Jae-On Kim and G. Donald Ferree. "Standardization in Causal Analysis." 1981. Sociological Methods and Research Vol. 10, No. 2, November 1981, pp. 187-210. Sage Publications Inc.

In comparative study, it is argued that (1) the standardization of variables and scales should be separated from the habitual use of standardized coefficients; (2) the use of standardized coefficients implies standardizing every variable using group specific standards, and, therefore, it is not appropriate even if some variables have group specific metrics or some variables do not possess commonly accepted metrics; and (3) the explicit standardization of some or all variables can be fruitfully combined with the use of unstandardized coefficients.

Standardization in Causal Analysis

JAE-ON KIM
University of Iowa
G. DONALD FERREE, Jr.
University of Connecticut

INTRODUCTION: THE CONTROVERSY

The long-standing debate in the social sciences over the desirability of using standardized coefficients in regression and path analysis seems, at first glance, almost a matter of style. Does one prefer to interpret causal coefficients as meaning, ceteris paribus, how many units the dependent variable will change if an independent variable is changed one unit? Or does one prefer a "scale-free" interpretation in which each variable is rescaled to have mean 0 and standard deviation 1.0, so that units, instead of being dollars (or one point on a 1-to-7 scale), are standard deviations? This apparent dispute of taste can also be seen in the light of the "sociology of social science": Psychologists, dealing with arbitrarily scaled attitude items, tend to standardize; economists, dealing with dollars, tend not to. Sociologists and

AUTHORS' NOTE: *We are grateful to Kathy Harmon, Gayle Scriven, Charles W. Mueller, and James Curry, who shared their thoughts with us on the subject, to anonymous reviewers for their valuable comments, and to Sidney Verba for providing a stimulating intellectual environment.*

© 1981 Sage Publications, Inc.

political scientists sometimes do the one, sometimes the other (Land, 1969; Blalock, 1968; Duncan, 1975; Hotchkiss, 1976; c.f. "path analysis," Duncan, 1966, with "dependence analysis," Boudon, 1965, 1968).

The controversy has been hard fought for many years. Tukey, for instance, had little use for standardized coefficients, maintaining that they should be used in only two circumstances: where dependent and independent variables had the same variance—and thus the coefficients are equal—or where measurement of at least one variable on a determinate scale was "hopeless" (1954: 39). Blalock (1964, 1967a) claimed that the unstandardized coefficient reflected the actual causal law and hence was to be preferred to a standardized coefficient, which not only obscured this law but also had a complicated sampling distribution. Blalock's point was reiterated by Cain and Watts (1970) and by Duncan (1975). Kim and Mueller (1976: 430-431), concurring, also specified four components that affect the sample standardized coefficient, b: (1) the estimate of the unstandardized coefficient; (2) the variance of x; (3) the variance of other variables, whether explicitly included in the equation or assumed implicitly, that affect y; and (4) covariances among the independent variables.

On the other hand, standardized coefficients have been championed on several grounds. First, because standardization attempts to put all the variables on the "same" scale, it is easier to compare the effects of different independent variables when standardized coefficients are used than when unstandardized coefficients are used (Land, 1969; Wright, 1960, 1934; Duncan, 1970; Hargens, 1976; Hotchkiss, 1976). Second, the standardized coefficients sometimes are easier to interpret than the unstandardized coefficient. For example, in a simple regression, the standardized coefficient is equivalent to Pearson's correlation coefficient, and the square of this coefficient can be interpreted as "proportion of variance explained." Third, the true (or theoretically relevant) metric may be relative to the distribution of a given subgroup or population and hence standardized coefficients reflect true structural parameters while unstandardized coefficients may not (Hargens, 1976). For instance, a comparatively well-off Indian might more appropriately be compared with a comparatively well-off American than with an American who had an equivalent dollar income but who, in U.S. terms, would be not at all well off. Fourth, many variables (such as indices of opinion) lack a clear metric basis and achieving comparable scale for these variables across populations or studies is almost impossible. Furthermore, it is rather easy to overestimate, particularly in comparative research, the degree to which apparently similar metrics are truly comparable (Felson, 1974; Przeworski and Teune, 1970). Under such circumstances, the use of standardized coefficients may introduce some degree of comparability which may not be obtainable otherwise (Forbes and Tufte, 1968; Stokes, 1974).[1]

These considerations indicate that the issues are more complex than apparent and they deserve a careful and systematic examination. We attempt to clarify the problem by (1) making a clear distinction between two logically different operations—the standardization of scales and the use of standardized coefficients—and then (2) relating these two distinctive operations to the following three important issues in causal analysis: how to define an invariance in causal relationships; how to obtain equivalent scales of measurement; and how to compare the causal effect of one variable with the causal effect of another variable.

STANDARDIZATION AND STANDARDIZED COEFFICIENTS: A CONCEPTUAL DISTINCTION

There are many ways to *standardize* scales of measurement. The most widely used and best known method of standardization is to transform (or rescale) each variable in such a way that the resulting mean is zero and the resulting standard deviation is one. The linear transformation used for this purpose is:

$$Z = (X - \bar{X})/S_x \quad [1]$$

$$= a + bX_i \quad [2]$$

where Z, X, $\bar{X}$, and S_x stand for the standardized variable, the original variable, mean of X, and standard deviation of X, respectively. Equation 2 puts the transformation into a more familiar form by replacing $-\bar{X}/S_x$ and $1/s_x$ with constants a and b, respectively. In fact, no information is lost in the transformation; the original metric can always be recovered given the mean and the standard deviation of the original variable.

