Potential Problems in the Statistical Control of Variables in Organizational Research: A Qualitative Analysis With Recommendations

THOMAS E. BECKER
University of Delaware

The author examines statistical control in a random sample of 60 articles published in four top journals during 2000 to 2002. Authors' bases for including control variables, clarity regarding measures and methods, and reporting of results were recorded. Potential problems included a lack of explanations for inclusion, unclear descriptions of measures and methods, incomplete reporting, and other flaws. Implications for interpreting results, replication, future reviews, and effect sizes are discussed. Twelve recommendations for addressing these issues are offered.

Keywords: *statistical control; control variables; nuisance variables; quantitative methods*

Control variables are factors that researchers include in their work to rule out alternative explanations for their findings (Schmitt & Klimoski, 1991) or to reduce error terms and increase statistical power (e.g., Schwab, 1999). There are two primary means of controlling variables in a study. The first is control by experimental design, whereby the researcher manipulates the nature of the sample or environment so that it is identical across participants (Keppel, 1991). For example, to control for gender effects, an experimenter might include only women in his or her study. The other is statistical control, whereby the researcher measures relevant variables and includes them in the primary analyses (Neter, Kutner, Nachtsheim, & Wasserman, 1996). For example, to control for gender effects, a researcher might dummy code gender (e.g., 0 = male, 1 = female) and then include the dummy-coded variable in regression

Author's Note: This project was supported by a summer research grant provided by the College of Business and Economics at the University of Delaware. I thank Karl Aquino, Daniel Freeman, Robert Kent, Mary Kernan, John Kmetz, Yasemin Kor, and Melissa Stacey for their help in coding articles to assess reliability. I also thank John Sawyer, Neal Schmitt, Daniel Sullivan, and Robert Vance for their comments on earlier drafts. Finally, I thank Larry Williams and two anonymous reviewers for their suggestions during the review process. Correspondence concerning this article should be sent to Thomas E. Becker, University of Delaware, Department of Business Administration, Newark, DE 19716-2710; e-mail: beckert@lerner.udel.edu.

Organizational Research Methods, Vol. 8 No. 3, July 2005 274-289
DOI: 10.1177/1094428105278021

analyses. This tactic mathematically partials the effect of gender from the other variables included in the analyses.

This article examines the use of statistical control in four major empirical journals that publish organizational research. The focus is on statistical rather than experimental control because statistical control is more common in this field and because reading these journals has led me to believe there may be problems in how statistical control is implemented. Reducing or eliminating such problems would benefit the field by justifying greater confidence in the conclusions generated by the analyses of research data. Because the concern is most relevant to nonexperimental studies, my emphasis is on the use of control variables for ruling out alternative explanations (i.e., addressing issues of internal validity) rather than simply reducing error terms. In sum, my goals are to analyze the treatment of control variables in recent organizational research and to evaluate how effectively issues of control have been addressed.

A central premise driving this study is that control variables are as important as independent and dependent variables (i.e., predictors and criteria). There are three reasons for this belief. First, the results of any mathematical analysis of data are directly driven by the variables included in the analysis, and improperly including control variables can produce misleading findings. For example, in the context of job stress research, Spector, Zapf, Chen, and Frese (2000) discussed the problems that can occur when negative affectivity is routinely treated as a bias factor to be statistically controlled. These authors pointed out that if, as some evidence suggests, negative affectivity plays a substantive role in stress-related phenomena, controlling for negative affectivity can lead to removing the effects of the variables (e.g., work conditions) that one wants to study. Just as with any predictor, decisions regarding which controls to include affect the significance levels and estimated effect sizes of the other variables. Indeed, unless there is absolutely no correlation between a control and the predictors and criteria, this must to some degree be the case.

Second, without knowledge of which variables were controlled, how they were measured, and the specific methods of control, effective replications of studies are not feasible. This is important because the replication and extension of findings across multiple settings, times, and samples of people is an essential part of ensuring that empirical results are generalizable (Sackett & Larson, 1990). Finally, one researcher's control variable is another's independent or dependent variable—or mediator or moderator. Thus, one researcher may look at the effect of X on Y controlling for Z. A second may be interested in the relationship between Y and Z and be interested in the first researcher's work for that reason. However, if in the first study, the statistics pertaining to Z are not reported or are insufficiently reported, the findings would be useless to the second researcher. This would also be true for a third researcher doing a qualitative or meta-analytic review of X-Y, X-Z, or Y-Z relationships. For instance, without knowledge of what controls were included in studies of the X-Y link, a reviewer could examine only simple bivariate relationships, leaving unknown the contexts in which these relationships exist. Of course, without adequate reporting of the control variable, Z, a study could not be included in reviews of X-Z and Y-Z, or, if it was included, the results could be misleading.

Reading the top journals in management and industrial-organizational psychology leads me to believe that the reasons for including control variables are often neglected, that authors' choices of measures are sometimes unexplained, that indices of reliability and validity for controls are often not provided (where such indices would be relevant), and that the reporting of control variable results is frequently incomplete. However, beyond these anecdotal data, evidence is sparse. Because this is an exercise in practical concept identification rather than theory testing, I offer no hypotheses about the kinds of problems associated with the use of control variables in the field or the ways they can best be resolved. Rather, I begin simply with this question: What kinds of problems, if any, exist with regard to the statistical control of variables in top journals in organizational research? The following study is intended to supply a concrete, inductive response to this question.

