, $r_{y \cdot cv} = r_{y \cdot x}$) or when cv is perfectly correlated with y ($r_{y \cdot cv}$ = 1, $r_{x \cdot cv} = r_{x \cdot y}$). The line for which the impact equals $r_{y \cdot x}$ corresponds to the 45-degree line in Figure 1. Above this line, the IT is associated with an $r_{y \cdot x}$ that is negative and just statistically significant, regardless of the initial sign of $r_{y \cdot x}$. Below the 45-degree line, the IT is associated with an $r_{y \cdot x}$ that is positive and just statistically significant.

Of course, the value of the IT depends in part on sample size. The range of the IT for positive values of $r_{y \cdot x}$ for $\alpha = .05$ is shown across a range of sample sizes in Figure 2. The sample sizes 784, 85, and 29 correspond to the sample sizes necessary for small (.1), moderate (.3), and large (.5) values of $r_{y \cdot x}$ to be statistically significant at $\alpha = .05$ with power = .80 for individual-level data (see Cohen and Cohen 1983:61, table G2).[6]

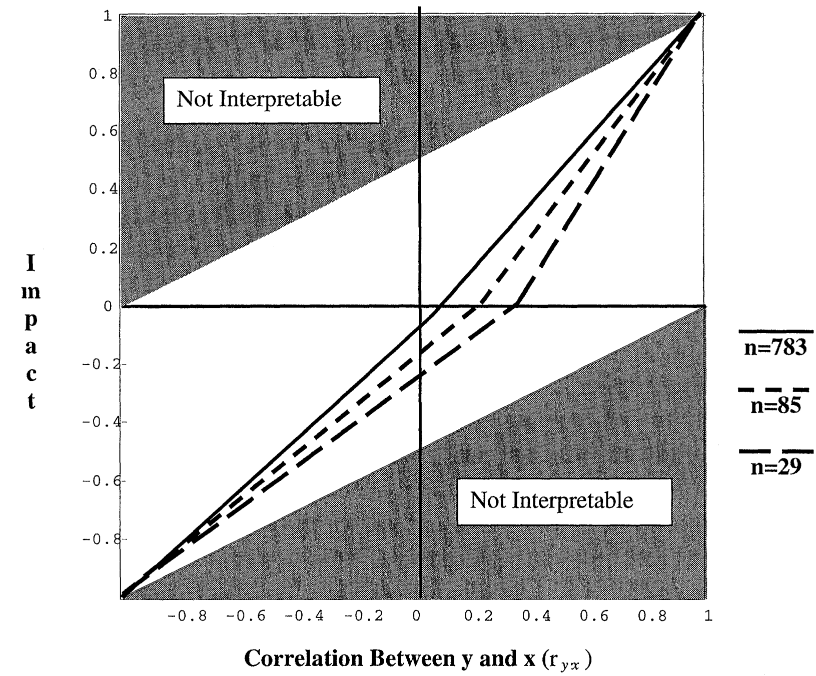

Figure 2: Impact Threshold and Correlation Between *y* and *x* for Different Sample Sizes

Each line crosses the reference line (where IT = 0) at the point at which $r_{y \cdot x}$ is just statistically significant. Furthermore, the intercepts (where $r_{y \cdot x} = 0$) indicate the threshold at which the impact of a suppressor is great enough to make an original correlation of zero become statistically significant. For example, for a sample of 784, an impact of –.08 is great enough to make $r_{y \cdot x} = 0$ statistically significant once the suppressor is included in the model.

All of the lines converge as $r_{y \cdot x} \rightarrow 1$ and the IT $\rightarrow 1$. When the correlation is strong, values of the IT must be near unity to alter the interpretation of $\hat{\beta}_1$, regardless of sample size. Values of the IT are larger for larger sample sizes, since it should be more difficult to alter an interpretation of $\hat{\beta}_1$ (and $r_{y \cdot x}$) for larger sample sizes. Furthermore, the slopes for the different sample sizes are not parallel. A given change in

$r_{y \cdot x}$ is associated with relatively more change in the ITCV for small sample sizes than for large sample sizes. In contrast, a given change in $r_{y \cdot x}$ is associated with relatively less change in the ITSV for small sample sizes than for large sample sizes.

The slopes for the ITCV and the ITSV converge for large sample sizes. That is, the ITCV and the ITSV both approach the line defined by IT $= r_{y \cdot x}$ as n increases. While this may seem a surprising result, it is contained within the expression for $\hat{\beta}_1$. When n is large, any value of $\hat{\beta}_1$ different from zero will be statistically significant. Therefore, $\hat{\beta}_1$ must approach zero to be not statistically significant. The numerator for $\hat{\beta}_1$ approaches zero when $r_{y \cdot x} = r_{y \cdot cv} \times r_{x \cdot cv}$. Thus, a simple rule of thumb when n is large is that the impact of a confounding variable ($r_{y \cdot cv} \times r_{x \cdot cv}$) must approach $r_{y \cdot x}$ for $\hat{\beta}_1$ to become not statistically significant. Using Cohen and Cohen's (1983) levels of small (.1), medium (.3), and large (.5) effects for correlations, the large sample rule implies that $r_{y \cdot cv}$ and $r_{x \cdot cv}$ both must be at least one level stronger than $r_{y \cdot x}$ to alter the inference of β_1 (assuming $r^2_{x \cdot cv} = r^2_{y \cdot cv}$).

The IT also is sensitive to the specified level of α. The relationships between the IT and $r_{y \cdot x}$ for levels of α of .10, .05, and .01 for a sample of 85 are shown in Figure 3. In general, the larger the value of α, the larger is the IT. This is intuitive, since it takes a greater impact for $t(\hat{\beta}_1)$ to cross the higher threshold. As in Figure 2, the slopes of the lines in Figure 3 are not parallel. A given change in $r_{y \cdot x}$ is associated with relatively more change in the ITCV for small values of α than for large values of α. In contrast, a given change in $r_{y \cdot x}$ is associated with relatively less change in the ITSV for small values of α than for large values of α.

I have developed the IT with respect to the significance test for $\hat{\beta}_1$ or $r_{y \cdot x}$. Because of the association between confidence intervals and significance tests, one could state the index in terms of the impact necessary to make the confidence interval for $\hat{\beta}_1$ have a boundary exactly at zero. Furthermore, one could state the impact in terms of standardized regression coefficients of cv on x ($\beta^*_{x \cdot cv}$) and of cv on y ($\beta^*_{y \cdot cv}$). The regression of cv on x controls for no other variable, so the standardized coefficient simply equals $r_{x \cdot cv}$. The regression on y includes both x and cv. Therefore, the impact expressed in terms of standardized coefficients is

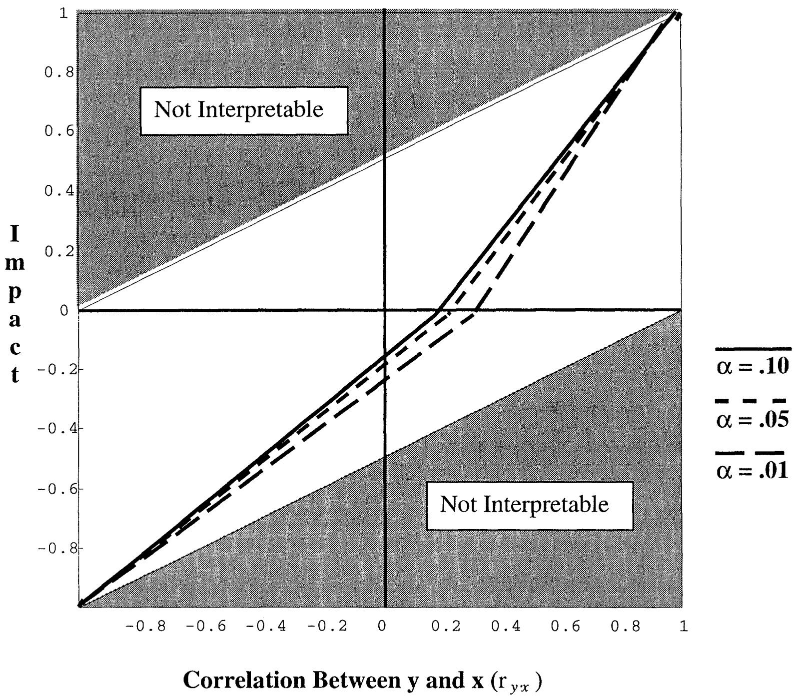

Figure 3: Impact Threshold and Correlation Between *y* and *x* for Different Significance Levels (*n* = 85)

$$k(\beta^*_{x\cdot cv}, \beta^*_{y\cdot cv}) = \beta^*_{x\cdot cv} \times \beta^*_{y\cdot cv} = r_{x\cdot cv} \times \frac{r_{y\cdot cv} - r_{y\cdot x} r_{x\cdot cv}}{1 - r^2_{x\cdot cv}} =$$

$$\frac{r_{x\cdot cv} r_{y\cdot cv} - r_{y\cdot x} r^2_{x\cdot cv}}{1 - r^2_{x\cdot cv}} = \frac{k(1 - r_{y\cdot x})}{1 - k}, \tag{7}$$

assuming $r^2_{x\cdot cv} = r^2_{y\cdot cv} = r_{x\cdot cv} \times r_{y\cdot cv} = k$. One can also calculate an impact in terms of unstandardized regression coefficients:

$$k(\beta_{x\cdot cv}, \beta_{y\cdot cv}) = \beta_{x\cdot cv} \times \beta_{y\cdot cv} = \frac{k(1 - r_{y\cdot x})}{1 - k} \times \frac{sd(x)sd(y)}{\text{var}(cv)}. \tag{8}$$

By replacing k in (7) or (8) with the values from (6), one can reexpress the ITCV and the ITSV in terms of standardized regression coefficients, $k(\beta^*_{x \cdot cv}, \beta^*_{y \cdot cv})$, or unstandardized regression coefficients, $k(\beta_{x \cdot cv}, \beta_{y \cdot cv})$.

RELATIONSHIP TO OTHER STATISTICAL APPROACHES/PROCEDURES

An extensive search of the literature including applied statistics textbooks (Chatterjee and Hadi 1988; Cohen and Cohen 1983; Draper and Smith 1981; Hanushek 1977; Montgomery and Peck 1982; Pedhazur 1982; Rice 1988; Ryan 1997; Snedecor and Cochran 1989; Weisberg 1985) did not reveal an index comparable to the IT. The presentation here has commonalities with many statistical approaches but differs in important ways.

SPURIOUS REGRESSION AND BIAS IN $\hat{\beta}_1$

Like the spurious regression coefficient and bias induced by correlation between x and e that could be attributed to cv (see Cohen and Cohen 1983; Draper and Smith 1981; Pedhazur 1982; Snedecor and Cochran 1989), the ITCV focuses on the impact a confounding variable may have on the interpretation of $\hat{\beta}_1$. But whereas bias and spurious regression focus on the coefficient $\hat{\beta}_1$ or on tests for misspecification (Godfrey 1988; Hausman 1978), the IT focuses on the statistical inference of β_1. Furthermore, the development here expresses impact in terms of $r_{y \cdot cv} \times r_{x \cdot cv}$.