This type of transformation is especially helpful when the variables under consideration do not have readily interpretable (or generally accepted) metrics or when variables have widely different scales of measurement. In these situations, we gain a certain degree of convenience and comparability across the variables by standardizing each variable according to common characteristics—means and standard deviations of the variable in the sample. Therefore, when one is dealing with a single sample (or population), the decision to use standardized variables or not to use them should be based on the relative convenience in interpretation provided by the use of standardization.

It is quite possible that for certain research issues and combinations of variables, it is better to use unstandardized variables (or the original metric), and for certain research issues, it is better to use standardized variables. Or, in certain type of research situations, it may be best to standardize some variables and to use original metric for other variables (see Felson, 1974). But the main issue is only a matter of convenience.

The use of standardized (regression or path) coefficients in a single sample is equivalent to (1) standardizing all the variables under consideration and then (2) using unstandardized coefficients for the transformed variables. Therefore, the decision as to whether to use standardized coefficients can be made on almost the same grounds as the issue of whether to standardize variables—except for one difference—the decision to use standardized coefficients implicitly binds one to the decision to standardize *all* of the variables. Still, the difference between the two strategies is small.

But the situation changes drastically if one begins to consider more than one sample or group. To make the argument concrete,

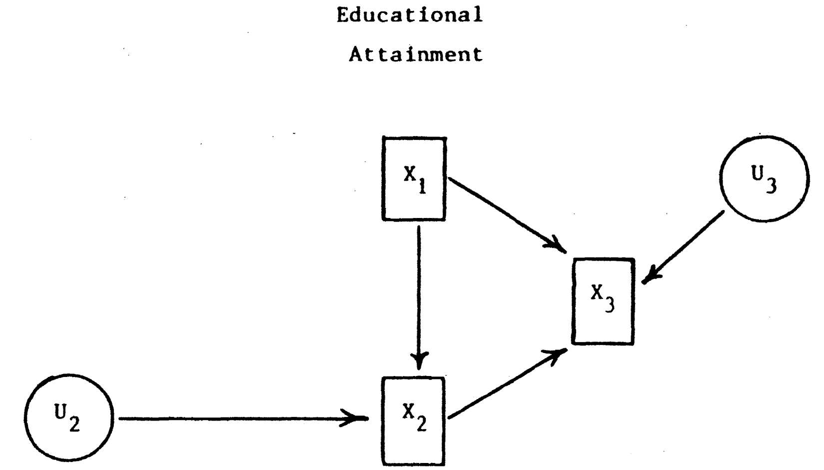

Figure 1: A Simplified Model of the Process of Stratification

suppose that one is dealing with the issue of whether the same status-attainment processes are operating for blacks and for whites. By illustration, assume that we have in mind the three-variable causal system shown in Figure 1. Even if we decide to use the usual standardization, as in equation 1, there are two options open to the researcher: (1) to use the means and standard deviations of the whole combined group or (2) to use group-specific means and standard deviations.

If one standardizes the variables on the basis of the means and standard deviations of the combined group, the resulting transformed variables in one group are comparable to the transformed variables in the other group because the same linear transformations are applied to both groups. If one standardizes variables according to the means and deviations of each (sub)group, one is using different transformations for the different groups, as long as means and standard deviations differ from one group to another.

The use of standardized coefficients—such as estimating standardized path coefficients for blacks and for whites separately—is equivalent to standardizing *all* the variables according to the group-specific means and standard deviations. What is implicit in the process of using standardized coefficients is that one standard deviation of the variable X_1 for one group is equivalent to one standard deviation of the same variable for the other group. An example may clarify these points: Say the standard deviation of income for a sample of whites is $5000 and that for a sample of blacks it is $2000. In using standardized coefficients, one would assume that $5000 for the whites was equivalent to $2000 for the blacks. In fact, one would be using different scales for measuring income, as if whites were paid in pounds sterling while blacks were paid in American dollars.

Suppose further that one wishes to evaluate the effects of formal education on earnings, and that there is little difference in the standard deviation of educational attainment between the samples of blacks and of whites. Which would one regard as indicative of equivalence of causal processes between the races: (1) six years of formal education being converted to $5000 for whites, but only to $2000 for blacks (which would, under the assumptions, result in equal standardized coefficients for whites and for blacks) or (2) six years of formal education being converted into $5000 for both blacks and whites (which would result in identical unstandardized coefficients but substantially disparate standardized coefficients)?

We think that standardization of variables in order to obtain theoretically meaningful or easily interpretable metrics has its place in comparative analysis, but the use of standardized coefficients as a means of achieving standardization carries with it too much unwanted baggage. What we propose is the combined use of standardization of variables and unstandardized coefficients. Once we separate the operation of standardization of variables from obtaining theoretically relevant coefficients, we also open up the possibilities of employing (1) other methods of standardization and data transformation than the one discussed above and (2) standardization of only selected variables.

