Title	Measuring differentials in communication research : issues with multicollinearity in three methods
Author(s)	Rosenthal, Sonny
Citation	Rosenthal, S. (2013). Measuring differentials in communication research : issues with multicollinearity in three methods. Communication methods and measures, 7(2), 106-125.
Date	2013
URL	http://hdl.handle.net/10220/18828
Rights	

Measuring Differentials in Communication Research:

Issues with Multicollinearity in Three Methods

Sonny Rosenthal

Wee Kim Wee School of Communication and Information

Nanyang Technological University, Singapore

Keywords: differential, raw difference score, index of change, partial variance, multicollinearity

Author Note

Please address correspondence concerning this article to Sonny Rosenthal, 31 Nanyang

Link #04-39, Singapore, 637718; or email to sonnyrosenthal@gmail.com.

Measuring Differentials in Communication Research:

Issues with Multicollinearity in Three Methods

Abstract

Models of communication processes sometimes require the computation of the difference between two variables. For example, *information insufficiency* is the difference between what people know and what they think they need to know about an issue, and it can motivate information seeking and processing. Common methods that compute this differential may bias model estimates as a function of the correlation between the differentiated variables and other variables in the model. This article describes Cohen and Cohen's (1983) analysis of partial variance for computing differentials, and analyzes simulated data in order to contrast that method with two alternative methods. The discussion recommends the use of the Cohen and Cohen method in other areas of communication research, such as studies of third-person perception.

Keywords: differential, raw difference score, index of change, partial variance, multicollinearity

Measuring Differentials in Communication Research:

Issues with Multicollinearity in Three Methods

Models of communication sometimes require the computation of the difference between two observed variables. Longitudinal studies generally involve the measurement of an initial condition and a subsequent condition (i.e., a pre-score and a post-score, respectively). Some cross-sectional studies measure analogous variables, but whose observations are simultaneous.[1] The current study considers the measurement of such differentials in communication research.

This study describes a statistical inadequacy of a common measure of information insufficiency, which uses a variant of Cohen and Cohen's (1983) analysis of partial variance (APV) in order to compute a differential. This variant method, which seeks to overcome limitations of using a raw difference score, tends to overestimate R^2 when information insufficiency is either a dependent or independent variable, and is sensitive to the degree of correlation between the pre-score—in this case, perceived current knowledge—and other variables in the model—for example, perceived behavioral control over seeking. Researchers continue to use this method with the justification that "prior researchers used it."

Thus, this article has specific and generic functions. The specific function is to provide guidance to researchers of information insufficiency and similarly structured models. The generic function is to draw attention to the measurement of differentials in communication research and to extend the discussion on the appropriateness of different methods in different research contexts.

First, I outline Cohen and Cohen's APV and how the variant method can produce potentially large bias in regression models. Then, I test a series of regression models using

---

[1] For example, information insufficiency is the difference between current knowledge and desired knowledge. Although these variables are measured at the same time, they imply a time order: current knowledge precedes future knowledge.

simulated data in order to evaluate the extent of this bias. Analyses compare three different measurements of change—(a) an index of change per Cohen and Cohen's APV, (b) a residual score per the APV variant method, (c) and a raw difference score—each as a dependent and independent variable at varying degrees of correlation between the pre-score and a covariate. For clarification, Table 1 distinguishes among the pre-score, post-score, differential, "covariate," and dependent variable. I use these labels throughout the manuscript to describe the variables in the regression models.

This study is tests the effects of a specific case of multicollinearity that can occur with certain differentials: when a differential contains variance due to the pre-score, the correlation between the pre-score and a covariate will confound the correlation between the differential and the covariate. Hence, multicollinearity among the pre-score, the differential, and the covariate may bias results. Cohen and Cohen's APV specifically controls for variance of the pre-score in the differential, which eliminates this potential bias. Residual scores from the APV variant and raw difference scores do not control for this variance and remain vulnerable to the bias of multicollinearity. I expand on this argument subsequently.

This study gives statistical examples in reference to models of information seeking and processing; thus, I begin with a brief overview of that line of research. The following literature review provides theoretical context and helps orient the reader to the statistical issue that motivates this study.

An Illustration of a Differential: Information Insufficiency

Information is an important basis of decision making, and its quality and quantity can affect the outcomes of a variety of behaviors, including public policy evaluations (Garber & Sox, 2010), business decisions (Kassim, 2010; Zacharakis & Meyer, 1998), consumer purchases

(Kowatsch & Maass, 2010), resource management (Dennis et al., 1996; Holmgren & Thuresson, 1998), and others. Information needs and uses are myriad: scientists need and use information to draw valid conclusions; policymakers need and use information to enact effective and reasonable policies; and people generally need and use information as the basis of important (and often unimportant) decisions.

When people perceive that their current knowledge about an issue is inadequate, they may respond by seeking and processing new information (Griffin, Dunwoody, & Neuwirth, 1999). The heuristic-systematic model describes such *perceived information insufficiency* as a precursor to effortful information processing (Chaiken, 1980; Chen & Chaiken, 1999). According to the heuristic-systematic model, perceived information insufficiency occurs when current knowledge falls short of the threshold at which knowledge is sufficient to pass confident judgment about an issue. When the issue has high importance, the need for confidence is relatively high; thus, desired knowledge is relatively high. When the issue is inconsequential, desired knowledge is relatively low.

