Journal of the Society for Social Work and Research
Vol.1, No.1, 28–40
June 2010
ISSN 1948-822X, DOI: 10.5243/jsswr.2010.2

Suppressor Variables in Social Work Research: Ways to Identify in Multiple Regression Models

Shanta Pandey
Washington University in St. Louis

William Elliott
University of Pittsburgh

Suppressor variables may be more common in social work research than what is currently recognized. We review different types of suppressor variables and illustrate systematic ways to identify them in multiple regression using four statistics: R^2, sum of squares, regression weight, and comparing zero-order correlations with respective semipartial correlations.

Keywords: suppressor variable; semipartial correlation

When selecting a set of study variables, social work researchers frequently test correlations between the outcome variables (i.e., dependent variables) and theoretically relevant predictor variables (i.e., independent variables). In some instances, one or more of the predictor variables are uncorrelated with the outcome variable. This situation poses the question of whether researchers' multiple regression analyses should exclude independent variables that are not significantly correlated with the dependent variable. Questions such as this are routine, and our article provides a systematic answer to these questions. In the multiple regression equations, suppressor variables increase the magnitude of regression coefficients associated with other independent variables or set of variables (Conger, 1974). A suppressor variable correlates with other independent variables, and accounts for or suppresses some outcome-irrelevant variation or errors in one or more other predictors, and improves the overall predictive power of the model. Given this function, some prefer to call the suppressor variable an *enhancer* (McFatter, 1979). A variable may act as a suppressor or enhancer—even when the suppressor has a significant zero-order correlation with an outcome variable—by improving the relationship of other independent variables with an outcome variable. This type of suppressor variable is more likely to be retained in a regression model than a variable that has a zero correlation with the outcome variable. However, this article aims to underscore the value of retaining these variables even when they are uncorrelated with outcome variables in zero-order correlation. To accomplish this goal, we first address the properties of suppressor variables and their prevalence in social work research, and then discuss strategies for using suppressor variables in multiple regression equations.

Item analysis is a common technique used to eliminate variables when the relationship of each predictor variable with an outcome variable is tested separately for statistical significance. Predictor variables that are not significantly related to outcome variables are often eliminated at the bivariate level. Bivariate results, such as zero-order correlation coefficients, provide only partial information about the relationship between a predictor and an outcome variable, and are an improper method for selecting variables for a multiple regression model. Some researchers have reported that when a multiple regression model included a predictor variable that was uncorrelated with the outcome variable in a bivariate model, the uncorrelated predictor variable sometimes significantly improved the explained variance (Courville & Thompson, 2001; Horst, 1941; McNemar, 1945; Meehl, 1945; Shieh, 2006). Under such circumstances, "the whole regression can be greater than the sum of parts" (Bertrand & Holdler, 1988, p. 371). Nevertheless, researchers often prematurely eliminate these variables during their variable selection process based on the variable's very low bivariate correlation with the dependent variable (Horst, 1941; Meehl, 1945; Shieh; 2006; Velicer, 1978). However, eliminating these uncorrelated variables will cause the researcher to underestimate some of the parameters, will undermine the predictive power of the model, and may yield regression equations which are overly sample-specific. Therefore, to accurately assess the contribution of each independent variable to the dependent variable, all theoretically relevant independent variables must be retained, including those variables that may not be correlated with the dependent variable at the bivariate level. Parsimonious use of a

Shanta Pandey is an associate professor at the George Warren Brown School of Social Work, Washington University in St. Louis.

William Elliott is an assistant professor at the University of Pittsburgh, School of Social Work, wie1@pitt.edu

Correspondence concerning this article should be addressed to Shanta Pandey at pandeys@wustl.edu

number of independent variables in regression models increases statistical power of tests (see for example, Cohen, Cohen, West, & Aiken, 2003; Tabachnick & Fidell, 2007), but elimination of theoretically relevant variables may result in underestimation of parameters.

Background

The concept of suppression is not new; various authors have explored this concept for more than 70 years. Mendershausen (1939) termed a set of variables *clearing variates* when the variables had no connection with the outcome variable but were connected with one or more predictor variables. These clearing variates helped determine the variance in the outcome variable because of their connection with other predictor variables. Horst (1941) separated the predictor variables into two sets: the first set had "appreciable" correlations with the outcome variable, whereas the second set included those variables that had "negligible" correlation with the outcome variable but substantially correlated with other predictor variables in the model. Horst termed the first set of variables *prediction variables* and the second set *suppression variables*. Horst explained that most prediction variables had components that were both related and unrelated to the outcome variable. Therefore, when predictor variables were included in the regression model, those variables explained not only part of the outcome variable but also parts of other predictors that were unrelated to the outcome variable. Further, Horst (1941) held that if a regression model included suppressor variables that were independent of the outcome variable but correlated with the components of the prediction variable that were independent of the outcome variable, "we should be able to suppress the irrelevant components of the prediction variables" (p. 434). Horst indicated that in a regression model with prediction and suppression variables, "we would in general have positive weights for the prediction variables and negative weights for the suppression variables" (pp. 434-435). Including these two sets of variables in multiple regressions should allow for the "most efficient and economical prediction" (Horst, 1941, p. 435).

Following Mendershausen (1939) and Horst (1941), many others have explored the role of suppressor variables in multiple regression (see for example, Conger, 1974; Lubin, 1957; McNemar, 1945; Meehl, 1945; Shieh, 2006; Tzelgov & Henik, 1991; Velicer, 1978). Most authors describe suppressor variables as predictor variables that are not correlated (or correlated minimally) with the dependent variable, even though such predictor variables correlate with one or more of the independent variables. The correlations of suppressor variables with the independent variables increase the multiple correlations by partialling out invalid variance of the other predictors included in the regression equation (Cohen et al., 2003; Horst, 1941; Maassen & Bakker, 2001; McNemar, 1945; Meehl, 1945). In other words, suppressor variables are predictors that in isolation correlate weakly (or zero) with the outcome variable but are strongly correlated with one or more predictors that are correlated with the outcome variable. This definition implies that the component of the predictor variable associated with the outcome variable has *noise*, that is, irrelevant variance; the suppressor variable suppresses or explains the part of the predictor variable that is irrelevant and not associated with the outcome variable (Meehl, 1945). These variables are called suppressors because they suppress outcome-irrelevant variance in other predictors, causing the suppressed variables to obtain a substantial regression weight (Tzelgov & Henik, 1991).