Some of the obvious methods of standardizing variables to achieve comparability of measurement across the variables without sacrificing equivalence across the groups are: (1) using the means and variances of the whole population (or combined groups) or (2) using the means and variances of a reference group. The first method would be particularly appropriate if the subgroups under consideration constitute together a well-defined population and if the two sample sizes are proportional to the population distribution. The second method can take several forms: (a) the variables may be standardized on the basis of known population means and standard deviations (for instance, derived from census data); (b) the variables may be standardized on the means and standard deviations of one of the groups under consideration (for example, using the mean and standard deviations for whites); (c) the variables may be standardized on the basis of means and standard deviations at a particular time (for example, using a particular year as a reference, or using the values at the equilibrium state, etc.); or (d) the variables in a given study may be standardized using mean and standard deviations reported for another study.

Standardization of any kind may sometimes hamper rather than facilitate interpretation, as, for example, in the case of dichotomous variables, such as sex, for which there is no intrinsic metric. The unstandardized coefficients associated with such binary or dummy variables (as scored 0 or 1) immediately tell us the expected differences between the two groups (other things being equal), while standardization of such variables complicates the interpretation. As Felson (1974) argued cogently, there is no reason to standardize every variable.

We will now illustrate one particular method of standardizing variables and combining the standardized variables with unstandardized coefficients.[2] In order to highlight the differences between using standardized coefficients and using our strategy, we will reanalyze the data originally reported by Duncan (1968). In Figure 2, part of Duncan's path model is reproduced, with unstandardized path coefficients for the standardized variables and standardized path coefficients in parentheses. Before obtaining the unstandardized coefficients, each variable was standardized on the basis of its mean and variance in the combined sample.

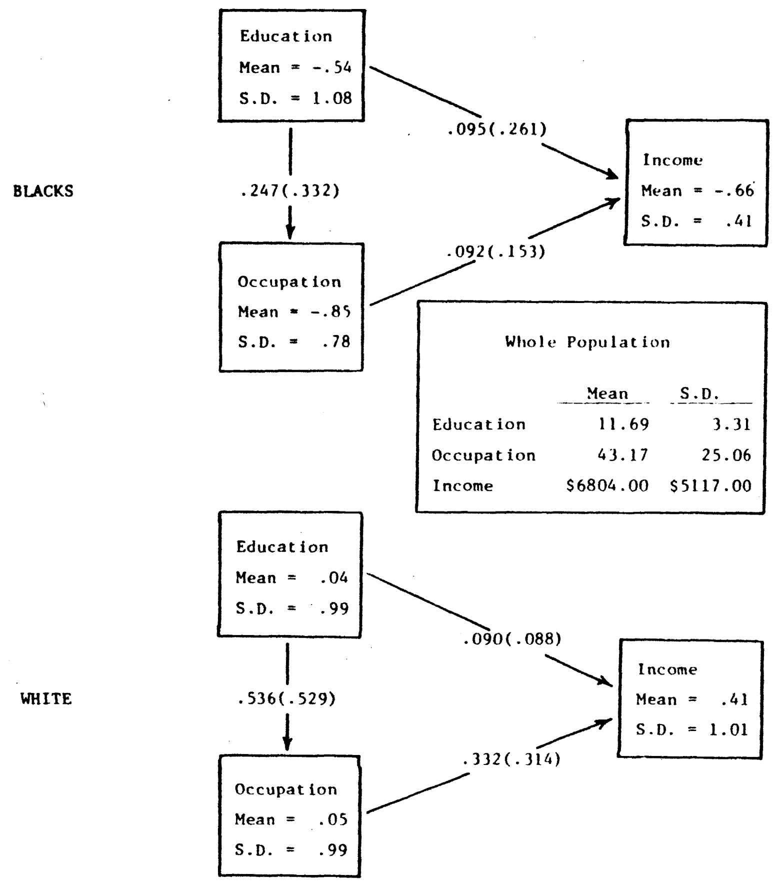

Figure 2: Simplified Path Diagrams: Unstandardized (and Standardized) Path Coefficients

A few introductory notes may be in order. Although only three variables are examined here, the coefficients are based on Duncan's full model, which also included father's education, father's occupation, and number of siblings. Thus, one should regard the path coefficients shown above as representing the effects, while holding the other three variables constant. Second, consider the means and standard deviations presented. Each variable has been transformed by subtracting the mean of the combined sample and by dividing the difference by the standard deviation of the combined sample. Thus, for example, we note that the mean income of blacks is .66 standard deviation unit below the national mean and that the respective standard deviation is less than one-half that of whites. (We think that this information on means and standard deviations should be given as a standard part of the path analysis.) Third, the analyses reported here are based on the reanalysis of the data for a different period and therefore the standardized coefficients are not the same as those of Duncan, although they are fairly close.

Turning our attention to the unstandardized coefficients, we note that there are differences between the two causal diagrams. Whites convert their investment in education into occupational status and their occupational status into income much more efficiently than do blacks: For the same unit of increase in educational attainment, blacks get less than one-half of the increase in occupational levels of whites (.247 vs. .536); for the same unit of increase in occupational status blacks get less than one-third the increase in income levels of whites (.092 vs. 332). There is clear evidence that, for the same type of qualifications, blacks are rewarded less at every stage of the stratification process than are whites.

The examination of standardized coefficients suggests a different conclusion, especially with respect to the path from education to income: Other things being equal, "blacks convert education into income at a higher rate than do whites" (see Duncan, 1968: 96). The source of the misleading conclusion lies in the fact that one is implicitly using different standards (or units of measurement) or earnings for blacks and for whites, because by using standardized coefficients one is actually using a group-specific standard of measurement.[3]

So far we have demonstrated (1) that standardization of variables is logically distinct from the use of standardized coefficients; (2) that comparisons of causal structure based on standardized coefficients can be misleading; and (3) that unstandardized coefficients may be fruitfully used along with selective standardization of variables. We have not shown fully why we must not view the standardized coefficients as measures of the invariant structural properties of causal processes. To this topic we now turn.