The Need for a Regression Approach

Through this theoretical lens, Griffin, Neuwirth, Dunwoody, and Giese (2004) examined antecedents of information insufficiency in a model of risk information seeking and processing. They measured information insufficiency using two variables: (a) current knowledge and (b) desired knowledge (p. 36):[2]

Current knowledge was measured as follows:

Now, we would like you to rate your knowledge about [a topic]. Please use a scale of 0 to

100, where 0 means knowing nothing and 100 means knowing everything you could

---

[2] The literature uses the terms "perceived knowledge" and "sufficiency threshold" to describe current knowledge and desired knowledge, respectively. However, for readers who are not familiar with that line of research, the latter two terms may have more intuitive meaning.

possibly know about this topic. Using this scale, how much do you think you currently

know about [the topic]?

Desired knowledge was measured as follows:

Think of that same scale again. This time, we would like you to estimate how much

knowledge you would need to [pass confident judgment about the topic]. Of course, you

might feel you need the same, more, or possibly even less, information about this topic.

Using a scale of 0 to 100, how much information would be sufficient for you, that is,

good enough for your purposes?

Information insufficiency reflects the gap between current knowledge and desired knowledge. As

the latter increasingly outsizes the former, information insufficiency grows in proportion. For

subsequent discussion, it may benefit the reader to think of current knowledge as a pre-score and

desired knowledge as a post-score.

The Deceit of Statistical Intuition

An intuitive approach to computing information insufficiency and other differentials is to

subtract the pre-score from the post-score. The result is a *raw difference score*. For example, on

the scale of 0 to 100, current knowledge of 30 and desired knowledge of 75 would yield an

information insufficiency of 45. The greater the resulting number, the greater the information

insufficiency.

However, Cohen and Cohen (1983) suggest that computing a raw difference score "is not

the simple, straightforward proposition it appears to be. Indeed. This is a methodological area

fraught with booby traps, where intuitive 'doing what comes naturally' is almost certain to lead

one astray" (p. 413).[3] The problem they attribute to computing a difference score is that the

resulting differential often contains variance due to the pre-score. As a result, the correlation

between the pre-score and a covariate may confound the relationship between the covariate and

the difference score.

More recent research has attributed the extent of this problem to a number of factors,

including measurement error, skewness, and floor/ceiling effects (Jamieson, 1999; Jamieson &

Howk, 1992). Consistent with this body of research, Griffin et al. (2004) note that computing

information insufficiency as a difference score can limit the reliability of the differential. In

response to the limitations of raw difference scores, Cohen and Cohen specify the APV to

compute an *index of change*. This index contains the residual variance of the post-score

controlling for the variance of the pre-score. Thus, the index of change is unrelated to the pre-

score, which circumvents a key limitation of difference scores, and is less sensitive than raw

difference scores to floor and ceiling effects (Cribbie & Jamieson, 2004).

The previous point warrants brief clarification: computing a raw difference score does not

create statistical problems in all cases. The problem with a raw difference score occurs in

analyses of covariance (e.g., regression and ANCOVA), while simple comparisons of means

(e.g., *t*-test and ANOVA) do not have the same problem. For example, the computation of

information insufficiency as a raw difference score allows for a straightforward description of

the mean difference, as it retains the scale (i.e., 0 to 100) of the pre-score and post-score. If a

research goal is simply to report the difference between two variables, then computing the raw

difference score is a valid approach. Rather, an index of change is useful in analyses that

---

[3] I cite the second edition of Cohen and Cohen's work. The 2003 third edition, with coauthors West and Aiken, dedicates considerably less space to the APV and gives less extensive statistical derivations and examples. Compare currently cited pages with pp. 59–60 and 570–571 in the 2003 edition.

examine the correlations among a pre-score, a post-score, and at least one covariate that correlates with the pre-score.

An Index of Change as Partial Variance

Cohen and Cohen (1983, p. 416) describe the following procedure to compute an index of change:

1. Use the pre-score (PRE) to predict the post-score (POST), and note the unstandardized regression slope (B).

2. Compute the index of change (C) as C = POST – PRE*B

The first step determines the proportion of variance in the post-score that is due to the pre-score, and the second step subtracts that variance from the post-score. The resulting index of change reflects the variance of the post-score unrelated to the pre-score, and thus variance that is due wholly to the intervening condition and measurement error. The term *analysis of partial variance* refers to this residual variance. Furthermore, since the index of change is unrelated to the pre-score, it is also unrelated to the common variance between the pre-score and any covariates. Consequently, the relationship between the differential and the covariate is independent of the pre-score, which precludes the aforementioned bias of multicollinearity.

A detailed discussion of the APV (see Cohen & Cohen, 1983, pp. 413–423) is beyond the scope of this article; however, the APV is, per Cohen and Cohen's derivation, an appropriate and useful method in communication research for one key reason: When the regression of the post-score on the pre-score approaches unity and their variances are approximately equal—which often occurs in the physical sciences, as when measuring children's height at two different times—the simple difference score yields unbiased results.[4] It is only when the post-score and

---

[4] Consider the following example: A researcher measures a pre-score (PRE) and a post-score (POST). For the set of observations, it is given that POST = PRE + 1. Thus, $r_{pre,post}$ = 1.00, $\sigma_{pre} = \sigma_{post}$, and B$_{pre,post}$ = 1.00. The index of

pre-score are largely unrelated—as often occurs in the social and behavioral sciences—that the

APV may be preferable to using raw difference scores when conducting analyses that involve the

differential and at least one covariate (Cohen & Cohen, 1983, pp. 416-417).