Types of Suppressor Variables

Since Horst's (1941) introduction of the concept of suppression, many authors have expanded the definition of these variables (see for example, Conger, 1974; Darlington, 1968; Lubin, 1957; Lutz, 1983; Velicer, 1978). There are four types of suppressor variables: the classic suppressor, the negative suppressor, the reciprocal suppressor, and the absolute and relative suppressor. We briefly introduce each type below.

Classic suppression in multiple correlations was originally introduced by Horst (1941) and was later demonstrated mathematically by Meehl (1945) and McNemar (1945). Although a suppressor and an outcome variable have a zero correlation, the prediction in the outcome variable increases when a suppressor variable is added to the equation simply because the suppressor variable is correlated with another predictor (or set of predictors) that are correlated with the outcome variable. In this case, the suppressor variable removes irrelevant predictive variance from the other predictor (or set of predictors) and increases the predictor's regression weight, thus increasing overall model predictability. Sometimes the suppressor variable may also receive nonzero regression weight with a negative sign. However, "a variable is a suppressor only for those variables whose regression weights are increased. Thus, a suppressor is not defined by its own regression weight but rather by its effects on other variables in a regression system" (Conger, 1974, p. 37). For example, examine the following regression equation with two predictors:

$$Y = a + b_1X_1 + b_2X_2 + e \quad (1)$$

where Y represents the predicted value of the outcome variable, *a* is the intercept or the point in Y-axis where the slope originates, and *e* represents the error or the proportion of the variance in Y not associated with the independent variables in the model (X_1 and X_2 in this equation). X_1 is a predictor variable; X_1 and Y are positively correlated both in bivariate and multiple regression equations. X_2 is another predictor variable that is positively correlated with $X_{1,}$ but X_2 is not correlated with Y. Including X_2 in the equation will increase the regression weight of $X_{1.}$ Sometimes X_2 may have an improved but negative regression weight; X_2 is a classic suppressor.

Negative suppression was introduced by Lubin (1957) and later explained mathematically by Darlington (1968) and Conger (1974). A negative suppressor works in a manner similar to that of a classic suppressor by removing irrelevant variance from a predictor (or set of predictors), increasing the predictor's regression weight, and increasing overall predictability of the regression equation. The difference between these two types of suppressors is the negative suppressor's positive zero-order correlation with other predictor variable (s) and with the outcome variable; however, when entered in multiple regressions, the negative suppressor has a negative beta (β) weight (Conger, 1974; Darlington, 1968; Lutz, 1983; Maassen & Bakker, 2001). In other words, contrary to what is expected, the regression weight of the negative suppressor has an opposite sign. For instance, under negative suppression in the equation (1), X_2 (negative suppressor) will be positively correlated with both Y and X_1 at the bivariate level. Because of the correlation of X_2 with one or more other predictor variables, including X_2 in a multiple regression equation produces two changes: the regression weight of X_1 increases and the regression weight of X_2 will be significant although this value will have a negative sign. The negative sign indicates that X_2 correlates highly with the error in X_1 (Darlington, 1968).

Reciprocal suppression was introduced by Conger (1974). Some authors have also called this concept *suppressing confounders* (Cohen et al., 2003). Here, both the predictor and the suppressor are positively correlated with the outcome variable but negatively correlated with each other. Recall Equation 1: Under reciprocal suppression both X_1 and X_2 correlate with Y and with each other, and the part that X_1 and X_2 share with each other is the part that is irrelevant to Y. In this situation, X_1 and X_2 will have a negative zero-order correlation. When Y is regressed on these two variables, X_1 and X_2 will suppress some of their irrelevant information, increase the regression weight of the other, and thus improve model R^2.

Absolute and relative suppression was originally introduced by Conger (1974) and further clarified by Tzelgov and Henik (1991). According to Tzelgov and Henik, "absolute suppression is defined by the relationship between the predictor's weight in bivariate regression equation and its weight in multivariate equations. It exists whenever adding predictors increases the weight of the variable relative to its weight in the bi-variate equation" (p. 527). On the other hand, if the regression weight of a predictor increases when a new variable is added to a regression equation, but the increase is not beyond the respective weight of the predictor in the bivariate mode, then the new variable is a relative suppressor (Tzelgov & Henik, 1991). Therefore, relative suppression is tested hierarchically, and the researcher must compare the standardized beta (β) weights of the predictors in the equation before and after the inclusion of the variable that may be a potential relative suppressor. Hence, relative suppression should be tested only when there are three or more predictors (Tzelgov & Henik, 1991). Given that most regression analyses involve more than two predictors, some researchers think that suppression situations are always relative (Tzelgov & Henik, 1991). In addition, it is "possible for a variable to act as a negative suppressor for one variable while simultaneously acting as a reciprocal suppressor for another" (Conger, 1974, p. 43).

Suppressor Versus Mediator Variables

Suppressor variables should not be confused with mediating variables, even though the statistical models for testing mediation and suppression are identical. MacKinnon, Krull, and Lockwood (2000) explained these two categories of variables using a three-variable model. In Mackinnon et al.'s first equation (Equation 2), the outcome variable Y is regressed on the independent variable X_1. In their second equation (Equation 3), Y is regressed on X_1 and another variable X_2, which could be either a mediator or a suppressor. In both equations, Y is the predicted value of the outcome variable, *a* is the intercept and *e* is the unexplained variance, b_1 and b_2 are the regression coefficients associated with X_1 and b_3 is the regression coefficient associated with $X_{2.}$

$$Y= a+ b_1X_1+e \quad (2)$$

$$Y= a+ b_2X_1+ b_3X_2 +e \quad (3)$$

If X_2 is a mediator, then X_1 would be hypothesized to affect the outcome variable directly or indirectly through the third variable ($X_{2,}$ mediator). When a mediator is added in the model, b_2 will be absolutely smaller than $b_{1,}$ the effect of X_1 after controlling for the effect of the mediator (X_2) and that b_1 and b_2 will share the same sign. On the other hand, if b_1 is smaller than b_2 or has an opposite sign, then X_2 must be a suppressor variable. Thus, omission of the suppressor variable from the model will lead to an underestimation of the relationship between independent and outcome variables. Although mediating variables are widely used and discussed in social work research, suppressor variables are less well understood. In the following section, we review the wide prevalence of these variables, discuss the extent to which they are recognized, and document simple ways to use suppressor variables.