Method

Selection of Articles

The data for the study were gathered from a stratified random sample of 60 studies published in the *Academy of Management Journal*, *Journal of Applied Psychology*, *Administrative Science Quarterly*, and *Personnel Psychology* during the period 2000 to 2002. Articles were selected by using a random numbers table to choose articles from journal volumes, with the restriction that 5 articles be selected from each journal each year, for a total of 15 articles per journal. I chose the period of the study so as to include recent literature within the analysis. Regarding the number of articles, although the sample size of 60 is somewhat arbitrary, I believed this number would be large enough for the limited quantitative analyses I planned on conducting yet not so large as to make the qualitative analysis prohibitive.

Analyzing and Coding Articles

Primary coding. The data were analyzed largely qualitatively using the techniques of open and axial coding (Miles & Huberman, 1994; Strauss & Corbin, 1998). Open coding is the process of identifying central concepts or categories and their properties, and axial coding involves relating categories to their subcategories. The major aim of coding was to identify categories and subcategories representing problems in the treatment of control variables and examples of positive treatment of control variables. In completing this task, I distinguished between conceptual control variables (CCVs), that is, the latent variables that authors were trying to control, and measured control variables (MCVs), that is, the variables as they were operationalized within a particular study. For instance, organizational size might be a CCV, and the number of employees might be the corresponding MCV. Distinguishing between conceptual and measured variables was important because some issues pertain more to one than the other. For example, reasons for why a control variable is included in a study is an issue relevant to CCVs whereas the reporting of descriptive statistics pertains to MCVs. Furthermore, the number of conceptual and measured variables often differs within a study, and I planned a few quantitative analyses of each.

To accomplish the goal of the study, each article had to first be coded. This involved reading and carefully studying each article and then coding it using 10 previously developed assessment dimensions. These dimensions were classes for evaluating articles and were created by reading and open coding several articles published prior to 2000 in the same four journals identified above. The dimensions addressed issues of inclusion of control variables, methods of measurement and control, reporting of descriptive statistics, relationships with other variables, and results in primary analyses. More specifically, they were the following:

1. Basis for inclusion: the extent to which the author(s) explained why the particular CCVs were included as controls.
2. Clarity, method of measurement: the extent to which methods of measuring CCVs was clear.
3. Basis for method of measurement: the extent to which a reasonable explanation was given for measuring the CCVs a certain way.
4. Clarity, method of control: the extent to which the author provided a clear description of how the MCVs were controlled.
5. Basis for method of control: the extent to which a reasonable explanation was given for controlling the MCVs in a certain way.
6. Basis for differential inclusion in analyses: the extent to which a reasonable explanation was given for including different MCVs in different analyses.
7. Reporting, descriptive statistics: the extent to which meaningful descriptive statistics (means and standard deviations for continuous MCVs; percentages within categories for categorical MCVs) were provided for the control variables.
8. Reporting, individual relationships between continuous MCVs and other variables: the extent to which descriptions were given regarding the direction and magnitude of relationships between continuous MCVs and the independent and dependent variables.
9. Reporting, individual relationships between categorical MCVs with other variables: the extent to which descriptions were given regarding the direction and magnitude of relationship between categorical MCVs and the independent and dependent variables.
10. Reporting, primary results: the extent to which results, such as significance levels and effect sizes, for MCVs were reported.

To evaluate the articles, I created a primary analysis form including the 10 assessment dimensions, along with an open-ended section for other information unique to a given article. A copy of the form, completed for a sample, randomly chosen article, is available on request. The basic categories for coding the articles were provided by the assessment dimensions, but subcategories were identified by axial coding of the completed analysis forms. This involved an iterative process whereby a potential problem or positive example in the treatment of control variables was placed into a category, and then as further problems or examples of a similar and different type were put in the category, a subcategory became apparent. For example, one article might provide no explanation for including a control variable in the study, so this would be placed in the "basis of inclusion" category. Additional articles might have the same characteristic (no explanation for inclusion), whereas others may offer an explanation, but a very unclear one, or offer a reasonably clear explanation but no evidence. Each of these cases can be reasonably put in the "basis of inclusion" category but fall into separate subcategories: "unclear explanation," "no evidence," and so on. The final categories and subcategories, and the placement of articles within them, constitute the main results of this study.

Types of studies. Although describing all the types of studies contained within the four journals was not a major goal of this project, I coded for two of the most common kinds: correlational field studies and archival studies. Study type was coded as 1 = correlational field, 2 = archival, and 3 = other (e.g., experiments, quasi-experiments).

Table 1
Interrater Agreement Regarding the Treatment of Control Variables for 20 Articles

Treatment Issue	T
Did the author(s) explain why all the control variables were included in the study?	.73
Did the authors describe how all of the control variables were measured?	.80
Were correlations for all [continuous and dichotomous] control variables reported in the correlation matrix?	.78
Were all the control variables contained in the table(s) and/or figure(s) of primary analyses?	.72
Is an effect size provided for each control variable in the table(s) and/or figure(s)?	.78

Note. *T* is the Tinsley and Weiss (1975) index of rater agreement.