Information Insufficiency as Partial Variance

If current knowledge and desired knowledge are strongly correlated, then using a raw

difference score should suffice to compute information insufficiency; given a weak-to-moderate

correlation, the APV is more statistically appropriate. In a recent study of information

insufficiency as a differential, Kahlor (2007) reported a moderate correlation between current

knowledge and desired knowledge ($r = .32$, $p < .001$); thus, the APV should provide an

appropriate estimate of information insufficiency. The resulting index of change reflects the

portion of variance in desired knowledge (i.e., the post-score) unrelated to current knowledge

(i.e., the pre-score) and wholly due to information insufficiency (i.e., the interceding condition,

or differential) and measurement error.

The purpose of the hitherto discussion of information insufficiency was to ground the

current discussion in existing research. However, I assume the readership comes from diverse

theoretical backgrounds. Thus, for the remainder of this manuscript, I will largely eschew

reference to "current knowledge," "sufficiency threshold," and "information insufficiency," and

use the more generic terms, "pre-score," "post-score," and "differential," respectively.

Study 1: Modeling the Differential as a Dependent Variable

Researchers have used at least three methods to study differentials: the APV; a variant of

APV, which I describe below; and analysis of raw difference scores. In this section, I test the

---

change thus equals POST – PRE*B = (PRE + 1) – PRE*1 = PRE + 1 – PRE = 1 for all observations. The raw
difference score would yield an identical result.

extent to which the correlation between the pre-score and the covariate biases the regression of

the differential on the covariate. I compare results across the three methods.

Although the APV is useful for calculating information insufficiency, researchers have

yet to fully implement the approach. Many studies of information insufficiency reference Griffin

et al. (2004), who examined predictors of the differential with hierarchical regression of the post-

score on the pre-score in the first step and on predictors of interest in the second step (see also

Griffin et al., 2008; ter Huurne, Griffin, & Gutteling, 2009; Yang et al., 2011). I label this

analysis *the two-step method* for reference purposes. Figure 1 contrasts the APV, the two-step

method, and regression of a raw difference score when a covariate predicts the differential.

The idea behind the two-step method is that by entering the pre-score as the first

predictor, the residual variance of the post-score is unrelated to the pre-score and thus bears an

index of change per the APV. However, equating the differential (i.e., the "residual" variable)

with the post-score (i.e., the dependent variable of the analysis) can inflate explained variance in

information insufficiency. Indeed, when the covariate predicts the post-score, controlling for the

pre-score, the resulting R^2 reflects explained variance in the post-score, and not in the

differential. The correct R^2 for the differential should account for the correlation between the pre-

score and any covariates. Without this correction, the two-step method over-explains variance in

the differential as a function of the partial-r of the post-score's regression on the pre-score. Put

differently, in the first step of the two-step method, the residual variance of the post-score is

identical to the variance of the index of change per the APV; however, in the second step, any

correlation between the pre-score and the covariate will confound the portion of explained

variance uniquely attributable to the pre-score. I evaluate the extent of this effect with analyses

of simulated data.

Method

I used an Excel workbook to generate the following variables: a pre-score, a post-score, a covariate,[5] a dependent variable,[6] the index of change, and the raw difference score. Appendix A contains a detailed account of the simulation method.

I estimated three least squares regression models with the intercept constrained to 0. The dependent variables were (a) the index of change, (b) the post-score controlling for the pre-score, and (c) the raw difference score. I use $r_{pre,cov}$ to denote the zero-order correlation between the pre-score and the covariate, which I manipulated to produce 20 degrees of correlation for each of the regression analyses.[7] I compared unstandardized regression slopes (B) using a procedure I describe in Appendix B.

Results and Discussion

First, the unstandardized slope of the regression of the index of change on the covariate exhibited a linear trend across values of $r_{pre,cov}$, declining from $B = .40$ to $B = .11$ (Figure 2), with R^2 values exhibiting a linear trend, declining from .17 to .01. Intuitively, B and R^2 should approach zero as $r_{pre,cov}$ approaches unity: Since the index of change is unrelated to the pre-score, it must also be unrelated to the covariate, which is perfectly correlated with the pre-score.

Second, the unstandardized slope of the regression of the post-score on the pre-score (as the control variable) and the covariate (as the predictor of interest) increased logarithmically across values of $r_{pre,cov}$ from $B = .40$ to $B = 1.17$ (Figure 2), and was significantly different than

---

[5] The covariate can be either a predictor variable of direct interest or a control variable; however, it must either be a predictor of the differential or a predictor with the differential of a dependent variable. In Study 1, the covariate predicts the differential. In Study 2, the covariate and the differential predict the dependent variable.
[6] I do not use the dependent variable until Study 2, in which the differential is an independent variable. For example, in the model of risk information seeking and processing, information insufficiency (the differential) predicts risk information seeking intention (the dependent variable).
[7] The different levels of multicollinearity reflect differences in the correlation between the pre-score and covariate, which ranged from 0 to .95 in increments of .05. The remaining correlations among the full set of variables remained constant.

the slope of the regression of the index of change for $r_{pre,cov} > .15$. This difference was

significantly larger at higher values of $r_{pre,cov}$. Relative to the APV, the two-step method

increasingly overestimated R^2 at higher levels of multicollinearity (ΔR^2 range: .07, .22; Figure 3).