Method

We undertook a review of social science literature and various databases to understand how often suppressor conditions are mentioned in social work research. Next, we designed a sample study for the purpose of illustrating and documenting ways to use suppressor variables in multiple regression models.

How Common Are Suppressor Variables?

The use of suppressor variables in multiple regressions is more common than currently recognized (Rosenberg, 1973; Tzelgov & Henik, 1991). This lack of recognition may stem from the fact that suppressor variables are not necessarily a special category of variables; they can be any predictor (or independent) variable in a multiple regression model, including variables for race/ethnicity, income, education, and self-worth. Using a multiple regression model to predict the salary of administrators at educational institutions, Walker (2003) found that the variable for *level of education attained* acted as a suppressor variable. The variable for level of education had a near zero (but positive) zero-order correlation with administrators' salaries (dependent variable) at both public and private institutions (r=.010 for public, and .014 for private institutions). However, the model's regression coefficient associated with the level of education was not only statistically significant but was also a negative. This finding prompted Walker to test the *level of education* variable for its suppression effect. He noted that at the bivariate level, the *level of education* variable was weakly correlated with the dependent variable (salary) but was significantly correlated with other independent variables, including respondent's age. To determine if *level of education* was a suppressor variable, Walker ran the regression model with and without the *level of education* variable included in the models predicting salary. In the model for public institutions, the addition of the *level of education* variable increased the R^2 from .26 to .28. In the model for private institutions that excluded the *level of education* variable, the R^2 was .22; the inclusion of the *level of education* variable increased the R^2 to .36. Walker concluded that the *level of education* was a suppressor variable in predicting salary of administrators for both public and private educational institutions.

In another study predicting employment among community college students with disabilities, Martz (2003) noted that *work experience* served as a suppressor variable. Martz used a series of psychological and demographic independent variables, including paid-work experience, to predict employment status among college students with disabilities. Although *paid-work experience* by itself was not a significant predictor of employment, the predictive power of the model improved substantially when this variable was included in the model with other variables. For instance, the model that included the *paid-work experience* variable had almost 3 times greater predictive power (Nagelkerke R^2 = 0.46) as compared to the model that excluded the *paid-work experience* variable (Nagelkerke R^2 = 0.13). Martz concluded that *paid-work experience* acted as a suppressor variable in the model, explaining variance in employment status.

The effect of suppressor variables has also been examined by Paulhus, Robins, Trzesniewski, and Tracy (2004) in their personality research. Specifically, these authors noted that in a study of aggression among more than 4,000 undergraduate students, the measures of shame and guilt were reciprocal suppressors in explaining aggression. Although the variables used for *shame* and *guilt* in the study analyses shared considerable variance (R^2 ranging from .43 to .48) and correlated positively at the bivariate level, these variables produced divergent outcomes. Variables for *guilt* and *shame*, with positive intercorrelations, were included in the regression one at a time as predictors of aggression. The effect of the *shame* variable on aggression was negative (beta = -0.13) but increased when the *guilt* variable was added to the equation (beta= -0.23). For this regression, R^2 increased from 3%, when *guilt* alone in the model, to 9%, when *shame* and guilt were both in the model. Similarly, when *shame* alone was present in the equation predicting aggression, the beta associated with *shame* was .10 and R^2 is .04. The addition of the *guilt* variable increased the beta associated with shame to .21 and R^2 increased to .11, confirming that *shame* and *guilt* had a mutual reciprocal suppression effect.

Similarly, the suppressor effect of a variable for cognitive ability was demonstrated by MacNeill, Lichtenberg, and LaBuda (2000) in a study examining outcomes of medical rehabilitation among older adults. Specifically, the study examined the probability of a patient's returning to independent living (i.e., living alone) versus living with others. MacNeill and colleagues noted that demographic variables for age and education became significant predictors of return to independent living only when the model included the variable for *cognitive ability*. Although the authors concluded the *cognitive ability* variable produced a suppressive effect, they did not analyze the nature of suppression.

Daniel (1996) conducted multiple experiments to determine if mathematical achievement during the high-school years served as a suppressor variable on gender differences in spatial performance measured among undergraduate students. Consistently in every experiment, he found that the *mathematical achievement* variable acted as a suppressor variable. Further, Daniel reported that in a series of experiments that included regressing a spatial performance variable on gender, controlling for mathematical achievement, and comparing the zero-order correlation between gender and spatial performance with partial correlation coefficient, the partial correlation (beta) associated with gender was higher (.46) than the zero-order correlation (.43). Thus, this finding confirmed the *mathematical achievement* variable as a suppressor variable on the relation between gender and spatial performance.

In Hannah and Morrissey's (1987) study of the development of psychological hardiness in Canadian adolescents, the authors regressed hardiness scale on five independent variables (i.e., *age*, *school grade*, *religion*, *sex*, and *feeling of happiness*) and concluded that the *age* variable acted as a suppressor. In their first set of analyses, Hannah and Morrissey regressed hardiness on variables for *age* and *school grade*. At the bivariate level, *age* and *hardiness* were not related whereas *grade* and *hardiness* were both positively and significantly related (standardized coefficient, b =.19). When *hardiness* was regressed on *age* and *school grade*, two changes occurred: (a) the standardized regression coefficient of *grade* on *hardiness* improved substantially (increased to .67), and (b) the affect of *age* on *hardiness* was both negative and significant. Thus, the authors suggested that hardiness in adolescents develops through a combination of age and schooling. This work is a classic illustration of suppression. Only after the outcome-irrelevant variance of each predictor is removed, does the regression reveal the real magnitude of the contribution made by the two independent variables to the development of hardiness.