Consistency check. Given the knowledge of research methods and statistics required by the task and the large amount of time required to code each article, I did the bulk of the coding myself. As a check on the consistency of coding for the major coding tasks, six faculty, one graduate student, and I independently coded 20 randomly selected articles. Six of the coders rated excerpts from three articles and one rated two articles. I provided my ratings on all 20. Because the purpose of this process was to estimate the consistency of coding for the 60 articles, in all cases I served as Coder 1 and one of the other seven coders served as Coder 2. To provide ratings, we used a coding form addressing issues of inclusion, method of measurement, descriptive statistics, and primary analyses. Items were answered “yes” or “no.”

To assess consistency, I used the Tinsley and Weiss (1975) index of interrater agreement. I chose a measure of agreement rather than reliability because I was concerned with the level of agreement possible among raters rather than the general correlation across ratings. The formula for the index is $T = (N_1 - NP)/(N - NP)$, where N_1 is the number of agreements between raters, N is the number of cases rated, and P is the probability that raters would agree by chance. Table 1 reports the specific questions addressed and the values of interrater agreement for the consistency check. Each value exceeds .70, indicating that independent coders agreed to a reasonable extent on how control variables were handled in the 20 randomly selected articles.

Results

Positive Treatment of Control Variables

Through coding the positive cases, I identified subcategories representing hallmarks of effectively handling control variables. Each of these is briefly discussed below.

Inclusion. With respect to explanations for including controls, positive cases offer rational explanations, citations, statistical/empirical results, or some combination of these. For instance, in justifying why they included job experience as a CCV, Van Scotter, Motowidlo, and Cross (2000) provided all three: a rational explanation based on empirical results with relevant citations:

> The relationships between experience, task performance, and contextual performance are relevant for three reasons. First, research shows that job experience

> explains considerable variance in task performance. . . . Second, differences in the way that task performance and contextual performance are related to experience provide additional support for distinguishing between the two kinds of behavior. . . . Third, the mean correlation between experience and contextual performance (mean $r = .17$) in these studies suggests that the relationship between experience and contextual performance should not be ignored. Therefore, Schmidt et al.'s (1988) warning that failing to control for differences in job experience could make it impossible to detect relationships between job performance and other variables seems relevant for this study. (p. 527)

Method of measurement. Good reasons for measuring CCVs in a certain manner were predicated on prior use, rational explanations, psychometric justifications, and face-valid measures. The latter subcategory involved cases in which the reasons for measuring a control in a given way was evident, for example, assessing employee age as current year minus year of birth or membership within departments as a set of effects-coded variables representing departments within a company. Researchers relying on prior use followed convention in measuring CCVs. An example is Dobrev, Kim, and Carroll's (2002) measure of organizational tenure, used because "we followed convention in modeling a firm's tenure in a particular organizational population rather than its organizational age" (p. 246). Citations were provided for the tenure measure. An example of rational explanation for this category is Aulakh, Kotabe, and Teegen's (2000) justification for their measure of industries: "Given that standard industry classifications were not available through secondary sources and that classification systems vary across countries, we asked the respondent to list the primary industries of their export products" (p. 352). Finally, authors providing a psychometric justification for their CCV measures relied on the same sort of evidence used to justify measures of independent and dependent variables, that is, results from factor analyses and other evidence for reliability and validity (e.g., Lam & Schaubroeck, 2000; Ruderman, Ohlott, Panzer, & King, 2002; Sherony & Green, 2002).

Method of control and differential inclusion. Rational explanation was a subcategory characterizing positive instances of the clarity and basis for method of control and the basis for differential inclusion. Regarding methods of control, authors in this class offered clear explanations for how MCVs were controlled in less common methods such as panel probit analyses (Ahmadjiian & Robinson, 2001), incorporation of measured controls into dependent variables (Ellstrand, Tihanyi, & Johnson, 2002), and confirmatory factor analysis (Sturman & Short, 2000). An illustration of a reasoned explanation for differential inclusion is Haveman and Nonnemaker's (2000) study of savings and loan associations. They controlled for geographic distribution of branch offices and across-market size of rivals only in the analyses of multimarket firms. As the authors pointed out, single-market firms cannot have a geographic distribution across markets or compete with rivals in different markets. In analyses of single-market firms, the authors controlled for the number of rivals in a given market. The subcategory, statistical/empirical results, contains articles offering observational justifications for differential inclusion. An example is Garonzik, Brockner, and Siegel's (2000) decision to drop two MCVs from a second study that they had included in the first: "Because two of the control variables in Study 1—previous overseas experience and predeparture training—were unrelated to the dependent variable in both correlation analyses and multiple regression, they were deleted from Study 2" (p. 17).

Reporting. There were no subcategories for reporting basic distributional information and intercorrelations for control variables. Positive cases of the former simply included the provision of means and standard deviations for continuous MCVs and percentages within categories for categorical MCVs. Many authors also reported correlations between the MCVs and other variables. One might think this would be unlikely in a study with a large number of controls, but this was not the case. For example, Li and Rowley (2002) reported correlations for all 17 of their MCVs, and Palmer and Barber (2001) reported correlations for 20. Positive cases of reporting control variable results in primary analyses included the subcategories of providing significance levels and effect sizes for sets of MCVs (e.g., Colvin, Batt, & Katz, 2001; Rotundo & Sackett, 2002) and providing significance levels and effect sizes for individual MCVs (e.g., Brown, 2001; Rau & Hyland, 2002).