SIZE RULE

Developed mostly for biomedical applications, the size rule accounts for the impact of a confounding variable on a coefficient in log-linear models for contingency tables (Bross 1966, 1967; see also Cornfield et al. 1959; Lin, Psaty, and Kronmal 1998; Schlesselman 1978). Like the size rule, the ITCV is based on the relationship between the confound and the predictor of interest, and between the confound and the outcome. But the IT applies to linear models and focuses on the impact on inference rather than on the change in the size of the coefficient.

INSTRUMENTAL VARIABLES

Like the instrumental variable (e.g., Angrist, Imbens, and Rubin 1996; Hanushek 1977; Hausman 1978; Heckman and Robb 1985), the IT concerns relationships between x and a second independent variable. Furthermore, statistical tests of $\hat{\beta}_1$ in the instrumental variables approach have been developed (Hausman 1978), and the instrumental variable has been considered for its effect on causal inference (Angrist et al. 1996). But in the development of the IT, the third variable (e.g., cv) is presumed correlated with both x and y, whereas the instrumental variable is assumed to be uncorrelated with the outcome y. The instrument is not considered an alternate cause of x and y but an alternate measurement of x. Thus, the instrumental variable is used to reduce bias in estimation of β_1, whereas the ITCV is used to assess the sensitivity of $\hat{\beta}_1$ and its standard error to the inclusion of a confounding variable.

COLLAPSIBILITY

A confounding variable is collapsible with respect to x and y if including the confounding variable does not statistically significantly alter $\hat{\beta}_1$ (Clogg, Petkova, and Shihadeh 1992). Thus, collapsibility concerns the statistical test of the difference between $\hat{\beta}_1$ in model (1a) without cv and $\hat{\beta}_1$ in model (1b) with cv (e.g., Clogg et al. 1992; Cochran 1970). Using tests for collapsibility and a given value of $t(\hat{\beta}_1)$ in model (1a), one could construct a test for $\hat{\beta}_1$ in model (2a). But this initial step has not been explored within the context of collapsibility, nor has an index been developed based on the components $r_{y \cdot cv}$ and $r_{x \cdot cv}$ as is done for the IT.

OMITTED VARIABLES

Many statistics test whether an omitted variable should be included in a model (e.g., Chatterjee and Hadi 1988; Ramsey 1974). Like the IT, these statistics address the impact of the omitted variable on parameter estimates and inferences. But statistics for omitted variables typically are expressed in terms of changes in sums of squares of residuals or other measures of fit and are not easily reduced to indexes of correlations associated with the omitted variable. In other words, it is assumed that measures of the omitted variables are available and their impact on parameter estimates can be calculated from available data. The expression here indexes the impact for unmeasured variables. Furthermore, the ITCV is developed to capture the impact of a special type of omitted variable, the confounding variable, that a skeptic has used to challenge an inference with regard to a regression coefficient.

CAUSAL MODELS AND INDIRECT EFFECTS

If the role of the covariate were that of a mediator instead of a confounding variable, then the "impact" could be interpreted as the indirect effect of *x* on *y* (Asher 1983; Holland 1988; Pedhazur 1982; Sobel 1998). That is, indirect effects can be thought of in terms of $r_{y \cdot m} \times r_{x \cdot m}$, where *m* represents a mediating variable (one that is considered to be caused *by* the predictor of interest). But even given this similarity, the indirect effect is typically evaluated for its impact on the direct effect, not for its impact on the statistical inference of the direct effect. Moreover, the development of the ITCV draws on a theoretical framework embedded in arguments of control for a confounding variable (e.g., the motivation for *minimizing* the expression for *t* to maximize the impact of the confounding variable).

SENSITIVITY ANALYSIS

Rosenbaum (1986) calculated the sensitivity of an estimated treatment effect to differences between treatment and control groups on a confounding variable, and Mauro (1990) calculated the sensitivity of inferences for regression coefficients for hypothesized impacts of confounding variables. Like the ITCV, these sensitivity analyses prospectively account for the impact of a confounding variable. But the ITCV neatly summarizes the threshold value as a function of the two critical pieces of information associated with the confound: $r_{y \cdot cv}$ and $r_{x \cdot cv}$. Defining the ITCV in terms of $r_{y \cdot cv} \times r_{x \cdot cv}$ greatly reduces the information required to assess the potential impact. Furthermore, the concise expression of the index facilitates the extension to the multivariate case and the exploration of a reference distribution, topics that are developed in subsequent sections.[7]

MULTIVARIATE EXTENSIONS

First, consider models in which there are a set of covariates, $z1$, $z2, \ldots, zp$, in the original model, but we still examine the impact of a single confound:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 z1_i + \ldots + \beta_{p+1} zp_i + e_i \tag{9a}$$

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 z1_i + \ldots + \beta_{p+1} zp_i + \beta_{p+2} cv_i + e_1. \tag{9b}$$

In the Catholic schools example, the set of covariates might include educational aspirations, family ability to afford private education, and so on. Model (9a) includes these covariates but still does not include the confounding variable, prior achievement. Thus, there is still the argument that the relationship associated with Catholic schools (represented by β_1) could be attributed to prior achievement. This suggests that β_1 is zero in model (9b), which includes the confounding variable.

Define **z** as a vector containing elements ($z1, \ldots, zp$) and define the partial correlation between y and x as $r_{y\cdot x|(\mathbf{z},cv)}$. This can be understood as the correlation between y and x, where both y and x have been controlled for **z** and cv. For model (9b), we have

$$t(\hat{\beta}_1) = \frac{r_{y\cdot x|(\mathbf{z},cv)}}{\sqrt{\dfrac{1 - r^2_{y\cdot x|(\mathbf{z},cv)}}{n-q-1}}}. \tag{10}$$

The expression for $r_{y\cdot x|(\mathbf{z},cv)}$ can be obtained as follows:

$$r_{y\cdot x|(\mathbf{z},cv)} = \frac{r_{y\cdot x|\mathbf{z}} - r_{y\cdot cv|\mathbf{z}} r_{x\cdot cv|\mathbf{z}}}{\sqrt{(1 - r^2_{y\cdot cv|\mathbf{z}})(1 - r^2_{x\cdot cv|\mathbf{z}})}},$$

where

$$r_{y\cdot cv|\mathbf{z}} = \frac{r_{y\cdot cv} - r_{y\cdot \mathbf{z}} r_{cv\cdot \mathbf{z}}}{\sqrt{(1 - r^2_{y\cdot \mathbf{z}})(1 - r^2_{cv\cdot \mathbf{z}})}} = \frac{r_{y\cdot cv}}{\sqrt{(1 - r^2_{y\cdot \mathbf{z}})}}$$

and

$$r_{x\cdot cv|\mathbf{z}} = \frac{r_{x\cdot cv} - r_{x\cdot \mathbf{z}} r_{cv\cdot \mathbf{z}}}{\sqrt{(1-r_{x\cdot \mathbf{z}}^2)(1-r_{cv\cdot \mathbf{z}}^2)}} = \frac{r_{x\cdot cv}}{\sqrt{(1-r_{x\cdot \mathbf{z}}^2)}}. \tag{11}$$

The latter two equivalences are based on the assumption that $r_{cv\cdot\mathbf{z}} = 0$. That is, the potential confounding variable is uncorrelated with the covariates already in the model. In general, if $r_{cv\cdot\mathbf{z}} > 0$, then $t(\hat{\beta}_1)$ in model (9b) is inflated, thus weakening the argument that including a confounding variable in the model would make $\hat{\beta}_1$ not statistically significant. As Pratt and Schlaifer (1988) argued, once x is independent of cv, regardless of whether cv is in the model, one can interpret $\hat{\beta}_1$ as consistent and unbiased and, therefore, supportive of a law. In other words, if the correlation between cv and x is absorbed through the correlation between x and $\mathbf{z}$, then the concern with regard to the potential impact of cv is nullified.[8] Therefore, to give full weight to the skeptic's concern, I assume $r_{cv\cdot\mathbf{z}} = 0$.[9]

Note that $r_{y\cdot x|\mathbf{z}}$ contains no terms involving cv and is readily estimated from the data. Using the expressions in (11), the expression for $r_{y\cdot x|(\mathbf{z},cv)}$ is then

$$r_{y\cdot x|(\mathbf{z},cv)} = \frac{r_{y\cdot x|\mathbf{z}}\sqrt{(1-r_{y\cdot \mathbf{z}}^2)(1-r_{x\cdot \mathbf{z}}^2)} - r_{x\cdot cv} r_{y\cdot cv}}{\sqrt{(1-r_{y\cdot \mathbf{z}}^2 - r_{y\cdot cv}^2)(1-r_{x\cdot \mathbf{z}}^2 - r_{x\cdot cv}^2)}} \tag{12}$$

and

$$t(\hat{\beta}_1) = \frac{\dfrac{r_{y\cdot x|\mathbf{z}}\sqrt{(1-r_{y\cdot \mathbf{z}}^2)(1-r_{x\cdot \mathbf{z}}^2)} - r_{x\cdot cv} r_{y\cdot cv}}{\sqrt{(1-r_{y\cdot \mathbf{z}}^2 - r_{y\cdot cv}^2)(1-r_{x\cdot \mathbf{z}}^2 - r_{x\cdot cv}^2)}}}{\sqrt{\dfrac{1-\left(\dfrac{r_{y\cdot x|\mathbf{z}}\sqrt{(1-r_{y\cdot \mathbf{z}}^2)(1-r_{x\cdot \mathbf{z}}^2)} - r_{x\cdot cv} r_{y\cdot cv}}{\sqrt{(1-r_{y\cdot \mathbf{z}}^2 - r_{y\cdot cv}^2)(1-r_{x\cdot \mathbf{z}}^2 - r_{x\cdot cv}^2)}}\right)^2}{n-q-1}}} \tag{13}$$

$$= \frac{r_{y\cdot x|\mathbf{z}}\sqrt{(1-r_{y\cdot \mathbf{z}}^2)(1-r_{x\cdot \mathbf{z}}^2)} - r_{x\cdot cv} r_{y\cdot cv}}{\sqrt{\dfrac{(1-r_{y\cdot \mathbf{z}}^2 - r_{y\cdot cv}^2)(1-r_{x\cdot \mathbf{z}}^2 - r_{x\cdot cv}^2) - (r_{y\cdot x|\mathbf{z}}\sqrt{(1-r_{y\cdot \mathbf{z}}^2)(1-r_{x\cdot \mathbf{z}}^2)} - r_{x\cdot cv} r_{y\cdot cv})^2}{n-q-1}}}.$$

Thus, $t(\hat{\beta}_1)$ can be represented as a function of $r_{y\cdot x|\mathbf{z}}, r^2_{y\cdot \mathbf{z}}, r^2_{x\cdot \mathbf{z}}, r_{x\cdot cv}, r_{y\cdot cv}$ and the degrees of freedom $(n - q - 1)$. If there are no covariates, then (13) becomes

$$t(\hat{\beta}_1) = \frac{r_{y\cdot x} - r_{x\cdot cv} r_{y\cdot cv}}{\sqrt{\dfrac{(1 - r^2_{y\cdot cv})(1 - r^2_{x\cdot cv}) - (r_{y\cdot x} - r_{x\cdot cv} r_{y\cdot cv})^2}{n - q - 1}}}. \tag{14}$$

Some algebra shows this latter expression is equivalent to the expression in (3) when we set $r_{y\cdot cv} \times r_{x\cdot cv} = k$.