CAUSAL LAWS, INVARIANCE, AND STANDARDIZATION

It is apparently as elusive to some as it is self-evident to others that the standardized coefficients cannot in general be viewed as structural parameters describing invariant causal processes. Hargens, for instance, reflects the feeling of many others when he argues that "most of the currently accepted arguments against standardized coefficients beg the question" and that "standardized coefficients can be structural parameters" (1976: 248).

According to Hargens, the proponents of the use of unstandardized coefficients state the invariant causal law (law that is uniform across groups) in terms of unstandardized coefficients as in equation 3,

$$X_0 = a + b_{01}X_1 \qquad [3]$$

and then show that group-specific standardizations do not preserve such an invariance across the groups (for samples) while the unstandardized coefficients do. Hargens then argues, if the invariant causal law is stated instead in terms of standardized variables, as in equation 4,

$$Z_0 = \beta_{01}Z_1 \qquad [4]$$

the unstandardized coefficient is likely to vary from group to group. Hence, the reasoning of the proponents of the unstandardized coefficients is circular. He then proceeds to show that it is plausible to state the causal laws in terms of standardized variables. For example, it is reasonable to postulate that "allocation of rewards within a social system is based on an individual's abilities or contributions *relative to those of the other members of the social system*" (Hargens, 1976: 251), and hence, variables standardized on the mean and variance of a given population may correctly reflect the theoretically relevant metric.

We concur with Hargens's arguments for the selective use of standardized scales; but when he argues, therefore, that the standardized coefficients can be viewed as structural parameters, we must disagree. His argument is based on an implicit assumption that standardization as a part of measurement operation is inseparable from the use of standardized coefficients.

One of the hidden sources of confusion is Hargens's reliance on a simple bivariate deterministic relationship, such as equation 3. Given such a relationship, there will be no sampling variability from sample to sample and the standardized coefficients will be constant from group to group or from sample to sample—the standardized coefficient will always be a unity. (We are assuming no measurement errors here.) But the situation changes drastically when one considers a multiple causation. To use the simplest case, consider that X_0 is completely determined by two causal variables, X_1 and X_2:

$$X_0 = a + b_{01}X_1 + b_{02}X_2 \qquad [5]$$

To simplify even further, also assume that X_1 and X_2 are uncorrelated in the population. In order not to beg the question, we assume that variables can be in some standardized form and a can be zero—that is, we leave the question open as to whether the structural parameters, the b's in equation 3, are standardized coefficients.

Given such an invariant causal law, any selected group or subset out of the same theoretical universe or population should reveal the same relationship. (Note that sampling errors with respect to b's do not exist if the system is deterministic and if there are no measurement errors.) To make the example concrete, assume that the underlying structural equation is given by

$$X_0 = .6X_1 + .8X_2 \qquad [6]$$

TABLE 1
Three Atypical Samples of Four from a Population in Which $X_0 = .6X_1 + .8X_2$ and X_1 and X_2 Have Unit Variances

Units	Subset 1 X_0	Subset 1 X_1	Subset 1 X_2	Subset 2 X_0	Subset 2 X_1	Subset 2 X_2	Subset 3 X_0	Subset 3 X_1	Subset 3 X_2
1	1.4	1	1	2	2	1	2.2	1	2
2	−.2	1	−1	.4	2	−1	−1.0	1	−2
3	.2	−1	1	.4	2	1	1.0	−1	2
4	−1.4	−1	−1	2	2	−1	2.2	−1	−2
S.D.	1	1	1	1.44	2	1	1.71	1	2

Standardized Coefficients

Subset 1 $X_0 = .6X_1 + .8X_2$
Subset 2 $X_0 = .833X_1 + .556X_2$
Subset 3 $X_0 = .351X_1 + .936X_2$

Now examine three subsets of this universe, each of which consists of only four units of observation. These three subsets are not meant to be random samples, although they can be viewed as atypical examples of random samples. Note once again that, given all three variables, X_0, X_1, and X_2, (1) the multiple R in every subset is unity, (2) the structural coefficients remain constant across the groups, and (3) X_0, X_1, and X_2 can be considered as standardized variables in the *population*. Although there is no reason to expect random samples to have the same means on these variables, or to have the same zero covariation between X_1 and X_2, we intentionally allow only one element to vary from one subset to another—the variance of each variable.

Now examine what happens if one uses standardized coefficients. Except for the fact that multiple R is unity for every group, there is no uniformity across the groups in the magnitude of standardized coefficients. The relationships as expressed in terms of standardized coefficients are no longer invariant from subgroup to subgroup (or over replications of experiments). This variation in the standardized b's is solely due to the fluctuation of the relative magnitude of variances in X_1 and X_2. The magnitudes of standardized b's can also vary if the covariation between X_1 and X_2 varies from subset to subset. Since these and other cases are clearly demonstrated elsewhere (Kim and Mueller, 1976), we will not repeat them here. We simply note that the causal law under consideration does not stipulate invariance of variance and covariance across the subsets, and such group-specific variation is irrelevant for the validity of the causal law. In particular, note the fact that we accept the possibility of the structural coefficients in the population being standardized coefficients. The important point is that given an invariant causal process, the unstandardized coefficients can identify its invariance across groups while the use of standardized coefficients cannot identify such invariance. Even if the structural coefficients are assumed to be standardized coefficients in the *population*, one should not assume that group-specific standardizations will reveal such an invariance.