In yet another study that examined factors contributing to the use of social and overt aggression among adolescents, Loukas, Paulos, and Robinson (2005) found that the *anxiety* variable acted as a suppressor and was positively associated with boys' and girls' social aggression. Although *anxiety* was not associated with *overt aggression for girls*, it was significantly and negatively associated with *overt aggression for boys*. These authors did not test for suppressor effect of the *anxiety* variable; however, because the zero-order correlation between *anxiety* and *overt aggression for boys* was positive and significant (r=.12), Loukas and colleagues implied that "the negative association between boys' reports of social evaluative anxiety and overt aggression may be the result of a suppressor phenomenon" (p. 342).

These and other examples (Caputo, 2000; Korn & Maggs, 2004; Onwuegbuzie, 2001; Voyer, 1996) point to the routine encounter of suppression variables in social work research, but recognition of this encounter is sparse. A search of the PsycINFO database for articles published between 1945 and 2007 with "suppressor variable" appearing in any field (including the title, abstract, or full text) yielded 103 articles. A search of three additional databases—Social Work Abstracts since 1977, Social Services Abstracts since 1980, and Sociological Abstracts since 1963—located eight more articles with the term "suppressor variable" appearing in any field. Among the journals that had published articles acknowledging the presence of "suppressor variable," in their analysis, most had published only one or two articles, with the exception of *Educational and Psychological Measurement*, which yielded 46 articles. The next highest count of articles with any reference to suppressor variables was in *Sociological Methods and Research* with four; *The Journal of Applied Psychology* with three; *European Journal of Personality*, *Journal of Youth and Adolescence*, *Psychological Bulletin*, and *Social Indicators Research*, with two each; and *Social Work Abstracts* with only one article. This search, however, does not account for social work researchers who published their work in nonsocial work journals, including *Educational and Psychological Measurement*; thus, our search might have underestimated the number of publications with a discussion on suppression effect produced by social work researchers. Our attempt here is to inform social work students and researchers who might be unaware of the methods available to control for these variables in their studies. We now turn to an illustration of how suppressor variables are understood in multiple regression models.

Sample Study Design to Control for Suppressor Variables

Several authors have suggested understanding suppressor variables by examining regression weights (Cohen et al., 2003; Conger, 1974; Darlington, 1968; Tzelgov & Henik, 1991). Instead of the regression weights, other authors have preferred squared semipartial correlation of the suppressor variable in evaluating suppressor effect of a variable (Pedhazur, 1997; Smith, Ager, & Williams, 1992; Velicer, 1978). In the current sample, we show how the suppressor variable is understood in multiple regressions by using four different statistics: R^2, hierarchical and simultaneous sum of squares, regression weights, and semipartial correlation coefficients.

Data

Solely for the purpose of illustration, our example that follows uses world data. These data are indicators of a country's social development between 1970 and 1983 compiled by Dr. David Gillespie of Washington University from various sources including the World Bank; the World Development Report; and the United Nations Educational, Scientific and Cultural Organization's Statistical Yearbook. The data have a sample size of 192 countries and 101 variables. A limitation of this analysis is that we have treated each country as an independent unit of analysis and have assigned the same weights to each country irrespective of its geographic location or population size. We have ignored limitations that are inherent in the use of such data. Readers should ignore all implications of our findings, taking away from this exercise only the discussion that pertains to the suppressor variable.

Hypothesis

We hypothesized that countries with poor health conditions will incur a higher level of external public debt. Specifically, we tested the following hypotheses:

- The higher a country's death rate, the higher will be its external public debt.
- The higher a country's population per physician ratio, the higher will be its external public debt.

Measures

We picked three variables from the world data: (a) *death rate of a country in 1978,* (b) *population per physician in 1980,* and (c) *external public debt in 1980 as a percent of gross national product* (GNP). We treated a country's *population per physician in 1980* and *death rate* as predictor variables and *external public debt* as the outcome variable. The variable, *external public debt in 1980 as a percent of GNP* was operationalized as debt owed by a country to foreign governments, banks, or international institutions including the International Monetary Fund and the World Bank. *Death rate of a country in 1978* represented number of deaths in a country per one thousand population in 1978 and p*opulation per physician in 1980* captured the ratio of total population per physician of a country in 1980.

Analysis and Results

The first step of analysis involved a bivariate Pearson's product-moment correlation of three variables for *death rate*, *population per physician*, and *external debt* (see Table 1). *Population per physician* was not correlated with *external public debt* but was significantly related to *death rate* (r=.70). The second analytic step involved examining any potential adverse effects of correlated independent variables on regression coefficients. To this end, we checked for multicollinearity in these two variables. The tolerance value associated with each of the two independent variables was .53. Most authors who have written on multicollinearity concur that a tolerance of this magnitude will not impair the precision of the parameter estimates (Fox, 1991; Lewis-Beck, 1980; Morrow-Howell, 1994). The third step involved assessment of *population per physician* as a potential suppressor. Although *population per physician* was not related to the outcome variable, the *population per physician* variable was significantly related to the other predictor variable (i.e., *death rate of a country in 1978*) and, therefore, it was important to explore the presence of a possible suppressor.

Table 1

Bivariate Pearson Product-Moment Correlation

	Death rate of a country in 1978	*Population per physician in 1980*	*External public debt in 1980 as a % of GNP*
Death rate of a country in 1978 (independent variable)	1.000	0.703***	0.259**
Population per physician in 1980 (suppressor variable)		1.000	-0.008
External public debt in 1980 as a % of GNP (dependent variable)			1.000

We ran three regression models (see Table 2). In Model 1, we regressed our outcome variable *external public debt* on the predictor variable *death rate of a country,* which was significant at the bivariate level. The model was significant and accounted for 6.69% of the variance (R^2=.0669) in the outcome variable. *Death rate of a country* was positively associated with the country's accumulated external public debt (b = .99, t = 2.51, p < .01). As a country's death per thousand population increased by one unit, its public debt increased by .99% of its GNP. In Model 2, we regressed *external public debt* on *population per* *physician*. As expected, the model was insignificant (R^2 = .0001), indicating that *population per physician* alone was not associated with *external public debt*.