Finally, the unstandardized slope of the regression of the raw difference score on the

covariate declined linearly from $B = .40$ to $B = -.55$ (Figure 2), and was significantly different

than the slope of the regression of the index of change for $r_{pre,cov} > .01$. Up to $r_{pre,cov} \approx .65$, the

regression of the raw difference score underestimated R^2 relative to regression of the index of

change; at higher values of $r_{pre,cov}$, it overestimated R^2. Notably, R^2 approached zero at $r_{pre,cov} =$

.40, which makes intuitive sense, since the correlation between post-score and the covariate was

constrained to $r = .40$ (see Table 2, Figure 3). Furthermore, as $r_{pre,cov}$ values increased, the R^2

values of the two-step method and regression of the raw difference score converged (Figure 3).

For additional reference, Figure 2 shows the regression of the post-score on the pre-score,

controlling for the covariate, which is marked with shaded squares.

These results show that at as the correlation between the pre-score and the covariate

increases, regression of the differential either per the two-step method or as a raw difference

score exhibits increasingly biased regression slopes in relation to the regression of the index of

change.

Study 2: Modeling the Differential as an Independent Variable

Researchers may also use a differential as an independent variable. In such a

configuration, the two-step method may lead researchers to misinterpret results. For example,

Hovick, Freimuth, Johnson-Turbes, and Chervin (2011) tested a regression model in which a pre-

score (current knowledge), a post-score (desired knowledge), and several additional independent

variables (e.g., education, worry, systematic message processing) predicted a dependent variable

(health protective action). They found that the post-score significantly predicted the dependent variable, and concluded that the differential (i.e., the post-score controlling for the pre-score) was a significant predictor. Their conclusion is logical, as the regression on pre-score controls for its partial effect on post-score.

However, as the earlier analysis of simulated data showed, the correlation between the pre-score and the covariate can confound the residual variance of the post-score. Specifically, as the degree of correlation increases, the differential will increasingly reflect the post-score controlling for the covariate; whereas, the differential should reflect only the post-score controlling for the pre-score. Figure 4 contrasts the APV; the two-step method; and regression on a raw difference score when the differential predicts a dependent variable.

Method

I conducted three least squares regression analyses of a simulated dependent variable in order to test the effects of the correlation between the pre-score and the covariate. In these analyses, (a) the index of change and the covariate; (b) the post-score, pre-score, and covariate; and (c) the raw difference score and covariate predicted the dependent variable. The model had an intercept of 0. I compared across the three methods (a, b, and c) the regression of the dependent variable on the differential, its regression on the differential controlling for the covariate, and its regression on the covariate controlling for the differential.

The first comparison tests whether, in the absence of the covariate, the two-step method and the APV produce identical regression slopes. This comparison includes the regression on the raw difference score for additional reference. The second comparison tests whether the addition of the covariate biases the results of the two-step method and regression on the raw difference

score in relation to the APV. The third comparison evaluates the extent of bias to the regression on the covariate as a function of its correlation with the pre-score.

Results and Discussion

First, the regression of the dependent variable on the differential—without controlling for the covariate—was insensitive to the correlation between the pre-score and the covariate (i.e., it did not vary across levels of $r_{pre,cov}$; Figure 5). Although regression of the dependent variable on the index of change ($\bar{B} = .18$; $\overline{R^2} = .03$) and on the post-score and pre-score ($\bar{B} = .18$; $\overline{R^2} = .12$) produced identical slopes,[8] the latter regression explained more variance. The higher R^2 from the two-step method is due to the partial-r of the pre-score predicting the dependent variable, which is marked with shaded squares in Figure 5. Relative to the regression on the index of change, regression on the raw difference score had significantly different slopes at all values of $r_{pre,cov}$ ($\bar{B} = -.04$) and had lower explained variance ($\overline{R^2} = .00$).

Second, I evaluated the effects of adding the covariate (Figure 6). The regression of the dependent variable on the index of change, controlling for the covariate, produced regression slopes that increased from $B = .06$ to $B = .14$. The regression on the post-score, controlling for the pre-score and the covariate, produced slopes that increased from $B = .06$ to $B = .18$, which were significantly different than those of the APV for $r_{pre,cov} > .07$. Recall that the correct comparison between these methods is between regression on the index of change, controlling for the covariate, and regression on the post-score, controlling for the pre-score and the covariate (reference B_2 and B_3, respectively, in Figure 4). Finally, regression on the raw difference score, controlling for the covariate, produced slopes that increased from $B = -.13$ to $B = .11$, which were

---

[8] Estimates of B were identical to more than 10 decimal places.

significantly different than those of the APV for all values of $r_{pre,cov}$. Figure 6 shows the regression of the dependent variable on the pre-score for additional reference.

Finally, I examined the regression of the dependent variable on the covariate controlling for the differential. When controlling for the index of change, the slope remained approximately static across all $r_{pre,cov}$ values ($\bar{B} = .28$; Figure 7). When controlling for the pre-score and post-score, slopes followed a logarithmic trend, declining from $B = .28$ to $B = -.05$, and were significantly different than those of the APV for $r_{pre,cov} > .03$. When controlling for the raw difference score, the relationship between $r_{pre,cov}$ and B appeared to follow a second-degree polynomial trend, achieving its minimum value ($B = .30$) at moderate levels of multicollinearity and its maximum values ($B \approx .35$) at both extremes. These slopes were significantly different than those of the APV for all values of $r_{pre,cov}$. Notably, R^2 values of all three methods converged as $r_{pre,cov}$ approached unity (Figure 8).