It is clear, then, that comparison of causal structures across subgroups should not be based on standardized coefficients. But we must face a possible argument along the line of Hargens: What if the theoretically relevant scale for a given variable is relative to the distribution of that variable for each subgroup? To make the argument more concrete, let us assume that X_1 in the preceding examples is such a variable, but X_2 is not. Comparison between subsets 1 and 3 shows that X_1 in both groups shows a standardized metric—each X_1 has a mean equal to zero and a standard deviation equal to 1. But the standardized coefficient associated with X_1 varies from one subset to the other, simply because the amount of variance associated with the other variable (X_2) varies from one subset to another. Unless we can assume that the relevant metrics for all the observed as well as *unobserved* variables that affect the dependent variable are relative to group-specific variations, standardized coefficients will never reveal the structural parameters.

The possibility of standardized coefficients reflecting structural parameters cannot be ruled out completely, but we can dismiss the argument easily because it is *unlikely* that every variable that has some effect on the dependent variable will have a group-specific metric. If such group-specific metrics are not relevant for *all* the variables, one must standardize only selected variables for which such group-specific standardization is warranted; one must not implicitly standardize all the variables using subgroup-specific standards, as is done in the use of standardized coefficients. The point is that one must deal explicitly with the problem of establishing equivalence in measurement scales across the groups (and societies) before attempting to establish equivalence in causal structures. It is poor strategy to deal with such a crucial issue as establishing equivalent measurement scales through implicit means.

EQUIVALENCE OF MEASUREMENT SCALES

A fruitful comparison of causal structures requires that the variables under consideration be measured equivalently across the groups to be compared (Schoenberg, 1972). This requirement confronts us with two problems we have not dealt with explicitly: (1) what to do if there is no common scale of measurement across the groups to be compared and (2) what to do if there is no well-defined or well-accepted scale for a given variable. The central issue here is how to obtain comparability of scales across groups; it is *not* how to compare the effects of several variables for a given group when different variables are measured using different scales, such as income and education (which will be discussed in a separate section).

Problem 1 arises if variables are measured using different scales across groups. Clearly, a variable measured differently across the groups cannot be directly compared. If told that it is 40° in El Paso and 70° in Moscow, and that the first is measured in Celsius while the second is measured in Fahrenheit, one cannot determine which city is hotter without bringing the measurement values to a common scale. In this case, given the underlying relationship between the two scales, one can easily convert one to the other, thereby obtaining comparable scale values. Unfortunately, however, measurement problems is social research are

not as simple as the above example. In many instances, it is difficult to achieve comparability of scales across groups, especially when one is comparing different societies. How does one establish equivalence across societies in such variables as social status, income, modernity, and attitude toward a particular issue? This question is beyond our present scope.[4]

For some variables, standardization on the basis of group-specific means and variances may turn out to be the best available solution for achieving comparability of variables across the groups. If so, one must use standardization as an explicit means of achieving comparability of scales before embarking on comparative causal analysis.

Problem 2 is often confused with the first, but needs to be separated from it. Tukey and others (Tukey, 1954; Forbes and Tufte, 1968; Stokes, 1974) have grudgingly admitted that standardized coefficients may be used when variables are measured on arbitrary scales. but we need to delimit the applicability of this advice. By arbitrary scales, we mean such attitude scales as those based on Likert-type scales or on the summation of positive answers to a series of agree-disagree questions. Suppose now that for the subgroups (populations or samples) to be compared all the variables are measured by the same scales, however arbitrary those scales may be. As an example, suppose that respondents from the North and South are each asked 20 race-related agree-disagree questions. For each respondent, a summary score is created by summing the number of items for which liberal responses were given. The resulting 0-to-20 scale is "arbitrary," but all the respondents in both groups are given scores that are based on the same scale. Suppose further that there is greater homogeneity in the South than in the North. A group-specific standardization of the variables will not make the arbitrary scale less arbitrary; it will only make it less comparable across the groups. Hence, the use of standardized coefficients, which implies standardizing all the variables in the analysis on the basis of group-specific means and variances, cannot be recommended when causal structures are compared across groups.

This brings us to the Miller and Stokes data dealing with constituency influence on the voting record of Congressmen, data which has long been the center of controversy (Miller and Stokes, 1963; Cnudde and McCrone, 1966; Forbes and Tufte, 1968; Blalock, 1967b; Stokes, 1974). Most of the variables involved in this analysis are attitude measures for which there is no generally accepted standard of measurement, and the measurement of each variable is based on a more or less arbitrary metric. But in comparing two potentially different subgroups—competitive congressional districts versus noncompetitive ones, or the South versus others—the crucial information that one must consider is not whether these variables are measured on an arbitrary scale, but whether they are measured using the same (comparable) scales across these potential subgroups. As in Duncan's analysis, in which the status-attainment process between blacks and whites is compared, Miller and Stokes use the same scales, however flimsy their metric may be, across these potential subgroups. Therefore, Forbes and Tufte are wrong in insisting on the use of standardized coefficients, which would destroy the comparability of the original scales by an implicit standardization based on group-specific scales.