There is a question of whether we can substitute $r^2_{x\cdot cv} = r^2_{y\cdot cv} = k$ as in the bivariate case. Given the constraint $r_{y\cdot cv} \times r_{x\cdot cv} = k$, note that (10) and (13) are minimized when $r_{y\cdot x|(\mathbf{z}, cv)}$ as defined in (12) is minimized. Given n, q, $r_{y\cdot x|\mathbf{z}}$, $r_{y\cdot \mathbf{z}}$ and $r_{x\cdot \mathbf{z}}$, the term $r_{y\cdot x|(\mathbf{z}, cv)}$ is minimized when $(1 - r^2_{y\cdot \mathbf{z}} - r^2_{y\cdot cv})(1 - r^2_{x\cdot \mathbf{z}} - r^2_{x\cdot cv})$ is maximized. Using a LaGrange multiplier to obtain the constrained maximum similar to the bivariate case, we have

$$\begin{aligned}
&(-2r_{y\cdot cv}(1 - r^2_{x\cdot \mathbf{z}}) + 2r_{y\cdot cv} r^2_{x\cdot cv})i + (-2r_{x\cdot cv}(1 - r^2_{y\cdot \mathbf{z}}) + 2r_{x\cdot cv} r^2_{y\cdot cv})j \\
&= \lambda(r_{x\cdot cv} i + r_{y\cdot cv} j) \\
&\Rightarrow \frac{r^2_{x\cdot cv} - (1 - r^2_{x\cdot \mathbf{z}})}{r^2_{x\cdot cv}} = \frac{r^2_{y\cdot cv} - (1 - r^2_{y\cdot \mathbf{z}})}{r^2_{y\cdot cv}} \\
&\Rightarrow \frac{r^2_{y\cdot cv}}{r^2_{x\cdot cv}} = \frac{(1 - r^2_{y\cdot \mathbf{z}})}{(1 - r^2_{x\cdot \mathbf{z}})}.
\end{aligned} \tag{15}$$

Thus, $t(\hat{\beta}_1)$ is minimized when the ratio of $r^2_{y\cdot cv}$ to $r^2_{x\cdot cv}$ corresponds to $(1 - r^2_{y\cdot \mathbf{z}})(1 - r^2_{x\cdot \mathbf{z}})$. Substituting into the original constraint, $r_{y\cdot cv} \times r_{x\cdot cv} - k = 0$, we have

$$r^2_{y\cdot cv} = k\sqrt{\frac{(1 - r^2_{y\cdot \mathbf{z}})}{(1 - r^2_{x\cdot \mathbf{z}})}} \tag{16}$$

$$r^2_{x\cdot cv} = k\sqrt{\frac{(1 - r^2_{x\cdot \mathbf{z}})}{(1 - r^2_{y\cdot \mathbf{z}})}}.$$

Note that if $r^2_{y \cdot \mathbf{z}} = r^2_{x \cdot \mathbf{z}}$ such that $(1 - r^2_{y \cdot \mathbf{z}}) / (1 - r^2_{x \cdot \mathbf{z}}) = 1$, then the last expression in (15) implies $r^2_{y \cdot cv} = r^2_{x \cdot cv}$, as was the case when there were no covariates in the model. Note also, as in the bivariate case, that this development applies to $r_{y \cdot x|\mathbf{z}} > 0$ and $k > 0$, while the complementary cases can be developed through symmetric arguments.

To solve for k, begin by substituting the equalities in (16) back into (13):

$$t(\hat{\beta}_1) = \frac{r_{y \cdot x|\mathbf{z}}\sqrt{(1 - r^2_{y \cdot \mathbf{z}})(1 - r^2_{x \cdot \mathbf{z}})} - k}{\sqrt{\dfrac{\left(1 - r^2_{y \cdot \mathbf{z}} - k\sqrt{\dfrac{(1 - r^2_{y \cdot \mathbf{z}})}{(1 - r^2_{x \cdot \mathbf{z}})}}\right)\left(1 - r^2_{x \cdot \mathbf{z}} - k\sqrt{\dfrac{(1 - r^2_{x \cdot \mathbf{z}})}{(1 - r^2_{y \cdot \mathbf{z}})}}\right) - \left(r_{y \cdot x|\mathbf{z}}\sqrt{(1 - r^2_{y \cdot \mathbf{z}})(1 - r^2_{x \cdot \mathbf{z}})} - k\right)^2}{n - q - 1}}}. \quad (17)$$

Letting $t = t(\hat{\beta}_1)$ and solving for k, we have

$$k = \frac{\sqrt{(1 - r^2_{x \cdot \mathbf{z}})(1 - r^2_{y \cdot \mathbf{z}})}\left[\begin{array}{c}(t^2(1 - r_{y \cdot x|\mathbf{z}}) - (n - q - 1)r_{y \cdot x|\mathbf{z}}) \pm \\ \sqrt{\begin{array}{l}(t^2(1 - r_{y \cdot x|\mathbf{z}}) - (n - q - 1)r_{y \cdot x|\mathbf{z}})^2 + \\ (n - q - 1)(t^2(1 - r^2_{y \cdot x|\mathbf{z}}) - (n - q - 1)r^2_{y \cdot x|\mathbf{z}})\end{array}}\end{array}\right]}{-(n - q - 1)}.$$

Again, define $d = t^2 + (n - q - 1)$. Then,

$$k = ITCV = \left(\sqrt{(1 - r^2_{x \cdot \mathbf{z}})(1 - r^2_{y \cdot \mathbf{z}})}\right)\left(\frac{t^2 + t\sqrt{d}}{-(n - q - 1)} + \left[\frac{-d - t\sqrt{d}}{-(n - q - 1)}\right]r_{y \cdot x|\mathbf{z}}\right)$$

or (18)

$$k = ITSV = \left(\sqrt{(1 - r^2_{x \cdot \mathbf{z}})(1 - r^2_{y \cdot \mathbf{z}})}\right)\left(\frac{t^2 - t\sqrt{d}}{-(n - q - 1)} + \left[\frac{-d + t\sqrt{d}}{-(n - q - 1)}\right]r_{y \cdot x|\mathbf{z}}\right).$$

The right-hand side of the first expression in (18) is equivalent to the right-hand side in (6) multiplied by $[(1 - r^2_{y \cdot \mathbf{z}})(1 - r^2_{x \cdot \mathbf{z}})]^{1/2}$ and replacing

$r_{y\cdot x}$ with $r_{y\cdot x|\mathbf{z}}$. Thus, k is again linearly related to $r_{y\cdot x|\mathbf{z}}$, and assigning $t = t_{\text{critical}}$ gives the ITCV and the ITSV as defined in the bivariate case.

Now consider the possibility of a set of confounds defined by the vector **cv**: *cv*1, . . ., *cvg*. The models are

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 z1_i + \ldots + \beta_{p+1} zp_i + e_i \qquad (19a)$$

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 z1_i + \ldots + \beta_{p+1} zp_i + \beta_{p+2} cv1_i + \ldots + \beta_{g+p+1} cvg_i + e_i . \qquad (19b)$$

This might be of particular interest when the skeptic proposes multiple confounds or when the confounds include a set of dummy variables to represent group membership. In the Catholic schools example, a set of dummy variables might be used to represent the religious affiliation of the student. The solution emerges directly from (18) by redefining k in terms of $r_{x\cdot\mathbf{cv}} \times r_{y\cdot\mathbf{cv}}$, where $r_{x\cdot\mathbf{cv}}$ represents the square root of the multiple correlation between x and **cv** and $r_{y\cdot\mathbf{cv}}$ represents the square root of the multiple correlation between y and **cv**. Note that in addition to $r_{x\cdot cv} \times r_{y\cdot cv}$, the terms involving cv in (13) are $r^2_{y\cdot cv}$, $r^2_{x\cdot cv}$. Each term will be positive provided $r_{x\cdot cv}$ and $r_{y\cdot cv}$ take the same sign, which will hold if k is defined as confounding for $r_{x\cdot y} > 0$.

EXAMPLE: FAMILY BACKGROUND AND EDUCATIONAL ATTAINMENT

The relationship between family background and educational attainment[10] has been the focus of considerable sociological attention over the years. For example, Featherman and Hauser (1976) estimated the effect of family background on educational attainment using Blau and Duncan's (1967) occupational changes in generation data. Recently, Sobel (1998) critiqued the inference made by Featherman and Hauser on several grounds. One of Sobel's primary arguments was that "it is not plausible to think that background variables satisfy the assumption of conditional random assignment" (p. 343). In particular, Sobel argued that "father's education should be a predictor of son's education" (p. 334). That is, father's education is a confounding variable for the relationship between background characteristics and educational attainment. Therefore, the coefficients estimated from Featherman and Hauser's data cannot be interpreted as effects.

Formally, define y to represent the outcome, educational attainment, and x to represent the predictor of interest, background characteristics. The cv represents Sobel's proposed confounding variable, father's education. Sobel's critique is represented in Figure 4. Begin with the standard representation of the relationship between x and y, referred to as β_1, associated with the arrow at the top of the figure. Then, introduce the concern for the confounding variable in terms of the relationships associated with the confounding variable, $r_{x \cdot cv}$ and $r_{y \cdot cv}$. The impact is then expressed in terms of the arrows emerging from $r_{x \cdot cv}$ and $r_{y \cdot cv}$ and converging to represent the product $r_{x \cdot cv} \times r_{y \cdot cv}$ which then affects β_1.

To assess Sobel's (1998) critique, begin by examining the initial correlation matrix for the variables in the regression analysis. The correlations for men in 1962 as reported in the upper left of Featherman and Hauser's (1976) Table 1 are represented in Table 1. Define father's occupation to be the specific predictor of interest and observe that father's occupation is strongly correlated with educational attainment (.427). Then, farm origin and number of siblings are defined as covariates in estimating the relationship between father's occupation and educational attainment (father's occupation is correlated –.437 with farm origin and –.289 with number of siblings).

When the covariates are controlled for, we obtain the regression results in Table 2 for married civilian men whose spouses were present.[11] From these and other analyses, Featherman and Hauser (1976) concluded that family background has an effect on educational attainment. Sobel (1998) countered, arguing that father's education affects both family background and educational attainment but has not been controlled for in the model. The question then is not whether father's education is related both to family background and to educational attainment. Undoubtedly it is. Focusing on father's occupation as an example of family background, a more specific question is "How large must be the correlations between father's education and father's occupation, and between father's education and educational attainment to alter the inference with regard to father's occupation?"

