Journal of Methods and Measurement in the Social Sciences,
Vol. 5, No. 1, 3-21, 2014

The Control Group and Meta-Analysis

John E. Hunter
Michigan State University

Jason L. Jensen
University of North Dakota

Robert Rodgers
University of Kentucky

Social scientists use a mixture of different methodologies, which creates problems for researchers attempting to review the cumulative results of all studies. Standard practice for review studies using meta-analysis is to adjust the findings of all studies that use control groups and to include studies not having control groups without adjustment for extraneous effects, or to not use studies that lack a control group, which could produce an erroneous result. Our study develops a novel meta-analytic procedure that combines the evidence on control group change with evidence on change from the intervention, making it possible to adjust for the effects of extraneous factors in all studies and bridges the gap between control group studies and other types of studies.

Keywords: meta-analysis; control groups in meta-analysis; organizational interventions; meta-analysis methodology

The design, implementation and evaluation of interventions are endeavors of many different academic fields. Interventions are often labeled as social if they take the form of policy or program interventions, so certainly encompass the social sciences, but interventions are also organizational, behavioral and medical. It is safe to say that the designers of interventions eventually will want empirical data on intervention impacts and outcomes. Intervention outcome studies will begin to occur, each based on one particular intervention in one particular setting. Although evaluation research in some fields, such as public policy, has become more standardized, there are still many different methods utilized.

Eventually many studies will have been done related to an intervention. At this point, scientific endeavors of any kind face similar challenges. Since the goal of science is "the cumulation of knowledge from the results of many studies" (Hunter & Schmidt, 1990 p. 13) at some point these individual studies must be aggregated. The challenge is to bridge the micro analysis with the macro. The existence of mixed-methods across the micro studies poses a significant challenge that must be confronted if a researcher is to consider the cumulative evidence on intervention outcomes.

Meta-analysis is a tool that has been developed and used extensively to confront the issue of accumulation of knowledge. Bausell et al. reported in 1995 that there had been 982 meta-analyses between 1980 and 1993 in the social sciences alone. While a newer estimate of usage could not be found, suffice to say that the use of meta-analysis has permeated the social sciences and continues to become more commonplace as a tool of research

synthesis. In this study we will present a solution to one of the problems that meta-analysts have faced related to mixed designs in primary studies. Many different types of primary studies will be available for the meta-analyst, ranging from those with strict experimental designs employing a control group, to those using non-experimental designs, likely lacking a control group. We develop a procedure that allows the meta-analyst to adjust studies without a control group based on studies with a control group. Such an adjustment not only alleviates the impact of extraneous factors (most often called threats to internal validity), but could serve the important function of bridging the gap between diverse research approaches, such as experimental/non experimental, and qualitative/ quantitative. We thus present one solution to resolving a lingering issue that meta-analysts confront: how to integrate a pool of primary studies that utilize multiple methodologies.

Researchers have often criticized case study research on the grounds that it does not add to scientific knowledge because the results are not generalizable. Following Jensen and Rodgers (2001) we feel that the criticism is shortsighted. They develop their argument around the fields of public administration and public policy (organizational and policy interventions), pointing out that case studies of all types, qualitative to quantitative, experimental to non experimental, can add evidence to the cumulative picture when meta-analysis is used. In a related vein, our technique allows us to confront the assumption that all primary studies (case studies or otherwise) must have used control groups to be included in a meta-analysis. In the interest of parsimony we refer to organizational interventions for the purposes of discussing and presenting our solution, but the logic can be applied to any field where interventions are studied.

Primary study researchers are only privy to the result from their own study, and perhaps a few others from the literature. A common approach to gauging the impact of an intervention is through the use of a pre-post design. Through the strict use of experimental controls, namely a pretest to posttest comparison and a control group to experimental group comparison, researchers can rule out many threats to internal validity[1]. Intervention effectiveness is estimated by the pre-post difference in the dependent variable and the control/experimental comparison. Internal validity is indeed important, but a key issue for the study of any

[1] We thank a reviewer for reminding us that experimental designs do not require a pretest. This is the case where proper randomization to control and experimental groups has occurred, which is more realistic in lab settings. We do not specifically address these situations, which would have a control group estimate but would not have a pretest to posttest estimate. Moreover, there are other definitions of intervention effect than pre/post change in the experimental group compared to pre/post change in the control group. Our technique does not directly include other definitions, but we expect compatibility.

intervention is related to external validity, or generalizability: To what extent can results generalize from one setting to other settings?