On the other hand, if a given variable is measured by using different items for different groups, the resulting scales cannot only be arbitrary but also noncomparable across the groups. Such a situation can arise in studies comparing societies with different languages and cultures and in studies on a given society conducted at different periods or by different scholars. The lack of comparable scales in such studies, or the inability to convert the measurements to a common scale, is, as mentioned earlier, the most serious stumbling block of comparative causal analysis. We can only reiterate that if the main focus of study is comparing causal structures across the groups, the crucial issue is not whether variables for given groups are comparable in their measurement scales, but whether the same variable is measured in a comparable scale across the groups. Whenever a variable is measured on a comparable scale across the groups, such comparability must be maintained. This means that one cannot rely

on the use of standardized coefficients, but this does not preclude standardizing variables beforehand on the basis of the standards that are common to all the groups under consideration.[5]

IMPORTANCE OF CAUSAL EFFECTS

Evaluating the relative importance of one causal variable against others is one of the perennial concerns of causal interpretation; and the apparent ease with which such an evaluation can be made with the use of standardized coefficients is probably the most important reason for their persistent use in the literature of the social sciences. However, upon close examination, the notion of causal importance itself is not easy to define (see Blalock, 1961, 1968).

The first issue is whether to define the causal importance of a variable (1) by the proportion of variance in the dependent variable explained by the variation in the given causal variable or (2) by the amount of *changes* or differences in the dependent variable that can be introduced by the *changes* or differences in the causal variable. If one were to use the second criterion—manipulability of the dependent variable—every causal variable can be considered as important as any other causal variable because by introducing sufficient changes in a given causal variable one can always introduce intended changes in the dependent variable. To make this criterion meaningful, one must therefore define the comparable amount of change in two or more causal variables; one may then compare whether a comparable amount of change in one independent variable brings about the same amount of change as does a comparable change in another causal variable. The comparable units may be defined in terms of an arbitrary standard, such as the standard deviation of each variable (see Blalock, 1961).

When this criterion is used, the choice of comparable units of measurement must precede the comparison of the relative contribution of two or more causal variables, and the justification of one's particular choices of units can and must be made on the basis of the particular purposes of the research at hand. Once such comparable scales across the variables are established, the easiest way to compare the causal effects is to standardize variables on these scales and to compare the relative magnitudes of the *unstandardized* coefficients associated with them. In some instances, when all the variables are standardized in terms of a group-specific standard deviation of each variable, such coefficients may be equivalent to the standardized coefficients. But such comparison will be different from an automatic reliance on the relative magnitude of standardized coefficients (1) if two or more groups are compared or (2) if some variables are standardized as usual but other causal variables are not. As an example, suppose one is comparing the effect of sex difference with the effect of family socioeconomic status: One may compare the coefficient associated with the dummy sex variable (coded 0 and 1), with the coefficient associated with the standardized socioeconomic status variable. In this type of comparison the researcher is aware of the nature of the comparison, because it is based on the particular and explicit choice of the units of "comparable" measurement across the variables. Consequently, the limited nature of the meaning of relative importance of causal variables is not overlooked.

Using the first criterion mentioned above (the proportion of variance explained) to determine relative importance has advantages as well as several disadvantages. Its use does not require establishing equivalent changes across the causal variables. What is compared is the relative amount of variation in the dependent variable that can be uniquely assigned to the existing variability of a given causal variable. For instance, when two or more causal variables are uncorrelated, the partial (as well as the bivariate) correlation squared between any causal variable and the dependent variable measures the proportion of variance explained by that causal variable. The relative magnitude of these proportions may be used as a measure of the relative causal importance of one given variable against that of another. Likewise, the magnitude of standardized coefficients can be directly compared for the same purposes, because the relative magnitude based on one type of information is a montonic function of that

based on the other.[6] This is perhaps the main attraction of the use of standardized coefficients. The contribution of each variable obtained in this way is scale-free; for instance, whether a given variable is measured on a Celsius scale or on a Fahrenheit scale, the resulting information will not change. Therefore, one can compare the effects of variables measured on different scales without worrying about the differences in the measurement scales.

But one must remember that causal importance measured by proportion of variance explained is dependent on the group-specific variance structure that is irrelevant or extraneous to the causal law under consideration, and consequently combines too much information into a summary measure to be useful for certain types of causal analysis (see Kim and Mueller, 1976). To elaborate the point that standardized coefficients cannot be causal parameters in a different context, consider once again a simple three-variable causal system:

$$X_0 = b_1X_1 + b_2X_2 \qquad [7]$$

The variance in X_0 is given by

$$\mathrm{var}(X_0) = b_1^2\,\mathrm{var}(X_1) + b_2^2\,\mathrm{var}(X_2) + 2b_1b_2\,\mathrm{cov}\,(X_1, X_2).$$

Suppose now that X_1 is a dummy variable standing for sex and X_2 is a "standardized" socioeconomic status variable (standardized on the variation in the population or the theoretical universe). If one is comparing the causal importance of one variable to the other using standardized coefficients, we are comparing

$$(b_1)\text{ times the square root of }\left(\frac{\mathrm{var}\,(X_1)}{\mathrm{var}\,(X_0)}\right)$$

with

$$(b_2)\text{ times the square root of }\left(\frac{\mathrm{var}\,(X_2)}{\mathrm{var}\,(X_0)}\right)$$

or

$$(b_1)(c)(S_1)$$

with

$$(b_2)(c)(S_2)$$

where c = a constant—the square root of $1/\mathrm{var}(X_0)$—and S_1 and S_2 are the square roots of $\mathrm{var}(X_1)$ and $\mathrm{var}(X_2)$, respectively. Note that in this case the relative magnitude is dependent on the relative values of two components: b_i, the change in X_0 associated with a unit change in X_i, and S_i, the amount of variation in the independent variable X_i. The first component is the criterion that may be used if the manipulability criterion is used as an indicator of causal importance. The second component is completely dependent on the group-specific variability—especially the relative magnitude of one to the other. In a sense, therefore, measures of causal importance based on the use of standardized coefficients contain more information than such measures based on unstandardized coefficients.