From the correlations in Table 1, the necessary impact for a confounding variable to alter the inference with regard to father's occupation

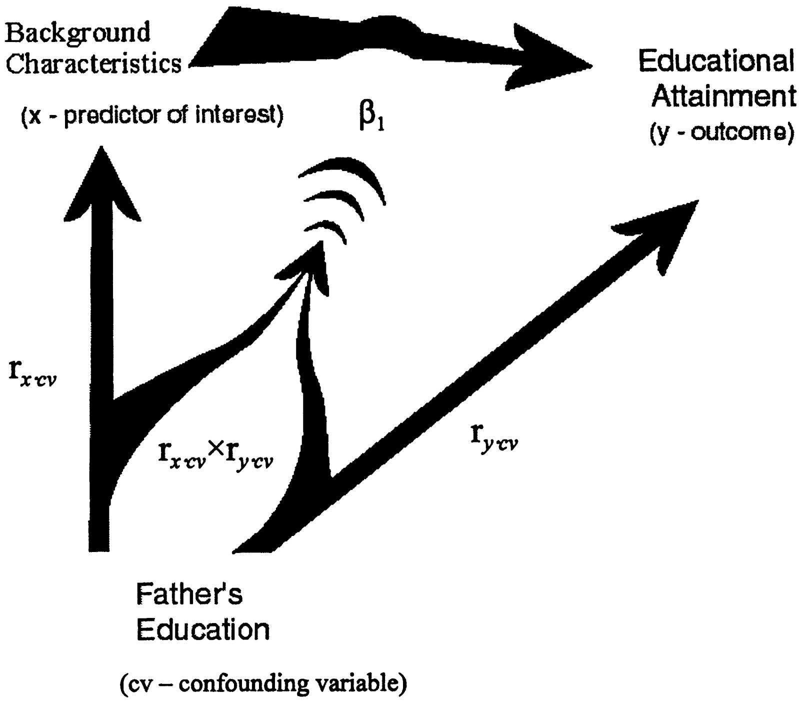

Figure 4: Path Diagram of the Impact of a Confounding Variable on a Regression Coefficient

TABLE 1: Correlations Between Background Characteristics and Educational Attainment for Featherman and Hauser's (1976) Data

	Father's Occupation	*Farm Origin*	*Number of Siblings*	*Educational Attainment*
Father's occupation	1.000	–.437	–.289	.427
Farm origin		1.000	.263	–.317
Number of siblings			1.000	–.350
Educational attainment				1.000

as captured by the ITCV from (18) is .228 (see Table 2). That is, for the coefficient associated with father's occupation to become not statistically significant, the impact of a confounding variable must be greater

TABLE 2: Regression of Educational Attainment on Background Characteristics in Featherman and Hauser's (1976) Data

Variable	*Coefficient*	*Standard Error*	t *Ratio*	p *Value*	*ITCV*	*Impact* k[a]
Intercept	10.9047	.0811	134.4	≤.0001		
Father's occupation	.0504	.0016	32.1	≤.0001	.228	
Farm origin	–.9213	.0714	–12.9	≤.0001		.058
Number of siblings	–.2871	.0111	–25.7	≤.0001		.097

NOTE: $N = 10{,}567$, $R^2 = .25$. ITCV = impact threshold for a confounding variable.
a. Values of k are calculated relative to the primary variable of interest, father's occupation.

than .228. Assuming that both correlations are positive, the minimum value of k for which t is not statistically significant is achieved when $r_{y \cdot cv|\mathbf{z}}$=.484 and $r_{x \cdot cv|\mathbf{z}}$=.470. The two values are not equal because of the different ratios defined in (16).

The partial correlations of at least .47 needed to alter the interpretation with regard to father's occupation are themselves close to large effects in the social sciences (Cohen and Cohen 1983). That is, the unique correlation between educational attainment and father's education and the unique correlation between father's occupation and father's education would each have to be large in social science terms to alter the inference with regard to the coefficient for father's occupation. Because it is likely that father's education is correlated with the other covariates in the model, farm origin and number of siblings, the zero-order correlations between educational attainment and father's education and between father's occupation and father's education would have to be larger than .47 for the corresponding partial correlations to be at least .47.

Although we cannot yet speak of the likelihood of the ITCV exceeding .228, we can now consider the magnitude of the impact necessary to sustain Sobel's (1998) concern with regard to Featherman and Hauser's (1976) interpretation of the estimated regression coefficient for father's occupation as an effect. Given the size of the ITCV, it is now incumbent upon the skeptic, in this case Sobel, to argue or demonstrate that partial correlations on the order of .47 are likely to be observed.[12]

A REFERENCE DISTRIBUTION FOR THE IMPACT OF A CONFOUND

Although the ITCV quantifies the concern of the skeptic, we only have a rough baseline in terms of "large effects in the social sciences" by which to evaluate the ITCV. But one can use the impacts of existing covariates to generate a more specific and informative reference distribution.[13] Here, covariates include measured variables considered to be confounding; that is, the covariates include variables that are correlated with both the predictor of interest and the outcome (Holland 1986, 1988). The use of existing covariates here is similar to Rosenbaum's (1986) for a sensitivity analyses in his Table 8.

For Featherman and Hauser's (1976) data, if the impact of the confounding variable, father's education, were represented by the existing impact of farm origin, the value would be .058:

$$r_{y \cdot farm\ origin|number\ of\ siblings} \times r_{x \cdot farm\ origin|number\ of\ siblings}$$

$$= -.29 \times -.20 = .058.$$

The impact for number of siblings would be .097 (these values are reported in the last column of Table 2). Note that these values are considerably smaller than the threshold value of .228 necessary to alter the inference regarding father's occupation.

These two data points can be supplemented using the correlation matrices reported by Duncan, Featherman, and Duncan (1972) for relationships between background characteristics and Educational Attainment for various samples and data sets. Covariates included number of siblings, importance of getting ahead, brother's education, and father in farming. Data sets included Blau and Duncan's (1967) occupational changes in generation study and the family growth in metropolitan America study. A distribution of Fisher's z transformation of 14 estimates of impact is shown in Figure 5.

Fisher's z transform of the ITCV of .228 is .232 (Fisher's z has little effect for values close to zero), and all of the values in Figure 5 are less than .232. The largest value, represented at roughly .2 in the figure, is .188.

We can use information such as in Figure 5 to formally test the hypothesis with regard to the ITCV. The general alternatives are as follows:

A_1: The impact of a confounding variable is greater than the ITCV (predictor of interest).
A_2: The impact of a confounding variable is less than or equal to the ITCV (predictor of interest).

In this particular case, we have the following:

A_1: The impact of father's education is greater than the ITCV (father's occupation).
A_2: The impact of father's education is less than or equal to the ITCV (father's occupation).

Here, A_1 represents the skeptic's claim that the inference with regard to father's occupation would change if father's education were in the model. We reject A_1 if we determine that P[impact of father's education $\geq$ ITCV (father's occupation)] $< \alpha$, where α is some prespecified level of probability, such as .05. The key to assessing the alternatives is to assume the impacts of the observed covariates define a reference distribution for the impact of the unobserved confounding variable. That is, operationally the alternatives are as follows:

A_1: The impact of a covariate is greater than the ITCV (father's occupation).
A_2: The impact of a covariate is less than or equal to the ITCV (father's occupation).

Thus, the reference distribution can be used to assess the likelihood that the impact of a covariate will exceed the ITCV of a predictor of interest. In Featherman and Hauser's (1976) data, all 14 observed estimates of impact in Figure 5 are less than the ITCV for father's occupation of .228, and therefore we can reject A_1 with $p \leq 1/14$ or $p \leq .08$.[14]

We could also assess the above alternatives in terms of a theoretical reference distribution for the impact (which is based on the product of two correlation coefficients) using statistics estimated from the observed impact values. One approximation is to apply Fisher's z transformation to each component of the impact, making each component approx-

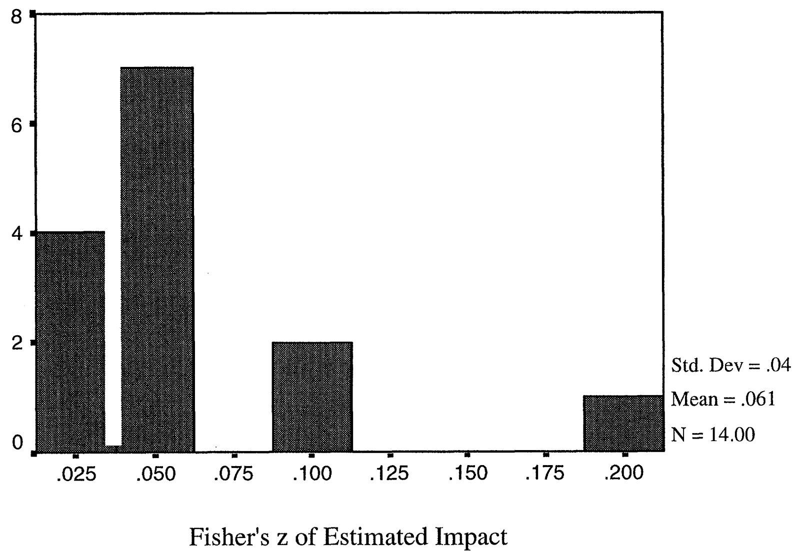

Figure 5: Reference Distribution of Impact for Covariates in Featherman and Hauser's (1976) Data

imately normally distributed.[15] That is, define a reference distribution for $k^* = z(\rho_{y \cdot cv|\mathbf{z}}) \times z(\rho_{x \cdot cv|\mathbf{z}})$.

To assess the probability of obtaining a given value of k^* or greater, begin by defining μ_1 and σ_1 as the mean and variance of $z(\rho_{x \cdot cv|\mathbf{z}})$, respectively, and μ_2 and σ_2 as the mean and variance of $z(\rho_{y \cdot cv|\mathbf{z}})$, respectively. The distribution for the product of two normal deviates is based on three parameters, $\delta_1 = \mu_1/\sigma_1$, $\delta_2 = \mu_2/\sigma_2$, and ρ, representing the correlation between $z(\rho_{x \cdot cv|\mathbf{z}})$ and $z(\rho_{y \cdot cv|\mathbf{z}})$. The estimates of these parameters for Duncan et al.'s (1972) covariates are $\hat{\delta}_1 = {}^{.2684}\!/_{.1029} = 2.608$, $\hat{\delta}_2 = {}^{.2431}\!/_{.1190} = 2.043$, and $\hat{\rho} = -.0125$.

Now, define $k^{**} = z(\rho_{y \cdot cv|\mathbf{z}}) \times z(\rho_{x \cdot cv|\mathbf{z}})/(\sigma_1\sigma_2)$, thus k^{**} is adjusted for the variances. The distribution of k^{**}, the product of two normally distributed variables (with nonzero correlation), is given by Aroian, Taneja, and Cornwell (1978), and the algorithm for integration is given by Meeker and Escobar (1994).[16] Meeker, Cornwell, and Aroian (1980) tabulated the percentages of the cumulative distribution for w, which is a standardized version of k^{**}.

In the Duncan et al. (1972) covariates, the sample mean of k^{**} is 2.25 and the sample standard deviation is 1.68. Given the other estimated parameter values, the critical value of w for $p \leq .05$ is 1.83. The transformed ITCV for father's occupation is 4.20, which is statistically significant at $p \leq .001$ (using no interpolation but instead rounding the value of $\hat{\delta}_1$ up to 2.4 and the value of $\hat{\delta}_2$ up to 2.8 to obtain a conservative p value).

Note that the percentages of the cumulative distribution for w do not account for the fact that several of the parameters were estimated, a critical issue for such a small ($n = 14$) sample size. Without the correction, the reported p value is likely to be too small. Note also the critical assumption that the impacts of the covariates are representative of the impact of the confounding variable. In this case, one could argue the impact of father's education would be most like the extreme impact (.188) of brother's education (taken from Duncan et al. 1972, appendix 1). If the impact of brother's education were most representative of the impact of father's education, it is quite possible that the impact of father's education would exceed the ITCV of .228. This highlights the importance of considering the representativeness of the covariates for making inferences with regard to an ITCV.