From the Single, Isolated Intervention Study to Meta-analysis

Many authors have written on the usefulness of control groups in research design (e.g. Campbell & Stanley, 1963). They have highlighted the importance of control groups in isolating the impact of the independent variable on the dependent variable and they identify a variety of threats to internal and external validity that can be allayed using one or more control groups. Unfortunately, in any social context, control groups are difficult to create. Consider the evaluation of organizational interventions. Since all parts of the organization are affected by a management intervention, there is usually no opportunity to assign some units of the organization (departments, divisions, or units) to a control group. If a particular study shows a change in the dependent variable, it may be a result of the intervention, but it could also be due to the influence of some extraneous factor which has nothing to do with the management intervention (Reichardt & Gollob, 1989).

Some challenges to the internal or external validity of the study can be eliminated by careful analysis concerning how the program was implemented (to eliminate the possibility of confounded treatments) and by careful analysis of the dependent variable that is used to measure outcomes (to avoid problems with construct validity). The more difficult challenges are those that stem from a consideration of factors outside the study, such as a productivity study being influenced by product demand or the fluctuating price of materials. Some such challenges can be eliminated with the assistance of managers who would know if they had occurred. It may be possible to eliminate most external threats if such factors are well known. There is always cause to worry, however, that the study finding might be influenced by a factor that both management and the researcher have failed to recognize.

The challenge of interpretation takes on an entirely different character when the findings of many studies that evaluate an intervention are analyzed cumulatively. There is a strong bias in meta-analysis for inclusiveness, so all study findings are routinely entered in the data base for analysis. Therefore, studies that do not have control groups are likely included in the meta-analytic data base.

Without evidence from a control group, it is not possible to correct a single study estimate for the influence of extraneous factors, which may have magnified or diminished its influence. The effect size that is entered in the meta-analytic data base is the observed effect of the intervention. In contrast, consider the character of the effect size in a study that does have a control group. Standard procedure is to adjust the effect size shown for

the intervention using the effect size for the control group. The influence of unknown extraneous factors is acknowledged by subtracting change in the experimental group from change in the control group.

The effects from non-control group studies and control group studies should not be combined and the difference simply ignored. Yet, there are inevitably some studies that are entered into the database that have no control group and no correction for extraneous factors. How many? For reasons that we will explain next, few studies that evaluate social outcomes use control groups.

Scarcity of Control Groups in Social Intervention Studies

Social interventions come in many different forms. But to explore the difficulty of arranging for a control group in a social intervention study, we will utilize the example of an organization-wide management intervention (or program, such as Total Quality Management). Productivity gains are expected when top management participates in the intervention process (Rodgers & Hunter, 1991). How could management justify leaving some of the second level managers out of the process while including others? Even if the workers who are excluded could be made to tolerate what is obviously an unequal chance at promotion, influence, and power, they would still know that they were not treated equally (diffusion of treatment), which could impact the intervention's success. For a typical intervention the whole organization is part of the study; thus, we can expect to find few that use a control group design (Mitchell, 1985).

Now consider the use of a control group in a social program intervention. If the program designers actually believe that the program will work and thereby improve the quality of life for participants, they would feel moral obligations to make the program available to all who are interested, rather than generating a random control group.

There is another way to generate a control group: study a perfectly matched entity that does not use the intervention. The reality is that organizations or jurisdictions can rarely be perfectly matched to one another, thus quasi-control groups are sometimes utilized (lacking random assignment thus questionable regarding comparability). Organizations, for instance, usually survive because they find a niche that competing organizations do not serve. This means that organizations with similar products or services usually differ in the specific niches that they serve and are thus subject to the influence of different external factors. Furthermore, matching assumes knowledge of the relevant matching factors; even if an organization could be perfectly matched on known extraneous factors, they may not necessarily be matched on unsuspected extraneous factors. Thus, because the two organizations are not perfectly matched, the results may be caused by the influence of extraneous factors

that were not controlled.

The obstacles confronted by researchers who do intervention studies are thus formidable. To demonstrate the issue, we consider whether studies evaluating organization interventions in the literature typically use control groups. Our brief, nonsystematic sampling of different organizational intervention research domains yields telling results. The proportion of studies that have control groups in the ten research domains considered range from 3.3% to 42.8% (see Table 1). Across domains, a mere 78 out of the 710 studies (11%) had control groups in pre-test posttest designs. Although our sampling is not representative of all study domains or intervention fields, it demonstrates our point that there is a dearth of control group studies in fields that study social interventions. If all listed studies are included in a meta-analytic data set, there will be no adjustment for the influence of unknown extraneous factors in 89% of the studies on average.