Table 2

Regression of External Public Debt on Two Predictors Separately and Together (N = 90)

Model 1 : Regression of a country's external public debt on death rate				
Model information				
Model Sum of Squares	2850.118			
F value	6.310**			
R^2	0.067			
Adj R^2	0.056			
Parameter estimates	b	β		
Intercept	15.508**	0		
Death rate of a country in 1978	0.992**	0.259		
Model 2: Regression of a country's external public debt on population per physician				
Model information				
Model Sum of Squares	2.842			
F value	0.010			
R^2	0.000			
Adj R^2	-0.011			
Parameter estimates	b	β		
Intercept	29.577**	0		
Population per physician in 1980	-0.014	-.008		
Model 3: Regression of a country's external public debt on death rate and population per physician				
Model Information				
Model Sum of Squares	5618.848			
F value	6.610			
R^2	0.132			
Adj R^2	0.112			
Type I Sum of Squares				
Population per physician in 1980	2.842			
Death rate of a country in 1978	5616.006			
Type III Sum of Squares				
Population per physician in 1980	2768.729			
Death rate of a country in 1978	5616.006			
Parameter estimates	b	β	Squared Semi-partial correlation Type I	Squared Semi-partial correlation Type II
Intercept	9.166	0		
Population per physician in 1980	-0.597**	-0.350	0.000	0.065
Death rate of a country in 1978	1.914***	0.499	0.132	0.132

*p <.05; **p < .01; *** p <.001; b = unstandardized reg. coefficient; β = standardized reg. coefficient.

In Model 3, we regressed *external public debt* on both *death rate* and *population per physician*. The model showed significance and accounted for 13.19% of the variance (R^2=.13; Adjusted R^2=.11) in the outcome variable. An R^2 change of 6.50% was statistically significant [F(1, 87)=6.52; p < .05)] because the critical value of F at the same degrees of freedom and alpha level was lower (F= 3.95) than the obtained sample F of 6.52. However, the R^2 is sensitive to the number of independent variables in the model and will always increase with the addition of variables. Hence, the adjusted R^2, which was less sensitive to the number of independent variables in the model, was used to assess the statistical significance of change in explained variance. Change in adjusted R^2 of 5.56% was also significant [$F(1, 87)=5.38$; $p < .05$)] because the critical value of F at the same degrees of freedom and alpha level was lower (F= 3.95) than the obtained sample F of 5.38.

The hierarchical or Type I sum of squares and simultaneous or Type III sum of squares for the two independent variables changed substantially. *Death rate of a country* alone was associated with 2850.118 of the total sum of squares in the outcome variable (Model 1). When *population per physician* was controlled, the sum of squares associated with *death rate of a country* almost doubled to 5616.006 (Types I or III sum of squares, Model 3).

When *population per physician* was present alone, the model sum of squares was 2.84 (see Model 2). However, when *population per physician* was present along with *death rate of a country*, the model sum of squares increased to 2768.73 (see Model 3, Type III sum of squares). Moreover, in the presence of the suppressor variable, the Type III sum of squares when summed up (2768.73+5616.01=8384.74) was higher than the total model sum of squares (=5618.848).

The parameter estimates associated with both the predictor variables were not only significant but were also larger than when these variables were correlated with the outcome variable individually (see Model 3 in Table 2). As a country's death per thousand population increased by one unit, its public debt increased by 1.91% of its GNP, while controlling for other independent variables in the model (b=1.91, t=3.63, $p < .000$). Although the relationship between the suppressor and the dependent variables was significant, the relationship was in the opposite direction. As population per physician increased by one unit, external public debt of that country decreased by approximately .60% of its GNP (b=-.60, t=-2.55, $p < .01$). We hypothesized that this relationship would be positive. This regression is a typical example of classic suppression, in which the suppressor has a positive zero-order correlation with other predictor variables, and the zero-order correlation with the outcome variables is not significant; however, when the suppressor is entered in multiple regressions, it has a negative beta (β) weight (Conger, 1974; Darlington, 1968; Lutz, 1983; Maassen & Bakker, 2001).

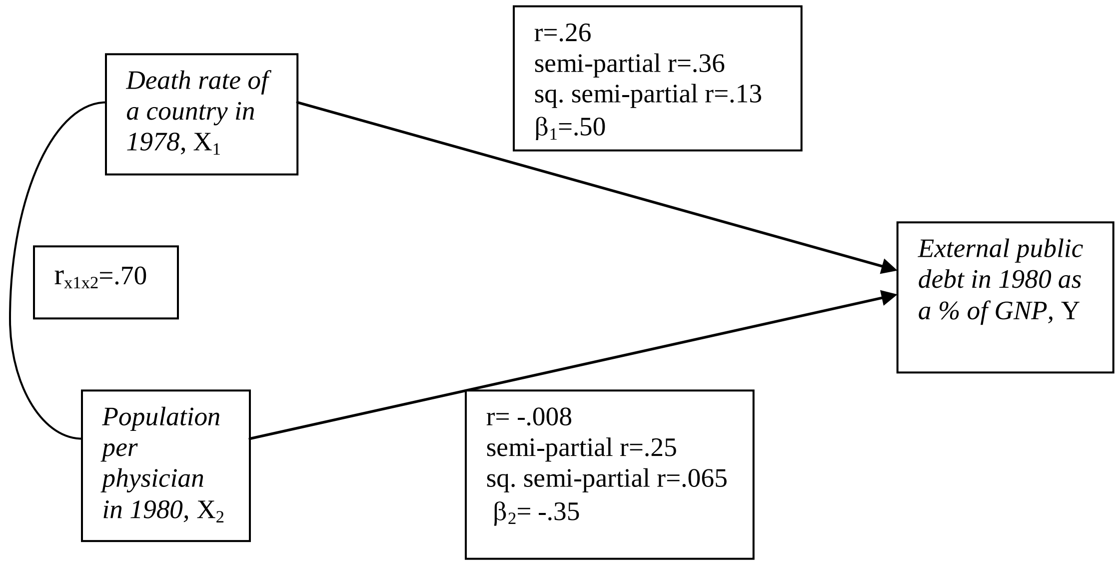
Finally, we compared regression coefficients and semipartial correlation coefficients with the zero-order correlation coefficients of the outcome variable and two predictor variables (see Figure 1). The zero-order correlations between the outcome variable and the two predictors were less than their respective semipartial correlation coefficients, confirming the suppressor effects. Comparing the zero-order correlation of a dependent variable and independent variables with their respective semipartial correlation helps to identify a suppressor variable. When a suppressor variable is present, a semipartial correlation will be larger than its respective zero-order correlation (see Figure 1) because a zero-order correlation is an unpartialed effect. Hence, a zero-order correlation coefficient does not reflect the true relationship. Instead, respective standardized regression coefficients (β) reflect the appropriate relationship between the independent and the dependent variables. Therefore, inspecting a zero-order correlation matrix is not sufficient to reveal the potential utility of variables that are simultaneously incorporated into a model. This situation illustrates the risk of using bivariate reduction of variables to select a set for inclusion in a model.