As it happens, the impact of father's education can be partially assessed in terms of data in Duncan et al. (1972, appendix A1). In these data, the correlation between educational attainment and father's education ($r_{y \cdot cv}$) was estimated to be .42 and the correlation between father's occupation and father's education ($r_{x \cdot cv}$) was estimated to be .49. The unadjusted impact is .206, and similar unadjusted impacts based on the appendix in Sewell, Hauser, and Wolf (1980) are .16 for women and .17 for men. Using Sewell et al., adjusting for number of siblings and farm origin, the correlation for men between educational attainment and father's education ($r_{y \cdot cv}$) was estimated to be .29 and the correlation between father's occupation and father's education ($r_{x \cdot cv}$) was estimated to be .47. The product gives an impact of .13. The impact for women was slightly smaller.

The adjusted, as well as the unadjusted, impacts are less than the ITCV for father's occupation of .228. Not surprisingly, in Sewell et al. (1980, tables 8, 9), father's occupation has a statistically significant direct effect on educational attainment for men or women when controlling for several background characteristics (including parental income, mother's education, mother's employment, farm origin, intact family, number of siblings, and mental ability, as well as father's education) and some mediating variables (including high school grades, parental encouragement, and teacher's encouragement).[17] Even though many of these variables have small to moderate correlations with both educational attainment and father's occupation, as a set their impact on the coefficient for father's occupation is only .424 because many of the covariates are intercorrelated. The impact of the set is less than the ITCV for the original correlation between father's occupation and educational attainment, and therefore, even after controlling for the set, the relationship between father's occupation and educational attainment is statistically significant. Nonetheless, we are not always fortunate enough to have estimates of correlations associated with the confounding variable. The critical point here is that the distribution of the impact of measured covariates may be used to assess the likelihood that a given impact will exceed the ITCV.

DISCUSSION

USE OF THE ITCV: INFORMING CAUSAL INFERENCE

One of the primary concerns in making causal inferences is that an observed relationship could be attributed to an alternate mechanism. But one can ask, "How large must be the impact of a confounding variable to alter an inference?" The question can be quantified by indexing the impact as $r_{y \cdot cv} \times r_{x \cdot cv}$ and then obtaining the threshold for which the original inference would be altered. No other technique indexes the impact in terms of $r_{y \cdot cv} \times r_{x \cdot cv}$ and applies the impact to unmeasured confounds.

I demonstrated that the bivariate ITCV depends on the original correlation between the dependent variable and the independent variable, $r_{y \cdot x}$, the sample size, and the specified level of α. In general, the larger the $r_{y \cdot x}$ the larger the sample, and the larger the α, the larger the impact to alter an inference must be. But Figures 1, 2, and 3 show that the relationship between the impact and $r_{y \cdot x}$ depends on sample size and α as well as whether one focuses on confounding or suppressing variables.

The ITSV is itself an important by-product of the development in this article, indicating the threshold at which a suppressor variable would alter a statistical inference.

Before applying the ITCV to an example, I generalized the ITCV to cases with multiple covariates or multiple confounds. This is necessary for most applications. Then, in the example of Featherman and Hauser's (1976) data raised by Sobel (1998), an ITCV for father's occupation of .228 indicates that a confounding variable would have to be correlated (partialing out covariates and maximizing the impact) both with father's occupation and with educational attainment at .47 or greater to alter the inference made with regard to father's occupation. The magnitude of the ITCV suggests that the component partial correlations would each have to be close to large by social science terms to alter the interpretation of the coefficient for father's occupation. By contextualizing the original inference, the ITCV informs a personal measure of uncertainty (Dempster 1988) or degree of belief (Rozenboom 1960) with regard to the regression coefficient. Thus, the ITCV takes causal inference beyond formal proofs and theorems (Freedman 1997) and embraces Duncan's "properly relativistic sociology" (cited in Abbott 1998:167).

Where the ITCV indicates the impact necessary to alter an inference, the reference distribution allows one to assess the *probability* of observing such an impact. Probability was introduced to causation through statements such as the following: "Assuming there is no common cause of *A* and *B*; if *A* happens, then *B* will happen with increased probability *p*" (Davis 1988; Einhorn and Hogarth 1986; Suppes 1970). But because probable causes do not absolutely link cause and effect, probable causes are open to challenges of alternative explanations associated with confounding variables. This has motivated many theories of causation (Dowe 1992; Mackie 1974; Reichenbach 1956; Salmon 1984, 1994; Sober 1988). Alternatively, use of the reference distribution provides a probabilistic response to challenges to "probable causes." In the Featherman and Hauser (1976) example, the values were $p \leq .08$ (for the observed distribution) and $p \leq .001$ (for the theoretical distribution) for the alternative that the impact of a covariate exceeds the ITCV of .228 for father's occupation. Thus, there is only a small probability that an alternate cause would account for the observed relationship between father's occupation and educational attainment.

Identifying the ITCV and use of the reference distribution does not absolutely rebut a skeptic's claim that the inclusion of a confounding variable would alter the inference of a regression coefficient. Rather, it allows researchers to quantitatively assess the skeptic's claim that the impact of the confound is large enough to alter an inference. (For a similar albeit nonquantitative argument, see Cordray [1986].) If the skeptic's arguments are not compelling, one can more strongly argue that a statistically significant coefficient is indicative of a causal relationship (although the size of the effect may still be undetermined). In this sense, causal inferences are neither absolutely affirmed nor rejected; rather, they are statements that are asserted and debated (e.g., Abbott 1998:164, 168; Cohen and Nagel 1934; Einhorn and Hogarth 1986; Gigerenzer 1993; Sober 1988; Thompson 1999).

Of course, causal inference cannot be asserted based on statistical criteria alone (see Dowe's [1992] critique of Salmon [1984]). A statistical relation combines with a general theory of causal processes (e.g., Salmon 1984, 1994), as well as a theory of a specific causal mechanism (see McDonald 1997; Scheines et al. 1998), to establish what Suppes (1970) described as a prima facie cause. Although Featherman and Hauser (1976) focused on gender differences, they generally argued that family background and resources could provide opportunities to pursue status attainment (including education). This theory combines with the statistically significant coefficients to establish family background as a prima facie cause of educational attainment. Prima facie causes are then separated into spurious causes and genuine causes depending on whether the effect can be attributed to a confounding variable. In the Featherman and Hauser example, the ITCV of .228 for father's occupation (and the small probability of observing such an ITCV of .228 or greater) is consistent with the assertion that father's occupation is a genuine cause of educational attainment.

Other research supports the claim that parental occupation has an effect on educational attainment and achievement net of father's education. At the micro level, Lareau (1989:110-14) found that parental occupation affected status differences between parents and teachers. These in turn affect the level and effectiveness of parental involvement in school, which in turn affects educational attainment. These effects are separate from effects of parental education, which are associated more with parental self-efficacy and competency. At the macro level, schooling institutions preserve and extend economic advantages into educational attainment, net of direct intergenerational transmission of educational levels (e.g., Bourdieu 1986; Bowles and Gintis 1976).

Beyond the specific example in Featherman and Hauser (1976), the ITCV is generally important for small or moderate effect sizes observed in large data sets. Pragmatically, we may not be able to alter those factors that have the strongest effects. (Holland [1986, 1988] would argue that such factors cannot even be considered causes.) But we may be able to alter several factors that have moderate effects. Thus, when the ITCV is large, we can commit further attention to studying and altering secondary, but nonetheless important, factors. An example of this is the effect of smoking on cancer, which has a small effect size, but for which there is a substantial accumulation of evidence (Gage 1978) and for which the statistical inference informs policy. Effect sizes from meta-analyses also can be enhanced through the ITCV, drawing on the strength of accumulated samples to establish the impact necessary to alter an inference.

The ITCV is also informative for large effect sizes that are statistically significant even in small samples. True, the boundary of a confidence interval close to zero also indicates caution in interpreting a coefficient. But a small ITCV more directly quantifies the concern that an interpretation may not be robust with respect to alternative causal explanations.

When the ITCV is *undisputedly* large, one can assert causality of one factor while acknowledging the possibility of other causal agents. In these circumstances, one need not contort theory or conduct experiments to defend the interpretation of a coefficient as an effect. This is consistent with how we approach causal inference from a cognitive perspective drawing on the philosophical positions of multiple causes and probable causes (Einhorn and Hogarth 1986; Mackie 1974; Meehl 1978; Mill [1843] 1973).

For the sake of clarity, I have presented most of the discussion in terms of the impact of a single confounding variable. Algebraically, extending the ITCV to include multiple covariates and confounds is straightforward, as demonstrated in equations (18) and (19). In terms of the debate on causal inference, the possibility of multiple confounding variables could overwhelm an ITCV of almost any value. Here, I believe the standard for the skeptic should be the same as for the researcher. Where strong and direct chains of causality are more persuasive in developing theory and presenting results (Einhorn and Hogarth 1986), so such chains are more persuasive in the critique. Where the researcher is called upon to consider theory, research standards, and parsimony in developing a model, so should the skeptic be required to consider a similar mix in proposing a set of confounding variables. Furthermore, multiple confounds are accounted for through the multiple correlation between the confounds and x, and therefore an argument must be made for the *unique* impact of each confound on the relationship between x and y. These unique impacts must be net of existing covariates in the model as well as other proposed confounds.

CONCERNS WITH REGARD TO SIGNIFICANCE TESTING

I have defined the ITCV relative to a change in statistical inference. Led by Cohen, many researchers have recently questioned the use of, or called for abandoning, significance tests (Cohen 1990, 1994; Gigerenzer 1993; Hunter 1997; Oakes 1986; Schmidt 1996; Sohn 1998),[18] although many of the arguments are not new (Bakan 1966; Carver 1978; Meehl 1978; Morrison and Henkel 1970; Rozenboom 1960). Very briefly, some concerns with regard to significance testing are as follows (for more extensive discussions, see Cortina and Dunlap 1997; Harlow, Mulaik, and Steiger 1997; Wilkinson and Task Force on Statistical Inference 1999):

- p values are misinterpreted as the probability the null hypothesis is true.
- p values are misinterpreted as the probability a statistical inference will be replicated.
- Small p values are mistaken to represent large effects (e.g., the more *'s, the better).
- The alpha level is an arbitrary cutoff, disrupting the incremental growth of science.
- The syllogistic reasoning of significance testing is illogical (incorporating probability into the deductive reasoning of the null and alternative hypotheses corrupts the logic) (see Cohen 1994:998).

There are important responses to many of these points. We can teach our students (and ourselves) how better to interpret p values (Schmidt 1996; Thompson 1999). We should consider changes in p values from prior to statistical analysis to posterior to statistical analysis (Grayson 1998; Pollard 1993). The null hypothesis is sensible for establishing a surprising effect (Abelson 1997) or the direction of an effect (Tukey 1991). Finally, the examples of the illogic of null hypothesis significance testing can be misleading (Cortina and Dunlap 1997).