Table 1
Number of Pre-test Posttest Control Studies of Organizations as a Percent of All Studies Reviewed

Domain Research Study	Number reviewed	Control Studies	%
Decentralization			
Yin & Heald, 1975	269	9	3.3%
Feedback			
Kopelman, 1986	28	5	17.9%
Flexitime			
Golembiewski & Proehl, 1980	32	2	6.2%
Formal Planning			
Armstrong, 1982	12	1	8.3%
Goal setting			
Mento, Steel & Karren, 1987	23	9	39.1%
Latham & Lee, 1986	28	12	42.8%
Job enrichment			
Berlinger et al., 1988	34	2	5.9%
Quality circles			
Barrick & Alexander, 1987	33	2	6.1%
Management by objectives			
Kondrasuk, 1981	71	5	7.0%
Rodgers & Hunter, 1991	70	6	8.6%
Organizational development			
Woodman & Wayne, 1985	50	16	32.0%
Participation in Decision Making			
Miller & Monge, 1986	60	9	15.0%
Totals and Overall Average	710	78	11.0%

A mathematical meta-analysis procedure will be developed in this article that introduces a novel solution to the problem of having a significant body of study evidence that does not use control groups. Our method adjusts the findings of intervention studies for the influence of unknown extraneous factors. This correction is accomplished by recognizing that the averaged change shown by control groups constitutes the best estimate of the influence of extraneous factors.

Base Rate of Change Caused by Extraneous Factors

Now let us reconsider the problem of extraneous factor influences in the context of a review study. Results might be affected by two very different kinds of extraneous influences: extraneous factors that are suspected by the meta-analytic reviewer or extraneous factors that are unknown. If an extraneous factor is hypothesized, then the possible confounding effect can be explicitly tested. In the context of the meta-analysis, this test can take the form of a "moderator variable" analysis in which the meta-analysis contrasts the observed treatment effects between studies in which the extraneous factor is present with studies in which that factor is absent. There should be a different amount of change in the studies where the factor is present than in the studies where the factor is absent. It is important to note that an explicitly hypothesized extraneous factor can be tested in a review context even if no study had a control group.

If an extraneous factor is something unsuspected by the reviewer, then it cannot be tested explicitly. However, the effects of such factors can be assessed using the collective body of study evidence. The key to our solution lies in the assumption that there will be a sufficient number of studies that do use control groups. It is possible to analyze the cumulative estimates for control group organizations to assess both the mean and variation of effects due to extraneous factors. The mean and variance for the intervention organizations and the control organizations must first be computed separately. Then, findings are considered in combination to derive an estimate of the effects of the intervention that are free of the potentially confounding effects of extraneous factors.

For simplicity, let us illustrate with an ideal case: the case where extraneous factors do not impact the dependent variable. Consider an analysis of change from studies that have control group data. Once the influence of sampling error is eliminated, control groups will show a mean change of 0 and a standard deviation of 0. That is, mean change will be zero and there will be no variation from zero change across control groups. Once this fact is known, we would know that it is not necessary to have a control group for any of the studies; change in the control group would

differ from zero only by sampling error. Use of control groups in further studies in that domain would be unnecessary.

There may be some research domains in which extraneous factors do play a role. In such domains, the analysis of control group change data will suggest something more complicated. The standard deviation in control group change may not be zero.

Consider a second important case: the case of zero mean change and change in the standard deviation that is not zero. Suppose that extraneous factors sometimes cause outcomes to increase and sometimes they cause them to decrease. If the two possibilities are equally likely, then the mean change for control groups across studies would be zero, but the standard deviation even after taking into account sampling error may not be zero. The size of the standard deviation is affected by the influence of extraneous factors. The larger the standard deviation, the stronger is the relationship between the extraneous factors and the effect of the intervention. The larger the standard deviation, the greater is the difficulty of interpreting the result in any one isolated study if it does not have a control group.

This difficulty vanishes, however, in the very different context of a cumulative meta-analysis of change. In particular, consider how the mean change in experimental groups across studies is interpreted. If the mean change produced by extraneous factors is zero, then the mean change in experimental groups is produced only by the intervention. Thus, the mean change in the dependent variable across experimental groups is the same as the mean effect of the intervention. In this case, control group data are not needed to compute the average effect of the intervention.