Discussion

In this section, we summarize the advantages of accurately identifying suppression effects and the benefits of using suppressor variables in multiple regression analyses. Using suppressor variables in multiple regressions will yield three positive outcomes: determining more accurate regression coefficients associated with independent variables; improving overall predictive power of the model; and enhancing accuracy of theory building.

First, the risks associated with excluding a relevant variable are much greater than the risks associated with including an irrelevant variable. The regression weight of an independent variable may change depending upon its correlation with other independent variables in the model. If a suppressor variable that should have been in the model is missing, that omission may substantially alter the results, including an underestimated regression coefficient of the suppressed variable, higher model error sum of squares, and lower predictive power of the model. An incomplete set of independent variables may not only underestimate regression coefficients, but in some instances, will increase the probability of making a Type II error by failing to reject the null hypothesis when it is false. In contrast, although including irrelevant variables in a model can contribute to multicollinearity and loss of degrees of freedom, those variables will not affect the predictive power of the model. Hence, the risk of excluding a relevant variable outweighs the risk of including an irrelevant variable.

Figure 1. A comparison of zero-order correlation, semipartial correlation, and standardized β in a three-variable situation.



To avoid underestimating a regression coefficient of a particular independent variable, it is important to understand the nature of its relationship with other independent variables. The concept of suppression provokes researchers to think about the presence of outcome-irrelevant variation in a independent variable that may mask that variable's genuine relationship with the outcome variable.

Only when a predictor variable that is uncorrelated with other predictors is included in a multiple regression, will the regression weight of other predictor variables remain stable and not change (Courville & Thompson, 2001). However, in most social work research, explanatory variables are intercorrelated, and regression coefficients are calculated after adjusting for all the bivariate correlations between independent variables. When the multiple regression model is altered by adding a variable that is correlated with other predictor variables, the usual outcome is that the correlated variable reduces the regression weight of the other predictor variable(s). The impact will be different if the added variable (or set of variables) is a suppressor variable. The suppressor variable will account for irrelevant predictive variance in some predictors and, therefore, will yield an increase in the regression weight of those predictors. Moreover, the regression weight of the suppressor may improve, thus improving the overall predictive power of the model (Courville & Thompson, 2001).

Another way to think about this situation is to remember that whenever a suppression effect exists in a model, the zero-order correlation between independent and outcome variables may be misleading because it provides only partial information about the relationship between a predictor and an outcome variable. It is a poor indicator of the potential value of a variable for a multiple regression model. Adding the suppressor variable to the model enhances the relationship between the independent variable and the outcome variable (Lancaster, 1999; Rosenberg, 1973). Suppression implies that the relationship between some independent variables of interest and the outcome variables are blurred because of outcome-irrelevant variance; the addition of suppressor variables clears or, "purifies" (Pedhazur, 1997, p. 186) the outcome-irrelevant variation from the independent variables, thus revealing the true relationship between the independent and outcome variables.

Our example using world data illustrates that the regression weight may change substantially when potential suppressor variables are included in models. If the regression weights of included variables improve dramatically due to the presence of a variable that was insignificant at the bivariate level, then one or more of the independent variables may be acting as a suppressor. In our example, the presence of *population per physician in 1980* improved the regression weight of *death rate of a country in 1978*. When *population per physician in 1980* (i.e., the suppressor variable) was partialed out from *death rate of a country in 1978* (i.e., the independent variable) the effect of *death rate of a country in 1978* on the *external public debt in 1980 as a percentage of GNP* (i.e., the dependent variable) was enhanced. The effect was enhanced because the suppressor variable had purified (or cleared) the relationship between the dependent and the independent variables by accounting for the outcome-irrelevant variation in the *death rate of a country in 1978*. Therefore, *population per physician in 1980* was a classic suppressor. In addition, the presence of the *death rate of a country in 1978* improved the regression weight of the *population per physician in 1980*. Thus, both independent variables acted as suppressor variables by removing the other's irrelevant variation; this action was not surprising considering that these two variables had a high zero-order correlation (r =.70).

Second, the most important benefit of controlling for the suppressor variable is to improve the predictive power of the model (Cohen et al., 2003; Tzelgov & Henik, 1991). The idea of suppression forces the researcher to think of predictor variables in multiple regression equations differently. Predictor variables may follow three patterns: (a) predictor variables may account for only the variance in the outcome variable and have zero or negligible correlation with other predictors; (b) predictor variables may clear out only the outcome-irrelevant variance from other predictors and have zero correlation with the dependent variable (i.e., classic suppression); or (c) predictor variables may explain some variance in the dependent variable as well as clear out outcome-irrelevant variance from other predictors, that is, relative suppression (Tzelgov & Henik, 1991). Thus, when a model includes a suppressor variable, total explained variance will increase even when it is not significantly related with the outcome variable.