A serious problem may, however, arise if this criterion is used in comparing two or more groups, because, as shown earlier, the standardized coefficients assume a group-specific metric for every variable. More specifically, the relative magnitude of standardized b_i will depend on three components

$$(b_i)(c_i)(s_i)$$

because c_i additionally will vary from group to group unless $\mathrm{var}(X_1)$, $\mathrm{var}(X_2)$, and $\mathrm{cov}(X_1, X_2)$ all remain invariant across the groups. Suppose that one is comparing the last two groups in Table 1. According to the magnitude of standardized coefficients, one would conclude that X_1 is more important than X_2 in group 2, while X_2 is more important in group 3. Such a result is obtained

because of the differences in the relative variability of the variables in the different groups:

	$B_1 = b_1\,C_1\,S_1$	$B_2 = b_2\,C_2\,S_2$
group 2:	.833 = .6(1/1.44) (2)	.556 = .8(1/1.44) (1)
group 3:	.351 = .6(1/1.71) (1)	.936 = .8(1/1.71) (2)

Therefore, statements about the relative importance of causal variables based on standardized coefficients lose their specificity and have nothing to do with the relative magnitude of structural parameters.

CONCLUSION

By noting the conceptual distinction between the standardization of scales and the use of standardized coefficients, this article demonstrates, among other things, that (1) the standardized coefficients cannot reveal invariant causal parameters even if *some* of the theoretically relevant metrics are relative to group-specific variations, (2) the use of standardized coefficients is self-defeating even if it is employed to achieve comparability of scales for variables without a well-defined or commonly accepted metric, and (3) there are severe limitations in using standardized coefficients as a means of evaluating the relative causal importance of variables. As an alternative to a habitual reliance on the standardized coefficients, this article proposes the combined use of explicit standardization of some or all variables and unstandardized regression coefficients.

NOTES

1. To convey the tone and subtlety of the controversy, we quote a few lines from political science journals:

Regression coefficients generally provide a better summary of a linear relationship than a correlation coefficient, although for some data this does not apply. For the Miller-Stokes data, unstandardized regressions are inappropriate because their metrics do not possess the substantive meaning required for the sensible use of such coefficients [Forbes and Tufte, 1968: 1270-1271].

Blalock's illustrative use of the very research I have discussed here is somewhat puzzling, to say the least, since the values of non-standardized coefficients would in this case be prey to wholly arbitrary choices of scale factors. Indeed, it is difficult to see how it would be possible to interpret the values of these coefficients without taking into account the variance of the seven measures incorporated in our model. More generally, the claim that non-standardized coefficients always give more fundamental information and are, in particular, more suitable for comparative inquiry seems to be quite false [Stokes, 1974: 197f.].

2. The method used here is an extension of what Hotchkiss (1976) called "standardized path regression." One of the authors developed, independent of Hotchkiss, a similar procedure that was included in the initial draft for an earlier article (Kim and Mueller, 1976).

3. Duncan (1975) emphatically notes the limitations of standardized coefficients. The quote from his writing is not meant to disparage his analysis but only to highlight the temptation to use standardized coefficients even when their limitations are appreciated. We often rely on standardized coefficients partly because the unstandardized coefficients are hard to compare and partly because, when both unstandardized and standardized coefficients are used, the analysis becomes cluttered with numbers. The method of display used in Figure 2 seems to provide a relatively tidy method of presentation.

4. There are certain strategies that can be used to improve the comparability of variables (Przeworski and Teune, 1970), but the final validation of a meaningful measurement scale requires the development of causal theories. In the long run, whether two scales are equivalent can be answered only in the context of theoretical expectation (see Verba et al., 1978).

5. There is a situation in which the equivalence of two causal systems can be established, at least in theory, without the knowledge of the underlying true metric if and only if variables are measured on an interval scale, or if one variable is a linear transformation of another variable, even though the exact form of the transformation may not be known. If two causal systems are equivalent—that is, each of the corresponding causal paths has the same underlying parameter—then the ratio between (1) the direct path from X_i to X_j and (2) the indirect path from one to the other should be invariant from group to group, regardless of the differences in the measurement metric (see O'Brien, 1977).

6. An anonymous reviewer suggested that the relative magnitude of F values can also be used for this purpose because F is also a monotonic function of the R^2 increment due to each variable. Such a use will obviate the need to rely on standardized coefficients. But the magnitude of F is also a function of sample size. Hence, its use in comparative study as a measure of the causal importance of variables should be confined to studies based on similar sample sizes.

REFERENCES

BLALOCK, H. M., Jr. (1968) "Theory building and causal inferences," pp. 155-198 in H. M. Blalock, Jr., and A. B. Blalock (eds.) Methodology in Social Research. New York: McGraw-Hill.