But, tellingly, the debate often boils down to a combination of the last two concerns. Those against significance testing argue that using an arbitrary cutoff to evaluate a null hypothesis inaccurately represents the data and falsely dichotomizes decision making. Instead, we should use confidence intervals to represent the lack of certainty in our belief about our data, power analysis to assess the probability of a Type II error (failing to reject the null hypothesis when the null hypothesis is false), and exploratory data analysis (Tukey 1977) and effect sizes (combined with meta-analysis) to represent the magnitude of effects. Those in favor of significance testing respond that making policy and determining courses of treatment require binary decisions, that an alpha level can be agreed upon for making such a decision, and that the conservative stance of an unknown relationship being nil accurately represents resistance to the implementation of a new program or treatment (Abelson 1997; Chow 1988; Cortina and Dunlap 1997; Frick 1999; Harris 1997; Wainer 1999).

The key point here is that the ITCV represents a middle ground. Based on the significance test, the framework of the ITCV applies to binary decisions. But like the confidence interval, the ITCV contextualizes a given p value. Like the effect size, the numerical value of the ITCV indicates an aspect of the strength of the relationship between x and y; the stronger the relationship between x and y, the greater must be the impact of a confounding variable to alter the inference with regard to x and y. But, linked to the t statistic, the ITCV also includes information about the uncertainty of an estimate. In fact, use of the ITCV is very much in the spirit of the recent guidelines for statistical methods in psychology journals, whose authors, including many of the most prominent statisticians in the social sciences, declined to call for a ban on statistical tests (Wilkinson and Task Force on Statistical Inference 1999:602). Instead, the task force recommended that results of statistical tests be reported in their full context. And sensitivity to the impact of a confound is part of that context. Perhaps it is best to consider the ITCV like other statistical tools that should be used based on the consideration of the researcher, referee, and editor (Grayson 1998).

Of course, the logic of the ITCV can be applied beyond the framework of null hypothesis significance testing. First, one can specify the impact necessary to reduce a standardized regression coefficient below a given threshold, v, as $k(v) \geq (\beta^*_{y \cdot x} - v) / (1 - v)$. For example, one could ask, "What is the impact necessary to reduce $\beta^*_{y \cdot x}$ to be no longer of substantive interest (defined by v) rather than statistical significance (as in the ITCV)?" Second, one can specify the impact necessary to change a standardized regression coefficient s units or more as $k(s) \geq s / (s + 1 - \beta^*_{y \cdot x})$.[19] The indices $k(s)$ and $k(v)$ are not defined by statistical inference, and thus can help social scientists discuss the conditions of causality in a population. But two critical observations in this article still apply: Characterizing impact in terms of $r_{y \cdot cv} \times r_{x \cdot cv}$ neatly quantifies the terms of discussion and the values of each index can be compared to a reference distribution based on impacts of existing covariates. Ultimately, however, the use of each index is tied to a binary decision, and thus may be controversial.

Most generally, I presented the ITCV in the context of statistical inference because contemporary theories of causation provide a sound basis for social scientists to use statistical inference. For all intents and purposes, unmeasurable differences among people force social scientists to accept probable causes and statistical relationships just as theoretical uncertainty in physical measurement forces physical scientists to accept the same (e.g., Suppes 1970). Philosophers of science have turned to probabilistic and statistical relations as essential and irreplaceable aspects of causality (Salmon 1998). Nonetheless, for social science to progress we must recognize that a statistical inference is not a final proclamation (e.g., Hunter 1997; Rozenboom 1960; Sohn 1998). This caution is consistent with Fisher's qualified interpretation of the p value (see Gigerenzer 1993:329). Therefore, instead of abandoning statistical inference we should expand upon it to recognize the limitations of the inference and the robustness of the inference with respect to alternative explanations. It is in this vein that I developed the ITCV.

CONCERNS WITH REGARD TO CAUSAL INFERENCE

There are those who are philosophically uncomfortable with causal inference[20] in the social sciences, especially from quantitative analyses of observational data. Such cautions have emerged particularly in response to claims associated with causal modeling (Abbott 1998; Freedman 1997; Pratt and Schlaifer 1988; Sobel 1996, 1998).[21] The cautions must be heeded. One cannot definitively infer causality from regression estimates in an observational study unless all confounds have been measured and controlled for. Consequently, some have emphasized the descriptive value of quantitative models from observational studies (e.g., Abbott 1998; Sobel 1998).

Even when these concerns are acknowledged, the ITCV has value. First, the ITCV quantifies the concern expressed by those who are philosophically uncomfortable with causal inference. After calculating the ITCV, we know how large the impact of a confounding variable must be to alter an inference. This informs our interpretations of findings by moving the discussion from a binary scale—either there is or is not a confounding variable—to the continuous scale of the ITCV. Use of the ITCV also is consistent with recent guidelines for psychological research (Wilkinson and Task Force on Statistical Inference 1999), which call for researchers to alert the reader to plausible rival hypotheses that might explain the results.

Second, the ITCV can be applied to any experimental design. Use of the ITCV can strengthen interpretations from quasi-experimental designs that control for many of the most important covariates; when one quantifies the impact of a confounding variable, one must acknowledge that the impact may be partially absorbed by existing covariates. In Featherman and Hauser's (1976) example, the existing covariates of farm origin and number of siblings absorbed some of the impact of the proposed confound, father's education; the coefficient associated with father's occupation was already partially adjusted for father's education. The ITCV also can be applied to statistical analyses of experimental data in which there is concern that important differences (not attributable to treatment) emerged between treatment and control groups.

Most generally, the ITCV can contribute to the accumulation of knowledge gleaned from any data design by helping researchers evaluate the robustness of an inference.

Third, those designs that offer the highest internal validity, such as experimental, may also be the most expensive and difficult to implement (in fact, there are many instances in which experiments are impractical in the social sciences). Thus, evidence from the weaker designs can be used as indicators for pursuing confirming evidence from more rigorous designs. Toward this end, use of the ITCV informs the research agenda.

In spite of the value of the ITCV, it is worth emphasizing that the ITCV does not definitively sustain new inferences when statistical control is relied upon to control for confounding variables. The ITCV does not replace the need for improved data collection designs, experiments (where possible), and theories. If one accepts the idea that causal inferences are to be debated, what the ITCV does do is quantify the terms of the debate. Therefore, instead of "abandoning the use of causal language" (Sobel 1998:345; see also Sobel 1996:355), we can use the ITCV to quantify the language, extending the deep history of causal inference in the social sciences (for references to Durkheim, see Abbott 1998; see also MacIver 1964).

CONCERNS WITH REGARD TO THE USE OF THE REFERENCE DISTRIBUTION

In Featherman and Hauser's (1976) example, the inference with regard to the probability of an impact of a covariate exceeding the ITCV was based on a small sample, suggesting caution. Furthermore, the covariate most representative of father's education had the largest impact of .18, relatively close to the threshold of .228. Perhaps it is best that the example of the reference distribution did not yield an unequivocal interpretation. It highlights the fact that in making any inference, researchers must determine the population represented by the sample. When the focus of the model is on explaining variance in the outcome, the reference distribution based on the covariates identified primarily for their correlation with the outcome may well underestimate the impact of a potential confound for a specific predictor. On the other hand, when a model is developed with a focus on a single predictor of interest, a reference distribution based on the most important covariates thought to be correlated with both the outcome *and* the predictor of interest may well overestimate the potential impact of any secondary confound.

Two other caveats with regard to the use of the reference distribution are critical. First, although in Sobel's (1998) example the ITCV was unlikely to be exceeded (and was not exceeded for observed correlations for father's education), it does not diminish concerns he and others have raised with regard to causal inference. The impact of father's education was close to the ITCV for father's occupation, and might well have exceeded the ITCV for number of siblings or farm origin, which were not as strongly correlated with educational attainment as father's occupation was. In fact, part of the value of the reference distribution is that it can be used to identify when a potentially confounding variable is likely to alter causal inferences.

Second, it may be tempting to argue that the cutoff value for assessing the alternatives with regard to the ITCV should be larger than the standard .05 because the ITCV is used to adjudicate between two equally held positions rather than between a "commonly accepted" null and an alternative. But a larger cutoff value might support erroneous causal inferences. Instead, I would advocate using a stringent criteria (perhaps even smaller than .05) that, if satisfied, would definitively rebut the skeptic.

EXTENSIONS

The concern with regard to the representativeness of existing covariates suggests it would be valuable to characterize the reference distribution of impacts across the social sciences based on different types of covariates. What is the reference distribution if pretests are included? What is it if only a core of covariates common to most studies are included?[22] What does it look like if all covariates are included? Perhaps a prediction could be obtained for k^* given characteristics of the covariates. This would support statements such as the following: "The ITCV for predictor x is on the order of a typical pretest impact." This type of reasoning may also have applications in the legal realm, where potential confounds are often proposed by an adversary and where there may be little basis for a specific reference distribution (see Dempster 1988:152).

Of course, if multiple effects are estimated in a single model, then each effect is adjusted for the others. In this case, a matrix could be used to represent the impacts of each independent variable on the others. That is, element $(x1, x2)$ would represent the impact of $x1$ on the inference of the coefficient for $x2$. Two such values are calculated in the last column of Table 2 for the impacts of farm origin and number of siblings on father's occupation. A scan of a full matrix could help one appreciate which independent variables were critical controls in interpreting estimated coefficients for other variables. But considering multiple independent variables as effects conflicts with the focus on one effect with many covariates in a model and, therefore, should be undertaken with caution.

Of course, the observed correlations on which the statistical control is based can be attenuated by measurement error. The concern here is that the observed relationship between x and y is underadjusted for the confounding variable due to the unreliability of the measure of the confounding variable. In this case we might ask, "How small must be the reliability of the confounding variable such that the observed relationship between x and y can be attributed to the unreliability of the measure of the confounding variable?" To answer this question, define κ as the impact of the confounding variable in the population if the confounding variable were measured with 100 percent reliability. Define $\alpha(cv)$ as the reliability of the confounding variable. The concern is that ITCV = κ, whereas $k <$ ITCV because $\alpha(cv) < 1$. Now, express κ in terms of the observed correlations and the reliability of the confounding variable, and then solve for the reliability. This gives the following relation for the bivariate case:

$$ITCV = \kappa = \rho_{x \cdot cv}\rho_{y \cdot cv} = \frac{r_{x \cdot cv}}{\sqrt{\alpha(cv)}} \frac{r_{y \cdot cv}}{\sqrt{\alpha(cv)}} \qquad (20)$$

$$\Rightarrow \alpha(cv) = \frac{r_{x \cdot cv} r_{y \cdot cv}}{ITCV} = \frac{k}{ITCV},$$

where the ITCV is calculated without the observed covariate in the model. Thus, the relationship between x and y can be attributed to the unreliability of the measure of cv in proportion to k/ITCV. Thus, large values of $\alpha(cv)$ indicate that the existing impact is quite close to the observed ITCV, and therefore only a small degree of unreliability in the confounding variable can cause an incorrect inference with regard to x. The expression in the multivariate case, assuming covariates ($\mathbf{z}$) are in the model, is

$$\alpha(cv) = \frac{(r_{y \cdot cv} - r_{y \cdot \mathbf{z}} r_{cv \cdot \mathbf{z}})(r_{x \cdot cv} - r_{x \cdot \mathbf{z}} r_{cv \cdot \mathbf{z}})}{ITCV \sqrt{(1 - r_{x \cdot \mathbf{z}}^2)(1 - r_{y \cdot \mathbf{z}}^2)}} + r_{cv \cdot \mathbf{z}}^2 . \tag{21}$$

Note that if there are no covariates, (21) reduces to (20).