These results show that when the differential solely predicts a dependent variable, the two-step method produces regression slopes that match those of the APV, while regression on the raw difference score produces significantly different slopes. Both the two-step method and regression on the raw difference score bias R^2 relative to the APV. The addition of a covariate biases the results of both the two-step method and regression on a raw difference score. As the correlation between the pre-score and the covariate increases, the slope of the regression of the dependent variable on the post-score, controlling for the pre-score and the covariate, diverges from that of the APV, and the slope of its regression on the raw difference score, controlling for the covariate, converges with that of the APV. At lower degrees of correlation between the pre-score and the covariate, the two-step method markedly overestimates R^2 relative to the APV.

General Discussion

Often, statistical models require the computation of the difference between (or differential of) a pre-score and post-score. Cohen and Cohen (1983) describe the APV as a useful method to compute the residual variance of a post-score controlling for a pre-score. Such a computation can be necessary when the pre-score is correlated with a covariate, and the relationship between the differential and the covariate is of interest to a researcher. As a bulk of literature has argued, simply calculating the raw difference between the two variables by subtracting one from the other can yield unreliable results (Cohen & Cohen, 1983; Griffin, et al., 2008; Jamieson & Howk, 1992). When the pre-score is correlated with the covariate, using the raw difference score does not account for this correlation. Rather than using the raw difference score, researchers have used the APV to remove the variance of the pre-score from the post-score. By controlling for this variance, the APV specifically addresses the problem of multicollinearity that can occur when using raw difference scores.

This study compared the APV with two alternative methods for computing difference scores. One method, the two-step method, is a variant of the APV, which I have encountered mostly in studies of risk information seeking and processing. The other method uses a raw difference score. Comparisons among these three methods address the necessity of computing a differential as the partial variance of a post-score controlling for a pre-score per the APV.

Analyses of simulated data showed that when a differential is either a dependent or independent variable in a regression model, regression slopes are sensitive to the magnitude of correlation between the pre-score and the covariate. Specifically, using the two-step method or the raw difference score produced significantly different slopes that did the APV at most or all degrees of correlation.

This study focused on the APV in models of information seeking and processing. However, the approach may find use in a variety of communication models that require computation of the difference between two cross-sectional or longitudinal variables. Consider, for example, third-person perceptions, in which people rate themselves as less influenced than others by the effects of media (Davison, 1983). Researchers often measure this perception by presenting study participants with a persuasive message and then asking participants to estimate the amount of influence it would have on them and on other people. Then, researchers compare the difference between the two estimations using t-tests or ANOVA (e.g., Davison, 1983; Gunther & Thorson, 1992; Lee, 2009; Perloff, 1989). As I described in the literature review, such analyses benefit from retaining the scale of the original items. The APV does not retain this scale, and may not be the ideal approach in such research settings. Furthermore, since these comparisons of mean differences do not consider the correlation between the pre-score and a covariate, the problem with multicollinearity does not arise.

However, some recent research of third-person perceptions has used a raw difference score as an independent variable in a regression model in which the pre-score correlated with a covariate (Ho, Detenber, Malik, & Neo, 2012; Lewis, Watson, & Tay, 2007; Shin & Kim, 2011; Wei, Chia, & Lo, 2011). As the current study shows, regression on a raw difference score can produce biased estimates at most degrees of correlation between a pre-score and a covariate. Researchers have also used a raw difference score as a dependent variable (Shin & Kim, 2011; Wei, Lo, & Lu, 2010), and at least one study (Wei, et al., 2010) analyzed a pre-score and post-score separately in accordance with the two-step method. As the current results suggest, either method can bias regression slopes significantly relative to the APV. For multivariate analyses that include a pre-score, post-score, and at least one additional variable that correlates with the

pre-score and is not dependent on the post-score, I recommend researchers use the APV to compute an index of change, which they can analyze as either a dependent or independent variable.

Limitations

This study had several limitations. First, I used a very large sample ($N = 100,000$) in order to minimize estimation errors and clarify differences. However, in most empirical studies, such a large sample would be impractical. In a real-world sample, the differences among the three methods might not be statistically significant.

Furthermore, with any sample size, the results might not be practically significant. When the differential was regressed on the covariate, the two-step method produced similar slopes to those of the APV at roughly $r_{pre,cov} < .50$, and regression of a raw difference score produced similar slopes to those of the APV at roughly $r_{pre,cov} < .10$ (see Figure 2). When the differential and the covariate are independent variables in a regression model, the two step-method produced similar slopes to those of the APV at all $r_{pre,cov}$ values, and using a raw difference score produced similar slopes to those of the APV at roughly $r_{pre,cov} > .80$ (see Figure 6). Thus, I would recommend that researchers generally avoid using raw difference scores in regression models; however, if imprecision is tolerable, researchers can achieve acceptable results with the two-step method when the pre-score and covariate have a weak-to-moderate correlation.

Second, this study treated the APV as the most statistically appropriate approach to computing differentials. One justification for using the APV is that it reduces problems related to floor and ceiling effects. However, in ANCOVA, between-group differences on a pre-score will bias post-score means in the direction of the difference, where a higher pre-score results in greater positive change (Jamieson, 1999). In other words, the correct interpretation of a

differential may depend on the level of the pre-score; though, researchers have not determined

the level of pre-score at which this bias is problematic. Furthermore, some studies eschew

separate pre-score and post-score measures entirely, and measure the differentiating construct

directly with Likert-type scale items (e.g., ter Huurne, 2008; Trumbo, 2002). For example,

Trumbo (2002) asked respondents to indicate their agreement with the statement, "The

information I have at this time meets all of my needs for knowing about the issue…." This scale

thus measures information insufficiency in a single step, and obviates the need to compute a

differential. Whether such measurement has greater validity than the APV is unknown.