——— (1967a) "Path coefficients versus regression coefficients." Amer. J. of Sociology 72 (May): 675-676.

——— (1967b) "Causal inference, closed populations, and measures of association." Amer. Pol. Sci. Rev. 61 (March): 130-136.

——— (1964) Causal Inferences in Nonexperimental Research. Chapel Hill: Univ. of North Carolina Press.

——— (1961) "Theory, measurement, and replication in the social sciences." Amer. J. of Sociology 66 (January): 342-347.

BLAU, P. M. and O. D. DUNCAN (1967) The American Occupational Structure. New York: John Wiley.

BOUDON, R. (1968) "A new look at correlation analysis," pp. 199-235 in H. M. Blalock, Jr., and A. B. Blalock (eds.) Methodology in Social Research. New York: McGraw-Hill.

——— (1965) "A method of linear causal analysis: dependence analysis." Amer. Soc. Rev. 30 (June): 365-374.

CAIN, G. and H. WATTS (1970) "Problems in making inferences from the Coleman Report." Amer. Soc. Rev. 35 (April): 228-241.

CNUDDE, C. F. and D. J. McCRONE (1966) "The linkage between constituency attitude and congressional voting behavior: a causal model." Amer. Pol. Sci. Rev. 60 (March): 66-72.

DUNCAN, O. D. (1975) Introduction to Structural Equation Models. New York: Academic.

——— (1970) "Partials, partitions, and paths," pp. 38-47 in E. F. Borgatta and G. W. Bohrnstedt (eds.) Sociological Methodology 1970. San Francisco: Jossey-Bass.

——— (1968) "Inheritance of poverty or inheritance of race?" pp. 85-110 in D. P. Moynihan (ed.) On Understanding Poverty. New York: Basic Books.

——— (1966) "Path analysis: sociological examples." Amer. J. of Sociology 72 (July): 1-16.

FELSON, M. (1974) "Standardization of scales in social science research: breaking the routine." Social Sci. Research 3 (September): 261-265.

FORBES, H. D. and E. R. TUFTE (1968) "A note of caution in causal modeling." Amer. Pol. Sci. Rev. 62 (December): 1258-1264.

HARGENS, L. L. (1976) "A note on standardized coefficients as structural parameters." Soc. Methods & Research 5 (November): 247-256.

HEISE, D. R. (1975) Causal Analysis. New York: John Wiley.

——— (1969) "Problems in path analysis and causal inference," pp. 38-73 in E. F. Borgatta and G. W. Bohrnstedt (eds.) Sociological Methodology 1969. San Francisco: Jossey-Bass.

HOTCHKISS, L. (1976) "A technique for comparing path models between subgroups: standardized path regression coefficients." Soc. Methods & Research 5 (August): 53-76.

KIM, J.-O. and C. W. MUELLER (1976) "Standardized and unstandardized coefficients in causal analysis: an expository note." Soc. Methods & Research 4 (March): 428-438.

LAND, K. C. (1969) "Principles of path analysis," pp. 3-37 in E. F. Borgatta and G. W. Borhnstedt (eds.) Sociological Methodology 1969. San Francisco: Jossey-Bass.

MILLER, W. E. and D. E. STOKES (1963) "Constituency influence in Congress." Amer. Pol. Sci. Rev. 57 (March): 45-56.

O'BRIEN, R. M. (1977) "A solution to the problem of 'differences in units of measure' in cross-population comparisons." Soc. Methods & Research 5 (May): 471-484.

PRZEWORSKI, A. and H. TUENE (1970) The Logic of Comparative Social Inquiry. New York: John Wiley.

SCHOENBERG, R. (1972) "Strategies for meaningful comparison," pp. 1-35 in H. L. Costner (ed.) Sociological Methodology 1972. San Francisco: Jossey-Bass.

SPECHT, D. A. and R. D. WARREN (1976) "Comparing causal models," pp. 46-82 in D. R. Heise (ed.) Sociological Methodology 1976. San Francisco: Jossey-Bass.

STOKES, D. E. (1974) "Compound paths: an expository note." Amer. J. of Pol. Sci. 18 (February): 191-214.

TUKEY, J. W. (1954) "Causation, regression, and path analysis," pp. 35-66 in O. Kempthorne et al. (eds.) Statistics and Mathematics in Biology. Ames: Iowa State College Press.

TURNER, M. E. and C. E. STEVENS (1954) "The regression analysis of causal paths." Biometrics 15 (June): 236-258.

VERBA, S., N. H. NIE, and J.-O. KIM (1978) Participation and Political Equality: A Seven Nation Comparison. London: Cambridge Univ. Press.

WRIGHT, S. (1960) "Path coefficients and regression coefficients: alternative or complementary concepts?" Biometrics 16 (June): 189-202.

——— (1934) "The method of path coefficients." Annals of Mathematical Statistics 5 (March): 161-215.

Jae-On Kim is Professor of Sociology at the University of Iowa. Areas of research interest are: political sociology, quantitative methods, and social stratification. His publications include coauthored books, Participation and Political Equality *(Cambridge University Press, 1978),* Introduction to Factor Analysis *(Sage, 1978), and* Factor Analysis: Statistical Methods and Practical Issues *(Sage, 1978).*

G. Donald Ferree, Jr., is Associate Director of the Roper Center, and Assistant Professor of Political Science at the University of Connecticut. He is presently co-authoring, with Sidney Verba and Gary Orren, a book on the attitudes of American elites toward equality.