Applications of the IT can be extended by drawing on the algebraic relationships of the general linear model. For example, one could present the ITCV in terms of the necessary impact of a confounding variable to position the border of a confidence interval at zero. Or one could assess the necessary impact of a *mediating* variable to reduce the direct effect of a predictor of interest to zero. In exploring such extensions, it is important to note that I developed the ITCV with regard to a skeptic who proposes a potentially confounding variable, with implications for my calculations. For example, I minimized the expression defining the index and calculated the index assuming the impact of the confound had not been absorbed by existing covariates in the multivariate case ($r_{cv \cdot \mathbf{z}} = 0$). But to the extent that the approach here is extended, other decisions might be made.

The ideas in this article also can be extended beyond the general linear model. For example, one could obtain the ITCV for multilevel models or logistic regressions. Indeed, I did not pursue the example of the effect of Catholic schools because of the absence of such extensions. Furthermore, there is an urgent need to consider the development of the ITCV for causal models cast in the framework of causal inferences (Holland 1988; Sobel 1998) and instrumental variables that have recently been reconsidered with respect to causal models (Angrist et al. 1996). Such developments will no doubt demand more complex representations than the simple scalar formulas developed here, but it is worthwhile to extend the concept to as broad an array of models as possible.

NOTES

1. See Sobel (1998) for a representation of the general linear model in terms of the expectation of y conditional on x.

2. Technically, the expression cv to represent a confounding variable is redundant, since it refers to a random variable. But I use the expression because the term *confounding variable* has nearly first-order meaning for social scientists.

3. The following presentation is developed for $\hat{\beta}_1$ instead of a partial correlation because $\hat{\beta}_1$ typically is interpreted by the social scientist.

4. The following expression for k has been modified from the standard quadratic because the 2's cancel and the negative sign typically associated with b is included in the expression of b.

5. The term k is presented as an index relative to an estimated regression coefficient. To the extent that k is considered an estimate of a population parameter, it would have all of the properties of an estimated correlation (e.g., bias, consistency) because it is a linear function of the estimated correlation. I have chosen not to simplify these expressions for the impact threshold (IT) so that I may represent the intercept and slope linking $r_{y \cdot x}$ and the IT.

6. I added one to the sample sizes represented in Cohen and Cohen (1983, table G2) to account for the extra parameter, β_2, associated with cv that would be estimated in model (1b).

7. Mauro (1990) described a potential index in terms of the volume within a cube defined by correlations associated with the confounding variable, but this measure was not fully developed or implemented.

8. I believe it is new to characterize the existing covariates as "absorbing" the impact of the confounding variable. The terms *impact* and *absorb* form a consistent physical metaphor and help us to think about the role of existing covariates as statistical controls.

9. It is possible that $r_{cv \cdot \mathbf{z}} < 0$ while the impact of cv is positive. Here, the skeptic would be arguing that the impact of the confounding variable is suppressed (in the traditional sense) by the existing covariates. That is, the skeptic argues "now that you have controlled for **z**, it is more important to control for cv." This line of argument is so unusual that I eliminate it by assumption to ease calculation.

10. I have chosen not to use the Catholic schools issue as an example because of complications with the multilevel nature of the data. The example developed here based on the effect of socioeconomic status on educational attainment underlies the debate with regard to the effect of schooling, controlling for socioeconomic status.

11. Note the effects reported here are estimated from the correlation matrix presented in Table 1 of Featherman and Hauser (1976). They are slightly different from regression results reported in Featherman and Hauser's Table 3, presumably due to differences in the use of pairwise and listwise deletion.

12. Sobel (1998) had other concerns with regard to the interpretation of the regression coefficients for the background variables as "effects." Namely, that a mediating variable, marital status, was essentially included in Featherman and Hauser's (1976) analyses. But Sobel's primary concern, as reflected in his theoretical development (pp. 322-30), and similar to Pratt and Schlaifer's (1988) argument, focused on the potential impact of a confounding variable.

13. I use the term *reference distribution* instead of *sampling distribution* to indicate that the observed estimates are based on the impact of covariates other than the potentially confounding variable. Use of the covariates in the reference distribution suggests that the product $r_{y \cdot cv|\mathbf{z}} \times r_{x \cdot cv|\mathbf{z}}$ is a natural indicator of the impact of an existing covariate on the inference of a regression coefficient. Furthermore, the product $r_{y \cdot cv|\mathbf{z}} \times r_{x \cdot cv|\mathbf{z}}$ can be interpreted relative to the ITCV. The impacts of farm origin and number of siblings are 26 percent and 44 percent of the ITCV for father's occupation, respectively, suggesting that these are indeed important covariates to have included in the model, although comparable covariates, in and of themselves, would not alter the inference with regard to father's occupation.

14. This departs considerably from a standard hypothesis testing framework. Here, I use the marginal distribution of the observed impacts to make an inference about *observing* an impact of a given value. Thus the decision making domain is defined by the skeptic's challenge to a statistical inference from sample data. Using a frequentist approach in a more traditional hypothesis testing framework, one could obtain the one-tailed probability of observing a given distribution of impacts under the null hypothesis that the population value of impact is greater than or equal to the ITCV. If the probability is small, one would reject the skeptic's null hypothesis. But this decision rule is sensitive to sample size, and thus one could reject the skeptic's hypothesis even if a significant percentage of the observed impacts is greater than the ITCV. Therefore such a decision rule is unlikely to persuade a skeptic.

Alternatively, using a Bayesian approach, one could combine the conditional distribution of the observed impacts (under the null and alternative hypotheses) with a prior distribution of impact to obtain the posterior distribution of impact. From this one could define a decision rule in terms of the probability that a given impact exceeds the ITCV. But specification of the prior distribution for impact and the conditional and posterior distributions in terms of (δ_1, δ_2, and ρ) is beyond the scope of this article.

15. When there are many estimates of k, the normality of each set of transformed coefficients can be checked through a q-q plot, but it is not informative to do so here because of the small sample size.

16. Special cases of k^* are discussed by Aroian, Taneja, and Cornwell (1978). For example, if $\rho = 1$ and $\delta_1 = \delta_2 = \delta$, then the distribution is χ^2 with noncentrality parameter $\lambda = \delta^2/2$.

17. In fact, the direct effect of father's occupation is no longer statistically significant only after controlling for college plans and occupational status aspiration, each of which clearly mediates the effect of father's occupation and, therefore, should not be used to alter the causal inference with regard to father's occupation.

18. Perhaps it is not surprising that recent concerns about p values have been voiced more strongly by psychologists than by sociologists. Sociologists have always had to be more circumspect about interpreting p values because of the limited possibilities for experimentation in sociology.

19. In applying either $k(s)$ or $k(v)$, it is important to note that the formulas are based on assumptions similar to those used for the ITCV, in particular that the confounding variable is not correlated with covariates already in the model and that $r^2_{x \cdot cv} = r^2_{y \cdot cv} = k$.

20. Of course, the concern with regard to causal inference that there are no confounding variables is critical to statistical inference, but I address the issues separately because of the theme of the article.

21. See Abbott (1998:162) for a discussion of the link between statistical models and causal inference that emerged with causal modeling.

22. See Cochran (1965) for a similar categorization of types of covariates in the context of creating matched comparisons.

REFERENCES

Abbott, Andrew. 1998. "The Causal Devolution." *Sociological Methods & Research* 27:148-81.

Abelson, Robert P. 1997. "On the Surprising Longevity of Flogged Horses." *Psychological Science* 8 (1): 12-15.

Anderson, Sharon, A. Auquier, W. W. Hauck, D. Cakes, W. Vandaele, H. I. Weisberg, A. S. Bryk, and J. Kleinman. 1980. *Statistical Methods for Comparative Studies*. New York: Wiley.

Angrist, Joshua D., Guide W. Imbens, and Donald B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables [With Discussion]." *Journal of the American Statistical Association* 91:444-72.

Aroian, L. A., V. S. Taneja, and L. W. Cornwell. 1978. "Mathematical Forms of the Distribution of the Product of Two Normal Variables." *Communications in Statistics: Part A—Theory and Methods* 7:165-72.

Asher, H. 1983. *Causal Modeling*. Beverly Hills, CA: Sage.

Bakan, D. 1966. "The Test of Significance in Psychological Research." *Psychological Bulletin* 66:423-37.

Blalock, H. M., Jr. 1971. *Causal Models in the Social Sciences*. Chicago: Aldine.

Blau, P. M. and O. D. Duncan. 1967. *The American Occupational Structure*. New York: Wiley.

Bourdieu, P. 1986. "The Forms of Capital." Pp. 241-58 in *Handbook of Theory and Research for the Sociology of Education*, edited by J. G. Richardson. New York: Greenwood.

Bowles, S. and H. Gintis. 1976. *Schooling in Capitalist America*. New York: Basic Books.

Bross, Irwin D. J. 1966. "Spurious Effects From an Extraneous Variable." *Journal of Chronic Disease* 19:637-47.

———. 1967. "Pertinency of an Extraneous Variable." *Journal of Chronic Disease* 20:487-95.

Bryk, Anthony S., Valerie E. Lee, and Peter B. Holland. 1993. *Catholic Schools and the Common Good*. Cambridge, MA: Harvard University Press.

Carver, Ronald. 1978. "The Case Against Statistical Significance Testing." *Harvard Educational Review* 48:378-99.

Chatterjee, Samprit and Ali S. Hadi. 1988. *Sensitivity Analysis in Linear Regression*. New York: Wiley.

Chow, Siu L. 1988. "Significance Test or Effect Size." *Psychological Bulletin* 103 (1): 105-10.

Clogg, Clifford C., Eva Petkova, and Edward S. Shihadeh. 1992. "Statistical Methods for Analyzing Collapsibility in Regression Models." *Journal of Educational Statistics* 17 (1): 51-74.

Cochran, W. G. 1953. "Matching in Analytical Studies." *American Journal of Public Health* 43:684-91.

———. 1965. "The Planning of Observational Studies of Human Populations." *Journal of the Royal Statistical Society Series A*, 234-66.

———. 1970. "Performance of a Preliminary Test of Comparability in Observational Studies." Office of Naval Research Technical Report No. 29, Harvard University.

Cohen, Jacob. 1990. "Things I Have Learned (So Far)." *American Psychologist* 45:1304-12.

———. 1994. "The Earth Is Round ($p < .05$)." *American Psychologist* 49:997-1003.

Cohen, J. and P. Cohen. 1983. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum.

Cohen, M. R. and E. Nagel. 1934. *An Introduction to Logic and the Scientific Method*. New York: Harcourt Brace.

Cook, Thomas and Donald T. Campbell. 1979. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Boston: Houghton Mifflin.

Cordray, D. S. 1986. "Quasi-Experimental Analysis: A Mixture of Methods and Judgment." *New Directions for Program Evaluation* 31:9-27.

Cornfield, J., W. Haenszel, A. M. Lilienfeid, M. B. Shimkin, and E. L. Wynder. 1959. "Smoking and Lung Cancer: Recent Evidence and a Discussion of Some Questions." *Journal of the National Cancer Institute* 22:173-203.