Potential Problems With the Treatment of Control Variables

The above results document cases in which control variables were handled well. The following findings indicate problems in how controls are treated in organizational research. Because these problematic cases are the focus of later discussion, the percentages of articles falling within the subcategories are reported in Table 2.

Inclusion. In more than half the articles (18.3% + 33.3% = 51.6%), no explanation was provided for including one or more CCVs. In addition, as shown in Table 2, there were many cases involving unclear or incomplete explanations. Moreover, in more than two thirds of the articles (35% + 33.3% = 68.3%), no evidence or citations were provided for at least one CCV. For example, in one article, the authors suggested that "gender . . . may affect how people think about teams and the language they use to describe them." Although this may or may not be true, no rationale or evidence is given for why gender may have these effects, and no citations of supportive prior work are supplied.

Method of measurement. As reported in Table 2, in 18.3% of the articles, measures of one or more of the CCVs were not described, and 45% of the cases involved unclear descriptions of one or more measures. As an illustration of the latter subcategory, authors of one article stated that family income was measured categorically, but not described were how many categories were used and what they were. In half the cases, no explanation was given for why one or more controls were measured as they were (when some explanation seemed to be needed). For example, in one article, nursing home size was measured as the natural log of the number of beds the home operated at the start of each year. One might ask why this measure is preferable to, say, the number of patients or employees and why the number of beds operated at the beginning of the year—rather than, say, the end of the year or throughout the year—is better. There may certainly be good answers to these questions, but if so, they were not supplied in the article. Finally, a number of articles offered unclear explanations of the measures of CCVs.

Method of control and differential inclusion. There were a few cases involving missing or unclear identification of the method of control. For instance, in one study, how the MCVs were included in structural equation modeling was not described, and

Table 2
Percentage of Problematic Articles in Each Subcategory

Category/Subcategory	*Percentage*
Basis for inclusion	
No explanations for any control variables	18.3
No explanation for at least one control variable	33.3
Unclear explanation(s)	63.3
No cites/evidence	35.0
No cites/evidence for at least one control	33.3
Clarity and basis for method of measurement	
One or more measures not described	18.3
Unclear description of one or more measures	45.0
No explanation for why one or more controls were measured as they were (for nontrivial cases)	50.0
Unclear explanation for one or more controls	23.3
Clarity and basis for method of control	
Missing or unclear identification of method	8.3
Missing or insufficient explanation for choosing method	28.3
Basis for differential inclusion or treatment	
Unclear explanation for differential inclusion	25.0
Unclear explanation for differential treatment	8.3
Reporting: descriptive statistics	
No descriptive statistics for one or more continuous controls	21.7
Missing a common descriptive statistic for one or more continuous controls	11.7
No descriptive statistics for one or more categorical controls	18.3
No index of measurement adequacy (e.g., reliability, validity, accuracy) where such seems needed	46.7
Reporting: relationships between controls and other variables	
None provided for any continuous controls	20.0
Incomplete reporting for continuous controls	33.3
None provided for any categorical controls	10.0
Incomplete reporting for categorical controls	30.0
Reporting: primary analyses	
Effect size given for set but not individually	5.0
Significance but no effect size for at least one control	3.3
No significance or effect size for at least one control	15.0
No significance or effect sizes for any controls	16.7

Note. Numbers do not sum to 100% within categories because some articles fit multiple subcategories and others fit no subcategory.

in another investigation, supplemental analyses were run on a subset of MCVs, but the kinds of analyses were not identified. In a larger number of cases (28.3%), although the methods of control were reasonably clear, the basis for controlling the MCVs in a certain manner was missing or insufficient. In one structural equation analysis, for instance, the authors allowed paths from seven controls to problem drinking (a central endogenous variable). However, unlike the typical case in regression, they did not allow correlations with any predictors. The reasons for this were not explained.

With respect to differential inclusion, authors of one quarter of the articles provided unclear explanations. For example, demographic MCVs were included in regressions of leader outcomes on the Big 5 and transformational leadership. However, these MCVs were not included in an analysis of the effect of the Big 5 on transformational leadership. No reason was given. In another article reporting the results of three studies, organizational size was included in Study 1 but not Studies 2 and 3. This seemed odd because in Study 1, size was correlated with the dependent variables. Furthermore, in Study 3, employee tenure was included in one sample but not a second. The authors said tenure was unavailable from company records for Sample 2, but it is unclear why they did not ask for these data on the questionnaire as they did with the first sample. Finally, in a few cases (8.3%), the reasons for differential treatment of MCVs were unclear. For instance, in a study involving 19 MCVs, one control (competition) was included as part of an interaction term with a key independent variable, without explanation. No other MCVs were similarly treated. In another study using hierarchical regression, two MCVs were entered at Step 1 whereas the third, occupational role commitment, was entered in a separate step. No explanation was provided.