It is usually not sufficient to know the average effect of an intervention. We would also like to know the standard deviation of treatment effects. Because of the effects of extraneous factors, the standard deviation of treatment effects may be a poor estimate of the standard deviation of true treatment effects. Even though the mean effect of extraneous factors is zero, the variation in the effect of extraneous factors would still add to the variation in observed change.

This problem too can be solved with meta-analysis. We will show that if the mean effect of extraneous factors is zero, then the variance of actual treatment effects will be equal to the difference between the variance of change across intervention groups less the variance of change across control groups. Control group data are analyzed separately from the intervention group data in order to obtain the variance value that is needed to correct the variance of intervention effects that are observed.

If mean control group change is not zero, then the average effect of extraneous factors is not zero. We will show that the mean treatment effect in this case is the difference between the average change in the intervention groups less the average change in the control groups.

Thus, in the end, the meta-analysis handles all cases: (a) cases in which the extraneous factors are equally likely to cause positive as negative effects and cases in which there is some systematic bias in extraneous effects; (b) cases in which the extraneous factors have the same effect in all studies and (c) cases in which the effect of extraneous factors varies from one study to the next.

We stress the nature of the goal for our study. The issue is *not* the usefulness of control groups in research design. When feasible, we support the use of control groups. Our goal is to present a solution that allows a meta-analyst to cumulate information from studies of interventions when the control group design is not used. In so doing, we derive the formal meta-analytic procedures for mathematically estimating the distribution of true intervention effects for organization study domains where there are many studies without control groups.

The Mathematics of Intervention Studies

The Isolated Study

We will first present, in mathematical form, the situation where a control group is present in the context of the single, isolated study. We will then show how the cumulative results of all studies can be used to correct for the effects of extraneous factors even though most studies do not have control groups. This argument will assume that the individual intervention studies are pre-post studies so that the effect of the program is estimated by the change that is seen from before to after the intervention.

Let us denote the organization outcome pre-post change measure by X and write:

XE = pre-post change in the organization with the intervention
XC = pre-post change in a perfectly matched control organization

Let us denote the effects of the intervention by R and the combined effects of extraneous factors by E. We then have:

$$XE = R + E$$

Since the effects of the extraneous factors on the control organization are identical to the effects on the intervention group, we have:

$$XC = E$$

If we have control group data, then we can also compute the difference

between change for the experimental and control group:

$$XE - XC = (R + E) - (E) = R$$

Thus, the intervention effect R is the difference between the pre-post change for the experimental group and pre-post change for the control group:

$$R = XE - XC$$

Thus, if we have data from a perfectly matched control organization, we can exactly correct for the effect of extraneous factors by subtraction. If we have no data on a control organization, then we know only the number XE = R + E and we cannot subtract the unknown number E.

Meta-Analysis of Many Studies

Now let us switch contexts and consider the results of a large set of studies. Since the effect of a program may vary somewhat from one organization to another, the effect R may vary to some extent from one organization to another. If so, then we want to know the distribution of effects across organizations. In particular, we would like to know the average effect [i.e., Ave(R)]. We would also like to know how much the effect varies across settings [i.e., SD(R)]. If all studies had been done with control groups, we would first compute the difference R for each study. We could then compute the mean and standard deviation across organizations using the usual statistical formulas.

The mean effect. In order to find the average treatment effect Ave(R) across settings we would average the R value across studies. But most studies of organizational interventions do not have a control group. That is, for most isolated studies, the value R for that study is not known, so we cannot simply average R values across studies. We could however approximate the desired mean by averaging the effects for the experimental organizations. We would then have:

$$Ave(XE) = Ave(R + E) = Ave(R) + Ave(E)$$

Note that this average includes all experimental groups regardless of whether there was a control group in the study or not. The error in this estimate is the number Ave(E). Note that Ave(E) is not the effect of extraneous factors for one isolated study, but is the average of effects across all studies. There are two possible outcomes of this analysis. Either Ave (E) is zero (the "zero baseline" case) or it is not (the "nonzero

baseline" case). We will consider each possibility.

Unbiased extraneous effects: "zero baseline." With this special case, an organizational intervention is evaluated where the effect of extraneous factors occur randomly across studies. For unbiased extraneous effects of this sort, the baseline for meta-analysis is zero. That is, for this case we have:

$$Ave(E) = 0$$
$$Ave(XE) = Ave(R) + Ave(E) = Ave(R)$$

That is, the average effect across studies with only experimental group data is correct even though the value for any single value may be incorrect.