In the world data example, R^2 changed dramatically because the model included a suppressor variable that was not correlated with the dependent variable at the bivariate level. In the model with two predictor variables and a dependent variable, the zero-order correlation between *population per physician in 1980* and *external public debt in 1980 as a percent of GNP* was zero. However, when both *death rate of a country in 1978* (independent) and *population per physician in 1980* (suppressor) were included, the R^2 was significantly higher than when the model included only *death rate of a country in 1978*. In other words, R^2 increased from 6.69% to 13.19% by adding a variable in the model that was not related with the dependent variable at the bivariate level. The model would have suffered if the variable for *population per physician in 1980* had been eliminated after examining the bivariate results.

Change in sum of squares (hierarchical and simultaneous) upon adding an insignificant predictor variable is also a good indication that the added variable may be a suppressor. Sometimes, when a suppressor variable is present, the Type III sum of squares (simultaneous) may add up to more than the model sum of squares. Such was the case in our world data example. The sum of Type III sum of squares was more than the model sum of squares (see Table 2, Model 3). In the absence of suppressor variables, Type III sum of squares will not be higher than the model sum of squares. Indeed, most of the time, Type III sum of squares will be less than the model sum of squares because none of the independent variables get credit for the shared variance or the variance in the dependent variable explained jointly.

Third, understanding of suppression relations contributes to theory building (Tzelgov & Henik, 1991). When an independent variable lacks a significant association with a dependent variable or an independent variable does not have the expected sign, researchers are often tempted to discard such an independent variable from further analysis (Gelman & Hill, 2007). Horst (1941) criticized this procedure, saying, "it assumes that the best weight to give a predictive variable depends only on the association of that variable with the criterion [or outcome variable]. In general, however, the best weight to give a variable depends on what other predictive variables are included in the set." (p. 67). As most regression models involve more than two independent variables, the prevalence of relative suppressor variables in social work research may be greater than what is currently recognized in the field.

A closer examination of previous literature, theory, and regression coefficients may help to identify and document these variables. Sometimes knowledge of a particular variable acting as a suppressor, or "irrelevant-variance cleaner" may be evident while testing a multivariate relationship (Tzelgov & Henik, 1991) Such situations often yield unexpected results and provide authors with an opportunity to present meaningful theoretical interpretation of the results in light of new information (see, e.g., Hannah & Morrissey, 1987; MacNeill et al., 2000; Paulhus et al., 2004).

Moreover, in our world data example, contrary to our hypothesis, the suppressor variable (*population per physician in 1980*) received a negative weight in multiple regression. When this situation occurs, researchers often find it difficult to explain the results and tend to regard such variables "with suspicion and distrust" (Horst, 1941, p. 435). Thus, a strong temptation exists to eliminate these variables and retain a model that is in line with the researcher's theory and hypothesis (Gelman & Hill, 2007; McFatter, 1979). In reality, such results clearly imply the presence of a suppressor variable, and the negative sign is a result of the suppressor's low correlation with the outcome variable and high correlation with other predictor variables at the bivariate level. Thanks to the suppression effect of *population per physician in 1980*, the relationship between the *death rate of a country in 1978* and *external public debt in 1980* was enhanced and the predictive power of the model improved substantively. The negative sign of the regression weight of *population per physician* on *external public debt* reflects that those countries with very high value on the suppressor have a lowered predicted value because of multiplying a negative regression coefficient by a positive score. Whereas, predicted scores of those below the mean on the suppressor variable are increased because of multiplying a negative regression coefficient by a negative score (Pedhazur, 1997). Readers that are interested in the underlying mathematical equations that explain this phenomenon should refer to Conger (1974) and Darlington (1968).

Conclusion

Our goal in this paper was to alert readers to the presence of suppressor variables in social work research and to create awareness that suppressor variables in social work research are more prevalent than previously recognized. The idea that a variable, which is unrelated to the dependent variable, should be retained not only for theoretical purposes but also to improve overall predictive power of the model is appealing. Horst (1941) has recommended that researchers should retain a variable, even if it has negligible correlation with the dependent variable and has a significant correlation with other predictor variables. Further, other benefits accrue from including suppressor variables in multiple regression models. Including a suppressor variable will eliminate "the danger of rejecting a true hypothesis as false" (Rosenberg, 1973, p. 369). As shown in this research, suppressor variables enrich the results of a multiple regression model, whereas premature elimination of suppressor variables reduces the predictive power of a model. Ideally, including suppressor variables in a model should be theory based and every regression model should include using a test for suppressor effects. This approach allows researchers to become ware of the suppressor effect of a particular variable and to be better able to explain when regression results change drastically from one model to another.

We have shown that it is possible to increase the predictive power of a model by including a variable that was uncorrelated (or weakly correlated) with dependent variable, as long as the variable correlated with other independent variable(s) that correlated with the dependent variable. Given this discussion of suppression effects, we suggest that researchers retain their list of independent variables, even if those variables are not significantly related with the dependent variable at the bivariate level, until they examine their multiple regression results for any evidence of suppression effects.

It is important that researchers conduct a systematic search for suppressor variables, recognize, and use such variables whenever the opportunity presents itself in social work research. Social work researchers can check for the presence of a suppressor variable by examining the affect of variables on the four key statistics discussed in this article. At a minimum, a quick comparison of standardized coefficients before and after the addition of suspected suppressor variable in the model should be sufficient to test for the presence of a suppressor effect. If the strength of relationship between an independent variable and the outcome variable improves after the addition of the third variable, then the researcher can be assured of the suppression effect.

References

Bertrand, P. V., & Holder, R. L. (1988). A quirk in multiple regression: The whole regression can be greater than the sum of its parts. *The Statistician, 37*, 371-374. doi:10.2307/2348761

Caputo, R. K. (2000). Race and marital history as correlated of women's access to family-friendly employee benefits. Journal of Family and Economic Issues, 21, 365-385. doi:10.1023/A:1026480621340

Conger, A. J. (1974). A revised definition for suppressor variables: A guide to their identification and interpretation. *Educational and Psychological Measurement, 34*, 35-46. doi:10.1177/001316447403400105

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

Courville, T., & Thompson, B. (2001). Use of structure coefficients in published multiple regression articles: β is not enough. *Educational & Psychological Measurement, 61*, 229-248.