Reporting. The categories pertaining to the reporting of MCVs (descriptive statistics, relationships with other variables, and reporting in the primary analyses) involve missing or incomplete results. The subcategories pertain to how much of the results are missing (some or all) and what kind (measures of central tendency, dispersion; reliability and validity; significance levels and effects sizes). As is evident in Table 2, with regard to descriptive statistics, the majority of articles reported basic distributional measures for MCVs. Only 10% to 20% did not report such summaries. On the other hand, 46.7% of the publications contained no index of measurement adequacy (e.g., reliability, validity) in those cases in which such information appeared relevant, that is, when the MCVs were not simply demographic variables or others whose measurement seemed more or less objective, simple, and, at least in theory, easily verifiable. Regarding relationships with the independent and dependent variables, although authors typically reported correlations among continuous MCVs, failure to report occurred in 20% of the cases. Furthermore, in about one third of the cases, the reporting was incomplete. The situation was similar if not as pronounced for the categorical variables (see Table 2).

With regard to reporting the MCV results in the primary analyses, incomplete reporting of significance or effect size occurred in about one third of the cases, with no significance or effect sizes reported for any controls in 16.7% of the cases and none for at least one MCV in 15%. In a few cases, effect sizes were given for a set of MCVs but not individually (5%), and in a few others, significance level was given without an effect size (3.3%).

Ancillary Analyses

To identify variables related to the number of controls included in a study, I conducted several quantitative analyses. The reason for these analyses is that problems with control variables may become more common as the number of controls increases. Table 3 reports the mean and median number of CCVs and MCVs for each journal and year. The number of MCVs exceeds CCVs because there were many cases in which a set of variables was used to capture a CCV. For instance, in controlling for job category (a CCV), an author might effects code a set of four categories (clerical, production, technical, managerial) and then include the set in a hierarchical regression. The most extreme case of this was an article in which rater effects (the CCV) were controlled by entering 504 dummy-coded variables representing individual raters.

Table 3
Number of Conceptual and Measured Control Variables by Year

	AMJ		ASQ		JAP		PP	
Year	*CCVs*	*MCVs*	*CCVs*	*MCVs*	*CCVs*	*MCVs*	*CCVs*	*MCVs*
2000								
M	4.0	6.6	16.0	36.0	6.4	7.0	3.2	3.6
SD	2.1	1.5	11.6	34.5	3.8	4.5	2.9	1.8
Median	4.0	6.0	10.0	20.0	8.0	8.0	3.0	3.0
2001								
M	8.2	13.2	9.4	15.8	4.0	5.2	3.6	5.0
SD	3.4	12.5	8.0	9.6	3.5	6.1	2.5	1.3
Median	10.0	10.0	6.0	21.0	2.0	2.0	4.0	4.0
2002								
M	6.8	9.0	6.0	11.6	3.2	4.5	0.8	0.8
SD	2.5	5.3	4.3	10.5	2.2	2.7	0.5	0.5
Median	7.0	7.0	7.0	7.0	4.0	5.0	1.0	1.0

Note. AMJ = Academy of Management Journal; ASQ = Administrative Science Quarterly; JAP = Journal of Applied Psychology; PP = Personnel Psychology. Five observations are contained in each cell, with the exception of one outlier (of 504 MCVs) eliminated from *JAP* 2002.

To determine whether the number of MCVs and CCVs varied by journal or type of study, I conducted two analyses of variance (ANOVAs). In the first analysis, I dropped the article containing 504 MCVs because this observation produced an extreme skew in the distribution of MCVs. The results are shown in Table 4. As can be seen, there are main effects of journal and study type on both MCVs and CCVs. To pinpoint the nature of the effect on MCVs, I conducted comparisons, with Tukey's control for experiment-wise error, of the overall means for the four journals. For each comparison, the error degrees of freedom was 48, error mean square was 165.12, and critical value for the mean difference was 3.76. The number of MCVs in *Administrative Science Quarterly* (*ASQ*; $M = 21.0$, $SD = 22.88$) was greater than that in the *Journal of Applied Psychology* (*JAP*; $M = 5.64$, $SD = 4.52$), $M_{diff} = 15.36$, $p < .05$, and *Personnel Psychology* ($M = 3.13$, $SD = 3.78$), $M_{diff} = 17.87$, $p < .05$. No other differences among journals were significant. To identify the nature of the effect of type of study, I conducted a similar comparison of means. For these comparisons, the error degrees of freedom was 48, error mean square was 165.12, and critical value for the mean differences was 3.42. The number of MCVs in articles reporting results of archival studies ($M = 19.40$, $SD = 19.03$) was greater than that of articles reporting findings from correlational field studies ($M = 8.11$, $SD = 12.62$), $M_{diff} = 11.29$, $p < .05$, or other kinds of studies ($M = 4.31$, $SD = 4.35$), $M_{diff} = 15.09$, $p < .05$. No other differences were significant.

A parallel analysis of the CCVs revealed exactly the same pattern of findings for mean differences. The average article in *ASQ* contained significantly more CCVs than average articles in *JAP* or *Personnel Psychology*, and archival studies contained more CCVs than did correlational field studies or other kinds of investigations.