If the number of studies is small, then the average extraneous effect across studies will differ from 0 by second order sampling error due to chance variation in the extraneous factors from one study to the next, but the principle is the same: there will be much less error in the average effect than in the effect for single, isolated studies.

General case: "nonzero baseline." There may be some interventions or dependent variables for which the average extraneous effect is not zero. If Ave(E) is not zero, then we can still solve the problem using meta-analysis if at least some of the studies present control group data.

As an example, consider productivity of employees in an organization as a dependent variable. Improved productivity is a goal that is considered frequently by both managers and workers. Thus, it seems likely that over a period of time, there will be some improvement in productivity due to day-to-day efforts. In this case, the mean effect of the extraneous factor is not zero, but positive. If Ave(E) is not 0, then it may produce a considerable error in the average effect in experimental groups. This problem too can be solved in meta-analysis. Assume that some of the studies do use a control group. Suppose we average the change for the control organizations that are available. We would have:

$$Ave(XC) = Ave(E)$$

That is, the average effect of extraneous studies on the subset of control organizations would be the same as the average for the entire set of studies. Thus, we could subtract the two averages:

$$Ave(XE) - Ave(XC) = [Ave(R) + Ave(E)] - Ave(E)$$
$$= Ave(R)$$

That is, the difference between the averages would be the same as the

average difference (i.e., the difference between the averages would be the desired average treatment effect). Thus, meta-analysis can accomplish what individual studies cannot accomplish: estimate the mean effect of the intervention.

Extraneous factor by treatment interaction. Our procedure for estimating the mean treatment effect works even if there is an extraneous factor by treatment interaction. To see that our meta-analysis procedure still holds, consider an isolated organization. If there is an interaction, then the treatment effect R for that organization depends on the extraneous factors present in that organization at that time. However, from an applied point of view, the number Ave(R) still tells the average effect of the treatment across settings. Had that organization not been exposed to the intervention, the mean change would be the simple effect of the extraneous factors alone. Thus, for each organization, we still have:

$$XE = R + E$$
$$XC = E$$

If we average across experimental organizations, we have:

$$Ave(XE) = Ave(R + E) = Ave(R) + Ave(E)$$

If we average across control organizations, we have:

$$Ave(XC) = Ave(E)$$

Hence the difference between the averages is the desired estimate of the average effect across settings:

$$Ave(XE) - Ave(E) = Ave(R)$$

If there is an extraneous factor by treatment interaction, then the key problem is that there is variation in the treatment effect that is not explained by factors explicitly considered by the investigator. That is, if there is an interaction of extraneous factors with the treatment, then the extraneous factors are simply unrecognized moderator variables. If the variation is large, the investigator would want to identify such interacting extraneous factors so they could be analyzed as potential moderator variables. As this stage however, these factors should be considered as moderator variables rather than as unknown "extraneous" factors.

Note that the conventional control group design for isolated, single studies does make the assumption that there is no interaction between the extraneous factor and the intervention. If there is an interaction, then the

effect R depends on the extraneous effects present. If the variation due to the interaction is large, then the extent of generalization from that study would be limited.

Using meta-analysis, it is possible to solve the problem of the interaction, though a large number of studies may be necessary. The key to meta-analysis is that the extraneous factor varies across studies. The way to handle an interaction within a single study is to use a two-way design, which systematically varies the extraneous factor as well as the treatment factor. Note that this requires the extraneous factor to be known. If the extraneous factor is not known, then the two-way design cannot be used in the single study. But consider a meta-analysis with many studies. Suppose there is an interaction between some extraneous factor and the intervention. After all anticipated moderator variables are taken into account, there will still be variation in the treatment effect that is not explained. It may be possible at this point to go back to the studies and identify the extraneous factor. The extraneous factor could then be analyzed as a moderator variable. Within each subset of studies where the extraneous factor takes on the same value, the simple analysis presented here would yield the treatment effect for that level of the extraneous factor.

Note that in this moderator analysis, it is not the mean control group change which varies from one level of the extraneous factor to the next. It is the experimental group whose outcome varies depending on the level of the extraneous factor. The interaction does not affect the control group because the intervention is *not* introduced.