Additionally, the separate analyses of a post-score and pre-score can address some of the

complexities of communication processes that the APV would conceal. The benefit of the two-

step method is that it allows researchers to examine separately the antecedents of both the pre-

score and post-score, which was an original purpose of the method (R. Griffin, personal

communication). Models that draw this distinction can produce valuable results, and certain

research contexts can benefit from using the two-step method. For example, Kahlor (2007)

reported that current knowledge (the pre-sore) positively predicted desired knowledge (the post-

score), which then positively predicted information seeking intention (the dependent variable).

This finding comports with the adage, "the more you know, the more you know you don't

know," and extends it to include, "and the more you know you don't know, the more you want to

know." However, Kahlor's model labels the post-score as the differential, assuming that its

earlier regression on the pre-score resulted in its partial variance. Thus, I recommend that

researchers who opt for the two-step method understand exactly what their model specifies and

also consider the effects of multicollinearity. If a differential is an independent or dependent

variable in a regression model, and the pre-score has a moderate or high correlation with a

covariate, researchers should supplement their results with a separate analysis of an index of change per the APV.

A third limitation concerns the object of change in analyses of partial variance. Researchers describe the APV as a means to compute an index of change for the same measurement at two different times (Cohen & Cohen, 1983; Jamieson, 1995, 1999; Jamieson & Howk, 1992). However, information insufficiency and third-person perceptions are each a differential of two cross-sectional variables. Thus, the calculation of information insufficiency and third-person perceptions per Cohen and Cohen's approach does not bear a true index of change, since no change occurs. I am unaware of research that has compared APV of cross-sectional data with APV of longitudinal data. Such a comparison could be informative.

Finally, this study suggests that the correlation between a pre-score and a covariate biases results of using both the two-step method and a raw difference score. However, this study did not determine the threshold at which multicollinearity is problematic. Although regression slopes were significantly different at most values of $r_{pre,cov}$, some of the smaller differences might not have practical significance.

Appendix A: Data Simulation

Distribution

Following Mayes (2010), I used the Excel function "=NORM.INV(RAND(),0,1)" to

generate an array of 4 (variables) x 100,000 (observations) of normal, pseudorandom data. The

data could not be completely random, since each data point in a normal distribution has a higher

probability of being close to the mean than far from the mean. The "0,1" portion of the Excel

function specifies that the variables will have a mean of 0.00 and a standard deviation of 1.00.

The benefit of constraining standard deviations to 1.00 is that it controls for effects due to

heterogeneity of variance, and the correlation and covariance matrices are equivalent.

Structure

Next, I generated the Cholesky decomposition (C) of a reference correlation matrix (see

Table 2), which for any positive definite matrix (M) is a lower triangular matrix such that:

$$M = CC'$$

where C' is the complex conjugate transpose of C. The reference correlation matrix included a

pre-score, a post-score, a covariate, and a dependent variable. Thus, the matrix had dimensions

of 4 x 4, as did its Cholesky decomposition.

I multiplied the 4 x 100,000 array of simulated data by the transpose of the Cholesky

decomposition. The resulting array contained normal, pseudorandom data with variable means of

approximately 0.00, standard deviations of 1.00, and a correlation structure approximately equal

to the reference matrix. Based on pre-score and post-score values, I calculated the index of

change and difference score, resulting in an array of 6 x 100,000.

Finally, I manipulated the zero-order correlation between the pre-score and the covariate

($r_{cpre,cov}$) in the reference table and re-estimated the model in increments of .05, ranging from

$r_{cpre,cov} = 0$ to $r_{cpre,cov} = .95$. I constructed a table of unstandardized regression slopes (B) and R^2

values, with each row comprising data corresponding to a discreet $r_{cpre,cov}$ increment.

Appendix B: Comparing Slopes

The following equation gives an approximate *t*-value estimate of the difference between two regression slopes from independent samples (Wuensch, 2007).

$$t = \frac{b_1 - b_2}{s_{b_1 - b_2}} \tag{1}$$

This equation computes the difference between the two slopes divided by the standard error of the difference between the slopes (df = n - 4), where b_1 and b_2 are the regression slopes for group 1 and group 2, and $s_{b\text{-}b2}$ is the standard error of the difference between slopes. I calculated the standard error of the difference between slopes (the denominator) with the following equation,

$$s_{b_1 - b_2} = \sqrt{s_{b1}^2 + s_{b2}^2} \tag{2}$$

where s_{b1} and s_{b2} are the standard errors of the individual slopes. Thus,

$$t = \frac{b_1 - b_2}{\sqrt{s_{b1}^2 + s_{b2}^2}} \tag{3}$$

I used equation 3 to compare slopes at different $r_{pre,cov}$ values, and determined approximate values at which $t > 1.96$. I report for each comparison the range of $r_{pre,cov}$ values that produce a significant difference in regression slopes.

References

Chaiken, S. (1980). Heuristic versus systematic information-processing and the use of source

    versus message cues in persuasion. *Journal of Personality and Social Psychology, 39*(5),

    752-766.

Chen, S. & Chaiken, S. (1999). The heuristic-systematic model in its broader context. In S.

    Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology*. New York:

    Guilford Press.

Cohen, J. & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the*

    behavioral sciences. Hillsdale, NJ: Lawrence Erlbaum.

Cribbie, R. & Jamieson, J. (2004). Decreases in posttest variance and the measurement of

    change. *Methods of Pyschological Research Online, 9*(1), 37-55.