Finally, to determine if types of studies varied systematically by journal, I conducted two chi-squared analyses. An omnibus test of the number of each kind of study in each journal revealed a significant difference, $\chi^2(6, N = 60) = 20.99$, $p < .05$, with the appearance that the proportion of archival studies in *ASQ* (60%) was greater than that in the *Academy of Management Journal* (33.3%), *JAP* (6.7%), or *Personnel Psychol-*

Table 4
Analysis of Variance for Number of Control Variables

Source	df	F	p
Measured control variables			
Journal	3	5.63	.002
Type	2	5.88	.005
Journal × Type	5	0.00	.99
Error	48	(165.12)	
Conceptual control variables			
Journal	3	6.18	.001
Type	2	7.14	.002
Journal × Type	5	0.00	.99
Error	59	(27.23)	

Note. Values in parentheses represent mean square errors.

ogy (0%). To see if this apparent difference was statistically significant, I dichotomized journals into *ASQ* and others and dichotomized type of studies into archival and nonarchival. The result demonstrated that the proportion of articles with archival studies was greater in *ASQ* than in the proportion collapsed across the other three journals (13.33%), $\chi^2(1, N = 60) = 13.07$, $p < .05$. The phi coefficient for this result, $r_\phi = .47$, indicates a relatively strong effect size.

Discussion

Due to space constraints, in this section I will focus my attention on problems in the treatment of control variables and recommendations for their proper treatment. One serious problem is that without a good explanation for including control variables, consumers of our research cannot adequately assess the credibility of the results. This is because treating a variable as a control when it actually plays a substantive role in the phenomenon of interest (as an antecedent, moderator, or mediator) leads to treating relevant variance as error variance. This, in turn, produces incorrect inferences. Years ago, Meehl (1971) called the erroneous statistical control of variables "the commonest methodological vice in contemporary social science research" (p. 146), and there is reason to believe that this is true today. Spector and colleagues (2000) provided a lucid account of how the indiscriminant statistical control of variables can increase Type II errors by partialling true variance from the relationships of interest. Furthermore, controls can also increase Type I error if they are by chance associated with the predictors but not the criterion. In this case, controls serve as "spurious suppressors," thereby making it appear that the predictors are significantly related to the criterion when, in fact, the relationship is due to a random association among controls and predictors. Therefore, a clear and convincing statement regarding why certain variables are controlled is an essential hallmark of good science.

In addition, without such a statement, future authors seeking to replicate or extend the study may be confused. Should they include the variable even though it was uncorrelated with the dependent variable in this particular study? The answer may be "yes" if there was a good reason for including it but "no" if there was not. If no reason is given, making an informed choice may be difficult. Similar consequences may occur regarding issues of measurement and methods. Furthermore, to the extent readers seek reasonable explanations for research choices, the credibility of the author—and even the journal—may suffer if such explanations are missing.

As suggested in the introduction, the consequences of lack of reporting also affect authors of review articles. Without knowledge of what controls were included in studies of the X-Y relationship or of significance and effect sizes, authors of review articles are reduced to examining simple bivariate relationships: relationships that may not at all represent the contexts in which X-Y relationships exist. Finally, nonreporting of control variable findings hinder any meta-analyses that would have otherwise included the controls. For instance, in a study of the relationship between employee commitment and organizational citizenship behavior, a researcher might control for extraversion and agreeableness but not report the findings for the controls. As a result, later meta-analyses cannot include these findings in the assessment of connections between personality and organizational citizenship behavior. The consequences for the cumulative knowledge of science may be small for one such case but large in the face of systematic underreporting.

In sum, problems in the statistical control of variables can have meaningful consequences for research and practice. The following are my recommendations for authors on how to provide useful information about their control variables. These were developed from the positive examples such as those contained in the Results section and by my interpretation of the potential problems discussed throughout.

Selection of Control Variables

Recommendation 1. Provide at least a brief explanation for why each CCV was selected, including why the variable is a biasing factor rather than a substantive one (per Spector et al., 2000). Wherever possible, also include evidence, or citations that contain evidence, that support including each control in the study.

Recommendation 2. Beware of impotent control variables (i.e., ones uncorrelated with the dependent variable). Unless there is reason to believe that an MCV is a legitimate suppressor, including an MCV that is uncorrelated with the dependent variable in analyses reduces power.

Recommendation 3. Beware the “everything but the kitchen sink” approach. Inclusion of numerous MCVs could be misunderstood as an attempt at methodological legerdemain. To avoid this, make sure there is a logical reason, prior evidence, or both for including each MCV, and do not include impotent MCVs.

Methods of Measurement and Control

Recommendation 4. Clearly and concisely describe how each CCV was measured and why it was measured that way.

Recommendation 5. Whenever a less common method of statistical control is used (e.g., in covariance structure modeling, incorporation of an MCV into the dependent variable), take care to precisely describe the method and why it was used.

Recommendation 6. If certain MCVs were included in some analyses but not others, or some are treated differently than others in the same analyses, provide an explicit rationale for the differences.

Reporting

Recommendation 7. Report standard descriptive statistics (e.g., mean, standard deviation) for all continuous MCVs, including those controlled via incorporation into the dependent variable. In addition, provide summary descriptive statistics for categorical MCVs (e.g., percentage of observations in each category). Wherever possible, supply evidence for reliability and validity of MCVs.

Recommendation 8. Show correlations for all continuous and dichotomous MCVs.