Note that if the interaction is not discovered, the mean treatment affect would still be well defined. The average treatment affect would be the average of the treatment affects across different levels of the extraneous factor. Since these levels are determined by local study conditions, the frequency of the various levels should approximately match the ecological frequencies in the domain of real organizations (if that is the unit of analysis). Thus, the average treatment effect estimated by the meta-analysis is an ecologically correct estimate of the average treatment affect in organizations where the extraneous factor is unknown and thus uncontrolled. Even if there were some unknown extraneous factor that had an interaction with the intervention, the meta-analysis would still produce a best estimate of its affect. Of course, heterogeneity across the intervention group studies and control group studies may be unequal because of fewer control group studies. However, any bias in the estimates is minimized by large aggregate subject numbers, and the control group adjustment still allows for the reduction of bias in the intervention group estimate.

Variance of effects. Meta-analysis provides a straightforward estimate of the mean effect of the intervention. This is also true of the variance of

effect sizes. Consider first the case where there is no interaction between the treatment effect and the effect of the extraneous factors. Across studies the two effects R and E will be uncorrelated, so the covariance of R and E will be zero. The variance of change in experimental groups across studies will then be given by:

$$Var(XE) = Var(R + E)$$
$$= Var(R) + Var(E) + Cov(R,E)$$
$$= Var(R) + Var(E)$$

This variance is too large by Var(E), the variance of extraneous effects. This variance can be estimated by looking at the variance of effects for the control organizations. That is,

$$Var(XC) = Var(E)$$

By subtraction we have:

$$Var(XE) - Var(XC) = Var(R)$$

Now consider the unlikely case of an extraneous factor by treatment interaction, (i.e., Cov (R,E) is not zero). The variance of the treatment is increased by the covariation between the treatment and some unknown extraneous factor. But if the interaction is large, it should be feasible to identify a moderator variable. The variance of the treatment effect will be increased because the moderator variable is present in some studies and absent in others. Thus, in the final analysis, the increase in variance will be explained by the moderator variable and not by an extraneous factor by treatment interaction. Even if a moderator variable is not found, it is the variance of the treatment effect that is increased: the extraneous factor is present in some organizations and absent in others. Meta-analysis thus also produces a correct estimate of the treatment variance as long as there are a sufficient number of studies to insure a reliable representation of the extraneous factor in the organizations that are studied.

The variance of treatment effects across studies can thus be estimated by the difference between the variance of experimental group change and the variance of control group change. The variance difference as an estimate of treatment effect variance is more robust than it might appear. Many of the artifacts that alter effect sizes (such as sampling error) also increase both variances in similar ways and hence tend to cancel out in the difference.

Pooling data for control group cases. The key consideration in determining the accuracy of estimates for the mean and variance of the

intervention is the number of values that are available to derive averages for the control group change and intervention group change. In estimating mean control organization change, the relevant statistical "sample size" is the number of organizations (or other entities: groups, jurisdictions, etc.). For experimental groups, there may be a dozen or more organizations with change data. However, the number of control organizations is usually quite small. Recall that only 11% would be expected to have control unit data. Thus, within one research domain, the mean and variance of control group effects will likely be based on a relatively small sample of studies and thus subject to considerable sampling error.

However, it is possible to pool control organization data across research domains and thus produce the large samples needed for accurate statistical analysis. That is, whereas experimental organization change might differ drastically from one intervention to the next, control organization change should not change from one research domain to another because the intervention is not introduced in the control organization. For any given outcome variable, the effect of extraneous factors in the control organization (where the intervention is not used) does not depend on the nature of the intervention tried in the experimental organization.

For example, consider the following management question: how much does job satisfaction change in an organization that does not introduce a major intervention to its management system? Consider this question in two research domains: Total Quality Management (TQM) and organization development. A control group for TQM research is an organization where TQM was not introduced. Thus, there is no reason why change in that control organization should be altered by the fact that TQM is being introduced in the experimental organization. A control organization for organization development research is an organization where neither organization development nor TQM is introduced. Thus, there is no reason why change in that control organization should be altered by the fact that organization development is being introduced in the experimental organization, assuming the control and non control organizations are not systematically different.[2]

Now consider the two examples together. Since change in the control organizations does not depend on what is done in the experimental organizations, there is no reason to expect that change will be any different in the TQM control organization than in the organizational development control organization. Whereas experimental organizations in different research domains are quite different because different management systems are being tried out, there is no such difference between control organizations where new interventions are not being tried out. Thus, there

[2] We thank a reviewer for pointing out this assumption.

would be only random differences between control organizations in different research areas.

D-statistic Corrections

Sampling error. Evaluation studies generally use the more powerful pre-post design rather than the posttest-only design assumed in most meta-analysis texts. Thus, within- subject formulas for sampling error must be used in place of the more typical independent groups formulas. The appropriate formulas have been developed by Hunter and Schmidt (1990) and will be summarized here.