Cortina, Jose M. and William P. Dunlap. 1997. "On the Logic and Purpose of Significance Testing." *Psychological Methods* 2:161-72.

Davis, W. A. 1988. "Probabilistic Theories of Causation." Pp. 133-60 in *Probability and Causality*, edited by James H. Fetter. Dordrecht: D. Reidel.

Dempster, Arthur P. 1988. "Employment Discrimination and Statistical Science." *Statistical Science* 3:149-95.

Dowe, Phil. 1992. "Wesley Salmon's Process Theory of Causality and the Conserved Quantity Theory." *Philosophy of Science* 59:195-216.

Draper, N. R. and H. Smith. 1981. *Applied Regression Analysis*. New York: Wiley.

Duncan, Otis D., David L. Featherman, and Beverly Duncan. 1972. *Socioeconomic Background and Achievement*. New York: Seminar Press.

Einhorn, Hillel J. and Robin M. Hogarth. 1986. "Judging Probable Cause." *Psychological Review* 99 (1): 3-19.

Featherman, David L. and Robert M. Hauser. 1976. "Sexual Inequalities and Socioeconomic Achievement in the U.S. 1962-1973." *American Sociological Review* 41:462-83.

Freedman, David. 1997. "From Association to Causation via Regression." *Advances in Applied Mathematics* 18:59-110.

Frick, Robert W. 1999. "Defending the Statistical Status Quo." *Theory and Psychology* 9:183-89.

Gage, N. L. 1978. *The Scientific Basis of the Art of Teaching*. New York: Teachers College Press.

Giere, R. 1981. "Causal Systems and Statistical Hypotheses." Pp. 251-70 in *Applications of Inductive Logic*, edited by L. J. Cohen and M. Hesse. Oxford, UK: Clarendon.

Gigerenzer, G. 1993. "The Superego, the Ego, and the Id in Statistical Reasoning." Pp. 311-39 in *A Handbook for the Data Analysis in the Behavioral Sciences: Methodological Issues*, edited by G. Keren and C. Lewis. Hillsdale, NJ: Lawrence Erlbaum.

Godfrey, L. G. 1988. *Misspecification Tests in Econometrics*. Cambridge, UK: Cambridge University Press.

Granger, C.W.J. 1969. "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods." *Econometrica* 37:424-38.

Grayson, D. A. 1998. "The Frequentist Facade and the Flight From Evidential Inference." *British Journal of Psychology* 89:325-45.

Hanushek, E. 1977. *Statistical Methods for Social Scientists*. New York: Academic Press.

Harlow, L. L., S. A. Mulaik, and J. H. Steiger. 1997. *What If There Were No Significance Tests*. Mahwah, NJ: Lawrence Erlbaum.

Harris, Richard J. 1997. "Significance Tests Have Their Place." *Psychological Science* 8 (1): 8-11.

Hausman, J. A. 1978. "Specification Tests in Econometrics." *Econometrica* 46:1251-71.

Heckman, J. and R. Robb. 1985. "Alternative Methods for Evaluating the Impact of Interventions." Pp. 156-245 in *Longitudinal Analysis of Labor Market Data*, edited by J. Heckman and B. Singer. New York: Cambridge University Press.

Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81:945-70.

———. 1988. "Causal Inference, Path Analysis, and Recursive Structural Equations Models (With Discussion)." Pp. 449-93 in *Sociological Methodology*, edited by C. C. Clogg. Washington, DC: American Sociological Association.

Hunter, John E. 1997. "Needed: A Ban on the Significance Test." *American Psychological Society* 8 (1): 3-7.

Jencks, C. 1985. "How Much Do High School Students Learn?" *Sociology of Education* 58:128-35.

Kitcher, P. 1989. "Explanatory Unification and the Causal Structure of the World." Pp. 410-505 in *Minnesota Studies in the Philosophy of Science*, vol. 13, *Scientific Explanation*, edited by P. Kitcher and W. Salmon. Minneapolis: University of Minnesota Press.

Lareau, A. 1989. *Home Advantage: Social Class and Parental Intervention in Elementary Education*. New York: Falmer.

Lewis, S. 1973. "Causation." *Journal of Philosophy* 70:553-67.

Lin, D. Y., B. M. Psaty, and R. A. Kronmal. 1998. "Assessing the Sensitivity of Regression Results to Unmeasured Confounders in Observational Studies." *Biometrics* 54:948-63.

MacIver, R. M. 1964. *Social Causation*. New York: Harper Torchbooks.

Mackie, J. 1974. *The Cement of the Universe*. Oxford, UK: Oxford University Press.

Mauro, Robert. 1990. "Understanding L.O.V.E. (Left Out Variables Error): A Method of Estimating the Effects of Omitted Variables." *Psychological Bulletin* 108:314-29.

McDonald, R. P. 1997. "Haldane's Lungs: A Case Study in Path Analysis." *Multivariate Behavioral Research* 32:1-38.

McKinlay, Sonja M. 1975. "The Design and Analysis of the Observational Study—A Review." *Journal of the American Statistical Association* 70:859-64.

Meehl, Paul E. 1978. "Theoretical Risks and Tabular Asterisks: Sir Karl, Sir Ronald and the Slow Progress of Soft Psychology." *Journal of Consulting and Clinical Psychology* 46:806-34.

Meeker, W. Q., L. W. Cornwell, and L. A. Aroian. 1980. "The Product of Two Normally Distributed Random Variables." Pp. 1-256 in *Selected Tables in Mathematical Statistics*, edited by W. J. Kenney and R. E. Odeh. Providence, RI: American Mathematical Society.

Meeker, William Q. and Luis A. Escobar. 1994. "An Algorithm to Compute the CDF of the Product of Two Normal Random Variables." *Communications in Statistics—Simulations* 23:271-80.

Mill, J. S. [1843] 1973. "A System of Logic: Ratiocinative and Inductive." In *The Collected Works of John Stuart Mill*, vols. 7, 8, edited by J. M. Robson. Toronto: University of Toronto Press.

Montgomery, D. C. and E. A. Peck. 1982. *Introduction to Linear Regression Analysis*. New York: Wiley.

Morrison, D. E. and R. E. Henkel. 1970. *The Significance Test Controversy*. Chicago: Aldine.

Oakes, M. 1986. *Statistical Inference: A Commentary for the Social and Behavioral Sciences*. New York: Wiley.

Pedhazur, E. 1982. *Multiple Regression in Behavioral Research: Explanation and Prediction*. 2d ed. San Francisco: Holt, Rinehart & Winston.

Pollard, Paul. 1993. "How Significant Is 'Significance?' " Pp. 449-60 in *A Handbook for the Data Analysis in the Behavioral Sciences, Methodological Issues*, edited by G. Keren and C. Lewis. Hillsdale, NJ: Lawrence Erlbaum.

Pratt, John W. and Robert Schlaifer. 1988. "On the Interpretation and Observation of Laws." *Journal of Econometrics* 39:23-52.

Ramsey, J. 1974. "Classical Model Selection Through Specification Error Tests." Pp. 13-47 in *Frontiers in Econometrics*, edited by P. Zarembka. New York: Academic Press.

Reichenbach, H. 1956. *The Direction of Time*. Berkeley: University of California Press.
Rice, J. A. 1988. *Mathematical Statistics and Data Analysis*. Pacific Grove, CA: Wadsworth & Brooks.
Rosenbaum, Paul R. 1986. "Dropping Out of High School in the United States: An Observational Study." *Journal of Educational Statistics* 11:207-24.
Rosenbaum, Paul R. and Donald B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70 (1): 41-55.
Rozenboom, William W. 1960. "The Fallacy of the Null Hypothesis Significance Test." *Psychological Bulletin* 57:416-28.
Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66:688-701.
Ryan, T. P. 1997. *Modern Regression Methods*. New York: Wiley.
Salmon, Wesley. 1984. *Scientific Explanation and the Causal Structure of the World*. Princeton, NJ: Princeton University Press.
———. 1994. "Causality Without Counterfactuals." *Philosophy of Science* 61:297-312.
———. 1998. *Causality and Explanation*. New York: Oxford University Press.
Scheines, R., P. Spirites, C. Glymour, C. Meek, and T. Richardson. 1998. "The TETRAD Project: Constraint Based Aids to Causal Model Specification." *Multivariate Behavioral Research* 33:65-117.
Schlesselman, J. J. 1978. "Assessing Effects of Confounding Variables." *American Journal of Epidemiology* 108:3-8.
Schmidt, Frank L. 1996. "Statistical Significance Testing and Cumulative Knowledge in Psychology: Implications for Training of Researchers." *Psychological Methods* 1:115-29.
Sewell, W. H., R. M. Hauser, and W. C. Wolf. 1980. "Sex, Schooling, and Occupational Status." *American Journal of Sociology* 86:551-83.
Snedecor, G. W. and W. G. Cochran. 1989. *Statistical Methods*. 6th ed. Ames: Iowa State University Press.
Sobel, Michael E. 1995. "Causal Inference in the Social and Behavioral Sciences." Pp. 1-38 in *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, edited by Gerhard Arminger, Clifford C. Clogg, and Michael E. Sobel. New York: Plenum.
———. 1996. "An Introduction to Causal Inference." *Sociological Methods & Research* 24:353-79.
———. 1998. "Causal Inference in Statistical Models of the Process of Socioeconomic Achievement: A Case Study." *Sociological Methods & Research* 27:318-48.
Sober, E. 1988. "The Principle of the Common Cause." Pp. 211-28 in *Probability and Causality*, edited by James H. Fetter. Dordrecht: D. Reidel.
Sohn, David. 1998. "Statistical Significance and Replicability." *Theory & Psychology* 8:291-311.
Steyer, Rolf and Thomas Schmitt. 1994. "The Theory of Confounding and Its Application in Causal Modeling With Latent Variables." Pp. 36-67 in *Latent Variables Analysis: Applications for Developmental Research*, edited by Alex von Eye and Clifford Clogg. Thousand Oaks, CA: Sage.
Suppes, P. 1970. *A Probabilistic Theory of Causality*. Amsterdam: North-Holland.
Thompson, Bruce. 1999. "If Statistical Significance Tests Are Broken/Misused, What Practices Should Supplement or Replace Them?" *Theory & Psychology* 9:165-81.
Tukey, J. W. 1977. *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.
———. 1991. "The Philosophy of Multiple Comparisons." *Statistical Science* 6:100-16.
Wainer, Howard. 1999. "One Cheer for Null Hypothesis Significance Testing." *Psychological Methods* 4:212-13.

Weisberg, Sanford. 1985. *Applied Linear Regression*. New York: Wiley.
Wilkinson, L. and Task Force on Statistical Inference. 1999. "Statistical Methods in Psychology Journals: Guidelines and Explanations." *American Psychologist* 54:594-604.
Zellner, A. 1984. *Basic Issues in Econometrics*. Chicago: University of Chicago Press.

Kenneth A. Frank is an associate professor of measurement and quantitative methods within counseling, educational psychology, and special education at Michigan State University. He received a Ph.D. from the University of Chicago. His substantive interests are in the sociology of organizations and the sociology of education. These combine with his methodological interests in multilevel models and social network analysis.