Recommendation 9. For categorical MCVs with more than two levels, provide a summary of the relationships with the other variables, especially the dependent variable. For example, regress the dependent variable on the categorical MCVs and report the R^2 and the betas for the categories.

Recommendation 10. In reporting the primary findings, treat the MCVs just like the independent variables; for example, in regression, include all the MCVs (continuous and categorical) in the table(s) of results, and report their betas and significance levels.

Recommendation 11. Run and report the primary results both with and without the MCVs. If the results do not differ, then authors and readers can rule out the controls as a potential explanation for the findings. If the results differ, this suggests further study of the role of the controls in the phenomenon of interest. In the former case, only the analyses without controls need be reported, along with a sentence such as "Analyses were repeated controlling for [the set of controls], but the results were essentially identical." In cases in which the results differ with and without controls, both sets of analyses should probably be reported and the contrasts discussed.

Interpreting Results

Recommendation 12. The results of a study often depend on what control variables are included in the analyses. Therefore, where possible, follow the lead of several prior authors (e.g., Judge & Bono, 2000) and include controls in hypotheses (e.g., "Controlling for variables A, B, and C, the higher the level of X, the lower the level of Y"). At the very least, interpret and discuss the results vis-à-vis the controls included in a given study. This will normally be more meaningful if Recommendation 11 is followed.

Related Issues

For the above recommendations to be successful, editors and reviewers (as well as authors) must be aware of the issues regarding statistical control. To produce effective action, this heightened awareness needs to be converted into at least two specific policies. First, with respect to control variables, authors should be held to higher standards of explaining, justifying, and reporting. The Results section provides examples of what theses standards look like and illustrates what can occur when such standards are not applied. In addition, the 12 recommendations just presented could serve as the basis for journal expectations with respect to authors' attempts at statistical control. The principle here is simple: Control variables deserve as much attention and respect as do independent and dependent variables.

Second, editors must allow authors a reasonable amount of space to meet the standards. This need not mean a great deal more journal space per article. In many of the positive examples provided, and others not cited, clear, concise explanations were provided for issues of inclusion, measurement, and methods. Often, only a sentence or two was required for each. Furthermore, adequate reporting of controls usually means no more than a few additional lines in a correlation matrix or table of primary findings. However, for more rigorous standards to be applied, editors and reviewers, as with authors, may need a slightly different mind-set: one that places control variables on par with the "variables of interest."

Although I maintain that the above recommendations are relevant to all journals publishing empirical research, the results of the ANOVA of MCVs and CCVs suggest that they may be especially pertinent when control variables are more common, that is, for some journals (*ASQ*) more than others and for some types of studies (archival). A caveat here is that the rather small sample size (15 articles per journal) and correspondingly low power means that differences in addition to those found here, such as significant journal-by-study type interactions or mean differences among journals, may exist but were undetected. Regardless, with a greater number of control variables comes a potentially greater likelihood that one or more of the problems identified in this study will appear. For example, it takes less effort to explain the inclusion of one CCV than to explain a dozen, it is easier to overlook impotent MCVs hidden in a forest of other control variables than when it is standing alone, and it may be more tempting to omit results when one is dealing with a mass of controls rather than a few. Thus, I respectfully suggest that editors and reviewers be especially diligent in monitoring authors' use of statistical control as the number of control variables grows.

Although I hope this study may serve to encourage greater attention to issues of statistical control, there are some technical issues that it cannot definitively answer. For instance, how should control variables be treated in structural equation modeling? A number of authors in my sample freed paths from controls to the endogenous variables but fixed paths to other exogenous variables to zero, without explaining the rationale for doing so. Unless there are clear theoretical reasons for doing otherwise, I suspect the most justifiable course is to follow regression procedures and allow all predictors (including controls) to correlate. In regression, this results in isolating the relationship of a given predictor with the dependent variable, partialling out the effects of the other predictors. Without the same specification in structural equation modeling, the paths representing effects of the exogenous variables would appear to reflect not just their own effects but also those of the other exogenous variables. At any rate, this issue deserves further attention, and authors using structural equation modeling need to explain why they are treating control variables as they are.

Another issue deserving greater attention is the impact of mismanaging control variables on effect sizes. To adequately address this concern, one would need a number of studies with identical independent and dependent variables, measured the same way and under similar conditions. No such studies were included in my sample, but future work on control variables could attempt to find such a sample or to create one (e.g., in a Monte Carlo simulation). It should be noted, however, that the effects of mishandling controls are likely to be complex and dependent on the specific type of problem under consideration. For instance, in the case of inclusion of many impotent controls, uncorrelated with the independent variable, power might be reduced and could lead to a Type II error (i.e., concluding that there is no effect when, in fact, there is). On the other hand, failing to include relevant control variables could inflate the amount of explainable variance in the dependent variable, thereby increasing the chances of a Type I error (i.e., concluding that there is an effect when, in fact, there is not). To complicate matters further, the actual impact of improper inclusion of control variables depends on the causal connection between the controls and the independent and dependent variables (Meehl, 1971; Spector et al., 2000). Still other problems, such as unclear explanations of measures or underreporting results, would not necessarily have any impact on effect sizes. However, as discussed earlier in this section, they would likely have other negative consequences.