Since the sampling error for the d statistic is almost perfectly symmetric in distribution (for sample sizes greater than 10 or so), the mean effect size need not be adjusted for sampling error. However, the variance of the d statistic is greatly increased by sampling error and thus must be corrected to produce an accurate estimate of the actual variation in effect size (Hedges & Olkin, 1985). Let us denote the population value of d by δ. The sample value is related to the population value by

$$d = \delta + e$$

where e is sampling error. The variance of the sample statistic is related to the variance of the population values by:

$$Var(d) = Var(\delta) + Var(e)$$

That is, the sampling error variance Var(e) directly inflates the estimate of the variance of effect sizes. This leads then to the correction formula:

$$Var(\delta) = Var(d) - Var(e)$$

If the average sample size is greater than 30, the sampling error variance is approximately

$$Var(e) = \frac{2(1-r)(1+d^2/8)}{N}$$

where r is the pre-post correlation, d^2 is the square of the mean d and N is the average sample size (Hunter & Schmidt, 1990). Most studies do not report the pre-post correlation, but if the treatment by subjects interaction is not large, the pre-post correlation is approximately equal to the reliability of the outcome variable.

Measurement error. Error of measurement can have a very large impact on the mean effect size. Error of measurement causes a blurring of the distribution which in turn causes the observed d statistic to be less than the true value. The standard psychometric correction for error of measurement works for the within subject d statistic just as it works for the between subject d statistic. If d_t is the d statistic for the "true" (i.e. perfectly measured) dependent variable, and d_x is the d statistic for the observed (or "fallible") measure, then the correction formula is:

$$d_t = d_x / \sqrt{r_{xx}}$$

where $\sqrt{r_{xx}}$ is the square root of the reliability.

The standard deviation of effect sizes is also altered by error of measurement. If there is variation in the reliability of the outcome variable over studies, then that variation will cause an artifactual increase in the variance of observed effect sizes. This increase is usually very small in magnitude and will be ignored here. However, the fact that error of measurement causes observed effect sizes to be smaller than true effect sizes also causes the standard deviation in observed effect sizes to be smaller than the standard deviation of true effect sizes. Thus, the standard deviation of effect sizes must also be corrected for error of measurement. If SD_t is the standard deviation of effect sizes for the true dependent variable and SD_x is the standard deviation of effect sizes for the observed measure, then they are related by:

$$SD_t = SD_x / \sqrt{r_{xx}}$$

To summarize, there is a sequence of four steps that are necessary to perform all of the adjustments that are needed to correct for the influence of extraneous factors using meta-analysis.

1. Compute the mean and variance of change for the intervention organizations and the control organizations separately.
2. Following standard meta-analytic procedures, correct each estimate of change for the effects of sampling error, error of measurement and other artifacts (if possible).
3. Estimate the effect of the treatment itself by computing the difference between the mean estimate of the intervention and the control group mean estimate. Also compute the difference in standard deviations. The difference between the experimental mean and the control mean is the mean treatment effect across studies; the difference between experimental

and control variances is the variance of treatment effects across studies. The difference in variances can be negative because of sampling error in which case the standard deviation of treatment effects is zero (i.e., no variation across studies). If the variance difference is positive, then its square root is the estimated standard deviation of treatment effects across studies. Step 3 corrects for the effects of extraneous factors.

4. To correct for error of measurement, divide the mean and standard deviation derived in (3) by the square root of the reliability of the outcome variable.

Conclusion

In the single study, the potential for the influence of extraneous factors forces the need for a control group. If the single study has no control group, interpretation of that study's result is equivocal. In an ideal world, it would be desirable for each and every study to have a control group. In social intervention research however, it is seldom feasible to use a control group. Our sampling of review studies showed that only 11% of studies used control groups. Thus, when each study is considered one-by-one, there is no way to assess the experimental effects for 89% of research evidence. The key to our method is to use the data for those few studies that could use control organizations in order to generate an estimate of the baseline mean and variance of extraneous factors. The cumulative meta-analysis considers the results of many studies and thus contains far more information about the effects of extraneous factors. Thus, meta-analysis can solve the interpretation problem caused by of the effects of extraneous factors.