Davison, W. P. (1983). The third-person effect in communication. *Public Opinion Quarterly,*

    47(1), 1-15.

Dennis, R. L., Byun, D. W., Novak, J. H., Galluppi, K. J., Coats, C. J., & Vouk, M. A. (1996).

    The next generation of integrated air quality modeling: EPA's Models-3. *Atmospheric*

    Environment, 30(12), 1925-1938.

Garber, A. M. & Sox, H. C. (2010). The role of costs in comparative effectiveness research.

    Health Affairs, 29(10), 1805-1811.

Griffin, R. J., Dunwoody, S., & Neuwirth, K. (1999). Proposed model of the relationship of risk

    information seeking and processing to the development of preventive behaviors.

    Environmental Research, 80(2), S230-S245.

Griffin, R. J., Neuwirth, K., Dunwoody, S., & Giese, J. (2004). Information sufficiency and risk

    communication. *Media Psychology, 6*(1), 23-61.

Griffin, R. J., Yang, Z., ter Hume, E., Boerner, F., Ortiz, S., & Dunwoody, S. (2008). After the

flood: anger, attribution, and the seeking of information. *Science Communication, 29*(3),

285-315.

Gunther, A. C. & Thorson, E. (1992). Perceived persuasive effects of product commercials and

public-service announcements: Third-person effects in new domains. *Communication

Research, 19*(5), 574-596.

Ho, S. S., Detenber, B. H., Malik, S., & Neo, R. L. (2012). The roles of value predispositions,

communication, and third person perception on public support for censorship of films

with homosexual content. *Asian Journal of Communication, 22*(1), 78-97.

Holmgren, P. & Thuresson, T. (1998). Satellite remote sensing for forestry planning - A review.

Scandinavian Journal of Forest Research, 13(1), 90-110.

Hovick, S. R., Freimuth, V. S., Johnson-Turbes, A., & Chervin, D. D. (2011). Multiple health

risk perception and information processing among African Americans and Whites living

in poverty. *Risk Analysis, 31*(11), 1789-1799.

Jamieson, J. (1995). Measurement of change and the law of initial values: A computer simulation

study. *Educational and Psychological Measurement, 55*(1), 38-46.

Jamieson, J. (1999). Dealing with baseline differences: two principles and two dilemmas.

International Journal of Psychophysiology, 31(2), 155-161.

Jamieson, J. & Howk, S. (1992). The law of initial values: A four factor theory. *International

Journal of Psychophysiology, 12*(1), 53-61.

Kahlor, L. (2007). An augmented risk information seeking model: the case of global warming.

Media Psychology, 10(3), 414-435.

Kassim, N. A. (2010). Information needs of Malaysian Bumiputera would-be entrepreneurs. *Malaysian Journal of Library & Information Science, 15*(2), 57-69.

Kowatsch, T. & Maass, W. (2010). In-store consumer behavior: How mobile recommendation agents influence usage intentions, product purchases, and store preferences. *Computers in Human Behavior, 26*(4), 697-704.

Lee, F. L. F. (2009). The prevention effect of third-person perception: A study on the perceived and actual influence of polls. *Mass Communication and Society, 13*(1), 87-110.

Lewis, I., Watson, B., & Tay, R. (2007). Examining the effectiveness of physical threats in road safety advertising: The role of the third-person effect, gender, and age. *Transportation Research Part F-Traffic Psychology and Behaviour, 10*(1), 48-60.

Mayes, T. R. (2010). Generating correlated normally distributed random numbers in Excel. *The Journal of Financial Modeling and Educational Technology, 1*(1), 59-66.

McAlister, A. L. (2001). Moral disengagement: Measurement and modification. *Journal of Peace Research, 38*(1), 87-99.

Perloff, R. M. (1989). Ego-involvement and the third-person effect of television news coverage. *Communication Research, 16*(2), 236-262.

Shin, D. H. & Kim, J. K. (2011). Alcohol product placements and the third-person effect. *Television & New Media, 12*(5), 412-440.

ter Huurne, E. (2008, December 7-10). *To know or not to know? A framework of risk information seeking in the sphere of industrial risks.* Paper presented at the Society for Risk Analysis Annual Meeting, Boston, MA.

ter Huurne, E. F. J., Griffin, R. J., & Gutteling, J. M. (2009). Risk information seeking among

      US and Dutch residents: An application of the model of risk information seeking and

      processing. *Science Communication, 31*(2), 215-237.

Trumbo, C. W. (2002). Information processing and risk perception: An adaptation of the

      heuristic-systematic model. *Journal of Communication, 52*(2), 367-382.

Wei, R., Chia, S. C., & Lo, V. H. (2011). Third-person effect and hostile media perception

      influences on voter attitudes toward polls in the 2008 US presidential election.

      International Journal of Public Opinion Research, 23(2), 169-190.

Wei, R., Lo, V. H., & Lu, H. Y. (2010). The third-person effect of tainted food product recall

      news: Examining the role of credibility, attention, and elaboration for college students in

      Taiwan. *Journalism & Mass Communication Quarterly, 87*(3-4), 598-614.

Wuensch, K. L. (2007, Comparing correlation coefficients, slopes, and intercepts. Retrieved

      from core.ecu.edu/psyc/wuenschk/docs30/CompareCorrCoeff.doc

Yang, Z. J., McComas, K. A., Gay, G., Leonard, J. P., Dannenberg, A. J., & Dillon, H. (2011).