Daniel, V. (1996). The relation between mathematical achievement and gender differences in spatial abilities: A suppression effect. *Journal of Educational Psychology, 88*, 563-571. doi:10.1037/0022-0663.88.3.563

Darlington, R. B. (1968). Multiple regression in psychological research and practice. *Psychological Bulletin, 69*, 161-182. doi:10.1037/h0025471

Fox, J. (1991). *Regression diagnostics*. Beverly Hills, CA: Sage.

Gelman, A., & Hill, Jennifer. (2007). *Data analysis using regression and multilevel/hierarchical models.* New York, NY: Cambridge University Press.

Hannah, T. E., & Morrissey, C. (1987). Correlates of psychological hardiness in Canadian adolescents. *Journal of Social Psychology, 127*, 339-344.

Horst, P. (1941). The prediction of personal adjustment. *Social Science Research Council Bulletin*, *48*. New York, NY: Social Science Research Council.

Korn, M. E., & Maggs, J. L. (2004). Why drink less? Diffidence, self-presentation styles, and alcohol use among university students. *Journal of Youth and Adolescence*, *33*, 201-211. doi:10.1023/B:JOYO.0000025319.57979.13

Lancaster, B. P. (1999, January 21). *Defining and interpreting suppressor effects: Advantages and limitations.* Paper presented at the Annual Meeting of the Southwest Educational Research Association, San Antonio, TX. Retrieved from http://eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/17/2c/59.pdf

Lubin, A. (1957). Some formulas for use with suppressor variables. *Educational and Psychological Measurement. 17*, 286-296. doi:10.1177/001316445701700209

Lewis-Beck, M. (1980). *Applied regression: An introduction*. Beverly Hills, CA: Sage .

Loukas, A. Paulos, S. K., & Robinson, S. (2005). Early adolescent social and overt aggression: Examining the roles of social anxiety and maternal psychological control. *Journal of Youth and Adolescence, 34,* 335-345. doi:10.1007/s10964-005-5757-2

Lutz, J. G. (1983). A method for constructing data which illustrate three types of suppressor variables. *Educational and Psychological Measurement. 43*, 373-377. doi:10.1177/001316448304300206

Maassen, G. H., & Bakker, A. B. (2001). Suppressor variables in path models: Definitions and interpretations, *Sociological Methods & Research, 30,* 241-270. doi:10.1177/0049124101030002004

MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prevention Science, 1*, 173-181. doi:10.1023/A:1026595011371

MacNeill, S. E., Lichtenberg, P. A., & LaBuda, J. (2000). Factors affecting return to living alone after medical rehabilitation: A cross-validation study. *Rehabilitation Psychology*, *45*, 356-364. doi:10.1037/0090-5550.45.4.356

Martz, E. (2003). Invisibility of disability and work experience as predictors of employment among community college students with disabilities. *Journal of Vocational Rehabilitation*, *18*, 153-161. http://iospress.metapress.com/content/q4eam25lqhk1wymn/

McFatter, R. M. (1979). The use of structural equation models in interpreting regression equations including suppressor and enhancer variables. *Applied Psychological Measurement*, *3*, 123-135. doi:10.1177/014662167900300113

McNemar, Q. (1945). The mode of operation of suppressant variables. *American Journal of Psychology*, *58*, 554-555. doi:10.2307/1417771

Morrow-Howell, N. (1994). The M word: Multicollinearity in multiple regression. Research note. *Social Work Research, 18*, 247-251.

Meehl, P. E. (1945). A simple algebraic development of Horst's suppressor variables. *American Journal of Psychology*, *58*, 550-554. doi:10.2307/1417770

Mendershausen, H. (1939). Clearing variates in confluence analysis. *Journal of the American Statistical Association*, *34*, 93-105. doi:10.2307/2279167

Onwuegbuzie, A. J. (2001). Statistics anxiety and the role of self-perceptions. *Journal of Educational Research*, *93* (5), 323-330. doi:10.1080/00220670009598724

Paulhus, D. L., Robins, R.W., Trzesniewski, K. H., & Tracy, J. L. (2004). Two replicable suppressor situations in personality research. *Multivariate Behavioral Research, 39*, 303-328. doi:10.1207/s15327906mbr3902_7

Pedhazur, E. J. (1997). *Multiple regression in behavioral research: An explanation and prediction.* New York, NY: Holt, Rinehart & Winston.

Rosenberg, M. (1973). The logical status of suppressor variables. *Public Opinion Quarterly, 37*, 359-372. doi:10.1086/268098

Shieh, G. (2006). Suppression situations in multiple linear regression. *Educational and Psychological Measurement*, *66*, 435-447. doi:10.1177/0013164405278584

Smith, R. L., Ager, J. W., & Williams, D. L. (1992). Suppressor variables in multiple regression/correlation. *Educational and Psychological Measurement, 52*, 17-29. doi:10.1177/001316449205200102

Tabachnick, B. G., & Fidell, L. A. (2007). *Using multivariate statistics*. New York, NY: Pearson Education.

Tzelgov, J., & Henik, A. (1991). Suppression situations in psychological research: Definitions, implications, and applications. *Psychological Bulletin*, *109*, 524-536. doi:10.1037/0033-2909.109.3.524

Velicer, W. F. (1978). Suppressor variables and the semipartial correlation coefficient. *Educational and Psychological Measurement*, *38*, 953-958. doi:10.1177/001316447803800415

Voyer, D. (1996). The relation between mathematical achievement and gender differences in spatial abilities: A suppression effect. *Journal of*

Educational Psychology, *88*, 563-571. doi:10.1037/0022-0663.88.3.563

Walker, D. A. (2003). Suppressor variable(s) importance within a regression model: An example of salary compression from career services. *Journal of College Student Development, 44*, 127-133. doi:10.1353/csd.2003.0010