Even with the increased power of the pre-post design, it is difficult to interpret the results of a single, isolated study because of the possible effects of extraneous factors (Campbell & Stanley, 1963). In studies where the individual is the unit of analysis, it is easier to remedy this problem by using random assignment to create a control group. Comparison of the experimental group change to the control group change eliminates the effects of extraneous variables. But by the nature of social intervention studies, a control group is usually not feasible because the study takes place outside of a “lab” in a “real world” setting. Often, in these studies the unit of analysis is not on the individual level, but is on the group level (organizations, jurisdictions, etc.).

We have developed a method of analyzing social intervention research using meta-analysis. The method recognizes and resolves one of the lingering problems that intervention researchers face: the existence of mixed-methodologies across individual studies. As long as some of the studies are able to use a control group design, it is possible to correct for the effects of extraneous factors in all studies.

There are some situations where the applicability of the method we propose may be more limited. For instance, a meta-analyst may have a low number of control group studies. This could undermine the assumption that all extraneous factors are accounted for. Of course there is no prescribed number of studies that is adequate, since there is no statistical basis and it may vary across study domains and topic areas. Second, primary studies could contain vastly different design types, which may lead to the presumption that control group studies not be aggregated. If this is the case then we suggest testing whether there is a difference between the aggregate effect sizes for different control group types, by moderator variable testing. Third, it's conceivable that there may be unknown systematic extraneous factors that cut across all experimental groups or control groups. These could be classical threats to internal validity in experimental settings. This unlikely scenario would undermine the accuracy of the aggregate effect size adjustment. But, even here the aggregate effect size adjustment would be preferable to the unadjusted effect size, because it would still be more accurate.

Author notes: John (Jack) E. Hunter passed away in 2002. Robert Rodgers is retired from academia. The contact author is Jason Jensen. He can be reached at jjensen@business.und.edu.

References

Armstrong, J. S. (1982). The value of formal planning for strategic decisions: Review of empirical research. *Strategic Management Journal*, *3,* 197-211.

Barrick, M. R., & Alexander, R. A. (1987). A review of quality circle efficacy and the existence of positive-findings bias. *Personnel Psychology*, *40,* 579-592.

Berlinger, L. R., Glick, W. H., & Rodgers, R. C. (1988). 'Job enrichment and performance improvement'. In J. Campbell and R. Campbell (Eds.), *Productivity in Organizations (*pp 219-254). San Francisco: Jossey Bass.

Bausell, R. B., Li, Y., Gau, M., & Soeken, K. L. (1995). The growth of meta-analytic literature from 1980 to 1993. *Evaluation and the Health Professions*, *18,* 238-251.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.

Golembiewski, R. T., & Proehl, C. W. (1980). Public sector applications of flexible workhours: A review of available experience. *Public Administration Review*, *40,* 72-85.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis*. Newbury Park: Sage Publications.

Jensen, J. L., & Rodgers, R. (2001). Cumulating the intellectual gold of case study research. *Public Administration Review*, *61,* 235-246.

Kondrasuk, J. N. (1981). Studies in MBO effectiveness. *Academy of Management Review*, *6,* 419-430.

Kopelman, R. E. (1986). *Managing productivity in organizations: A practical, people oriented approach*. New York: McGraw-Hill.

Latham, G. P. & Lee, T. W. (1986). 'Goal setting', In E. A. Locke (Ed.), *Generalizing from laboratory to field settings: Research findings from industrial-organizational psychology, organizational behavior, and human resource management* (pp. 101-117). Lexington, Mass: D.C. Heath and Company.

Mento, A. J., Steel, R. P., & Karren, R. J. (1987). A meta-analytic study of the effects of goal setting on task performance, 1966-1984. *Organizational Behavior and Human Decision Processes*, *39,* 52-83.

Miller, K. I., & Monge, P. R. (1986). Participation, satisfaction, and productivity: A meta-analytic review. *Academy of Management Journal*, *29,* 727-753.

Mitchell, T. R. (1985). An evaluation of the validity of correlational research conducted in organizations. *Academy of Management Review*, *10,* 192-205.

Reichardt, C. S., & Gollob, H. F. (1989). Ruling out threats to validity. *Evaluation Review*, *13,* 3-17.

Rodgers, R., & Hunter, J. E. (1991). Impact of management by objectives on organizational productivity. *Journal of Applied Psychology, 76*, 322.

Woodman, R. W., & Wayne, S. J. (1985). An investigation of positive-findings bias in evaluation of organization development interventions. *Academy of Management Journal*, *28,* 889-913.

Yin, R. K., & Heald, K. A. (1975). Using the case survey method to analyze policy studies. *Administrative Science Quarterly*, *20,* 371-381.