In conclusion, my analysis of statistical control in organizational research reveals that there are many things we are doing right and some things we could do better. It is my sincere hope that this study leads us all to take the issue of statistical control a bit more seriously. In the words of one acquaintance, control variables are variables too.

References

Ahmadjian, C. L., & Robinson, P. (2001). Safety in numbers: Downsizing and the deinstitutionalization of permanent employment in Japan. *Administrative Science Quarterly*, *46*, 622-654.

Aulakh, P. S., Kotabe, M., & Teegen, H. (2000). Export strategies and performance of firms from emerging economies: Evidence from Brazil, Chile, and Mexico. *Academy of Management Journal*, *43*, 342-361.

Brown, K. G. (2001). Using computers to deliver training: Which employees learn and why? *Personnel Psychology*, *54*, 271-296.

Colvin, A. J. S., Batt, R., & Katz, H. C. (2001). How high performance human resource practices and workforce unionization affect managerial pay. *Personnel Psychology*, *54*, 903-934.

Dobrev, S. D., Kim, T., & Carroll, G. R. (2002). The evolution of organizational niches: U.S. automobile manufacturers, 1885-1981. *Administrative Science Quarterly*, *47*, 233-264.

Ellstrand, A. E., Tihanyi, L., & Johnson, J. L. (2002). Board structure and international political risk. *Academy of Management Journal*, *45*, 769-777.

Garonzik, R., Brockner, J., & Siegel, P. A. (2000). Identifying international assignees at risk for premature departure: The interactive effect of outcome favorability and procedural fairness. *Journal of Applied Psychology*, *85*, 13-20.

Haveman, H. A., & Nonnemaker, L. (2000). Competition in multiple geographic markets: The impact on growth and market entry. *Administrative Science Quarterly*, *45*, 232-267.

Judge, T. A., & Bono, J. E. (2000). Five-factor model of personality and transformational leadership. *Journal of Applied Psychology*, *85*, 751-765.

Keppel, G. (1991). *Design and analysis: A researcher's handbook* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Lam, S. S. K., & Schaubroeck, J. (2000). The role of locus of control in reactions to being promoted and to being passed over: A quasi-experiment. *Academy of Management Journal*, *43*, 66-78.

Li, S. X., & Rowley, T. J. (2000). Inertia and evaluation mechanisms in interorganizational partner selection: Syndicate formation among U.S. investment banks. *Academy of Management Journal*, *45*, 1104-1119.

Meehl, P. E. (1971). High school yearbooks: A reply to Schwartz. *Journal of Abnormal Psychology*, *77*, 143-148.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis* (2nd ed.). Thousand Oaks, CA: Sage.

Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). *Applied linear statistical models*. Boston: McGraw-Hill.

Palmer, D., & Barber, B. M. (2001). Challengers, elites, and owning families: A social class theory of corporate acquisitions in the 1960s. *Administrative Science Quarterly*, *46*, 87-120.

Rau, B., & Hyland, M. A. (2002). Role conflict and flexible work arrangements: The effects on applicant attraction. *Personnel Psychology*, *55*, 111-136.

Rotundo, M., & Sackett, P. R. (2002). The relative importance of task, citizenship, and counterproductive performance to global ratings of job performance: A policy capturing approach. *Journal of Applied Psychology*, *87*, 66-80.

Ruderman, M. N., Ohlott, P. J., Panzer, K., & King, S. N. (2002). Benefits of multiple roles for managerial women. *Academy of Management Journal*, *45*, 369-386.

Sackett, P. R., & Larson, J. R., Jr. (1990). Research strategies and tactics in industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial & organizational psychology* (Vol. 1, 2nd ed., pp. 420-489). Palo Alto, CA: Consulting Psychologists Press.

Schmitt, N. W., & Klimoski, R. J. (1991). *Research methods in human resources management*. Cincinnati, OH: South-Western.

Schwab, D. P. (1999). *Research methods for organizational studies*. Mahwah, NJ: Lawrence Erlbaum.

Sherony, K. M., & Green, S. G. (2002). Coworker exchange: Relationships between coworkers, leader-member exchange, and work attitudes. *Journal of Applied Psychology*, *87*, 542-548.

Spector, P. E., Zapf, D., Chen, P. Y., & Frese, M. (2000). Why negative affectivity should not be controlled in job stress research: Don't throw out the baby with the bath water. *Journal of Organizational Behavior*, *21*, 79-95.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory.* Thousand Oaks, CA: Sage.

Sturman, M. C., & Short, J. C. (2000). Lump-sum bonus satisfaction: Testing the construct validity of a new pay satisfaction dimension. *Personnel Psychology*, *53*, 673-700.

Tinsley, H. A., & Weiss, D. J. (1975). Interrater reliability and agreement of subjective judgments. *Journal of Counseling Psychology*, *22*, 358-376.

Van Scotter, J. R., Motowidlo, S. J., & Cross, T. C. (2000). Effects of task performance and contextual performance on systemic rewards. *Journal of Applied Psychology*, *85*, 526-535.

Thomas E. Becker is an associate professor of business administration at the University of Delaware, where he conducts research and teaches in the areas of organizational behavior and human resource management. Previous forays into issues of research methods include validation of measures of organizational citizenship behavior and the examination of additive versus multiplicative method effects in applied psychological research.