      Information seeking related to clinical trial enrollment. *Communication Research, 38*(6),

      856-882.

Zacharakis, A. L. & Meyer, G. D. (1998). A lack of insight: Do venture capitalists really

      understand their own decision process? *Journal of Business Venturing, 13*(1), 57-76.

Figure Captions

Figure 1. Regression of the differential on the covariate. Methods 1 and 2 have identical first steps. The regression of the index of change on the covariate (B_2) and the regression of the post-score, controlling for prescore, on the covariate (B_3) will have identical values only when the correlation between the prescore and the covariate ($r_{pre,cov}$) is equal to zero. Furthermore, B_2 will equal the regression of the raw difference score on the covariate (B_4) when the correlation between the prescore and post-score is equal to one (see footnote 4).

Figure 2. Regression of the differential on the covariate. The horizontal axis shows different levels of correlation between the pre-score and the covariate ($r_{pre,cov}$). The vertical axis shows the unstandardized slope (B) of the regression of the differential on the covariate. The data points indicate B of the three methods at each $r_{pre,cov}$ value. The regression of the post-score on the pre-score controls for the covariate and constitutes the first step of the two-step model (as integrated in the second step; see Step 2 in the second panel of Figure 1). The regression of the post-score on the covariate controls for the pre-score and constitutes, per the two-step method, the regression of the differential on the covariate. Direct comparisons should be made among the lines with non-shaded markers.

Figure 3. Regression of the differential on the covariate. The horizontal axis shows different levels of correlation between the pre-score and the covariate ($r_{pre,cov}$). The vertical axis shows the R^2 of the differential. The data points indicate R^2 of the three methods at each $r_{pre,cov}$ value. The regression of the post-score on the pre-score controls for the covariate and constitutes the first step of the two-step model (as integrated in the second step; see Step 2 in the second panel of Figure 1). The regression of the post-score on the covariate controls for the pre-score and constitutes, per the two-step method, the regression of the differential on the covariate. Direct comparisons should be made among the lines with non-shaded markers.

Figure 4. Regression of the dependent variable on the differential and the covariate. The regression of the dependent variable on the index of change (B_2) and the regression of the dependent variable on the post-score, controlling for prescore, (B_3) will have identical values only when the correlation between the prescore and the covariate ($r_{pre,cov}$) is equal to zero. Furthermore, B_2 will equal the regression of the dependent variable on the raw difference score (B_4) when the correlation between the prescore and post-score is equal to one (see footnote 4).

Figure 5. Regression of the dependent variable on the differential. The horizontal axis shows different levels of correlation between the pre-score and the covariate ($r_{pre,cov}$). The vertical axis shows the unstandardized slope (B) of the regression of the dependent variable on the differential. The data points indicate B of the three methods at each $r_{pre,cov}$ value. Note that pre-score controls for post-score and post-score controls for pre-score. Direct comparisons should be made among the lines with non-shaded markers.

Figure 6. Regression of the dependent variable on the differential, controlling for the covariate. The horizontal axis shows different levels of correlation between the pre-score and the covariate ($r_{pre,cov}$). The vertical axis shows the unstandardized slope (B) of the regression of the dependent
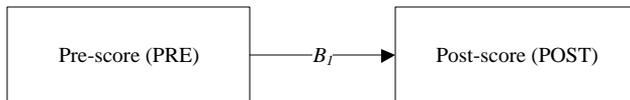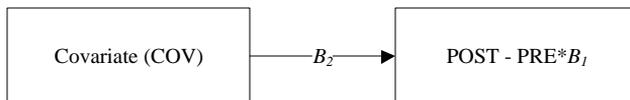
variable on the differential, controlling for the covariate. The data points indicate B of the three methods at each $r_{pre,cov}$ value. Note that, in addition to controlling for the covariate, pre-score controls for post-score and post-score controls for pre-score. Direct comparisons should be made among the lines with non-shaded markers.

Figure 7. Regression of the dependent variable on the covariate, controlling for the differential. The horizontal axis shows different levels of correlation between the pre-score and the covariate ($r_{pre,cov}$). The vertical axis shows the unstandardized slope (B) of the regression of the dependent variable on the covariate, controlling for the differential. The data points indicate B of the three methods at each $r_{pre,cov}$ value.

Figure 8. Regression of the dependent variable on the differential and covariate. The horizontal axis shows different levels of correlation between the pre-score and the covariate ($r_{pre,cov}$). The vertical axis shows the R^2 of the dependent variable. The data points indicate R^2 of the three methods at each $r_{pre,cov}$ value.
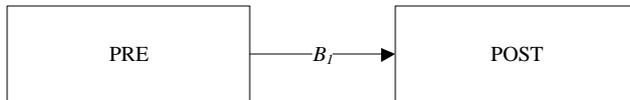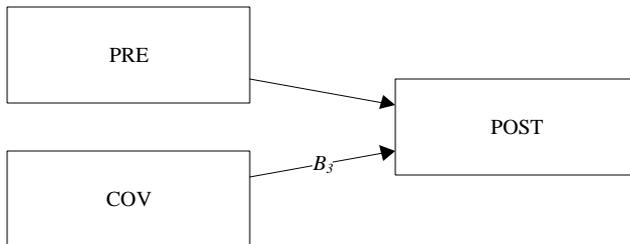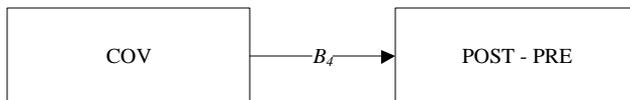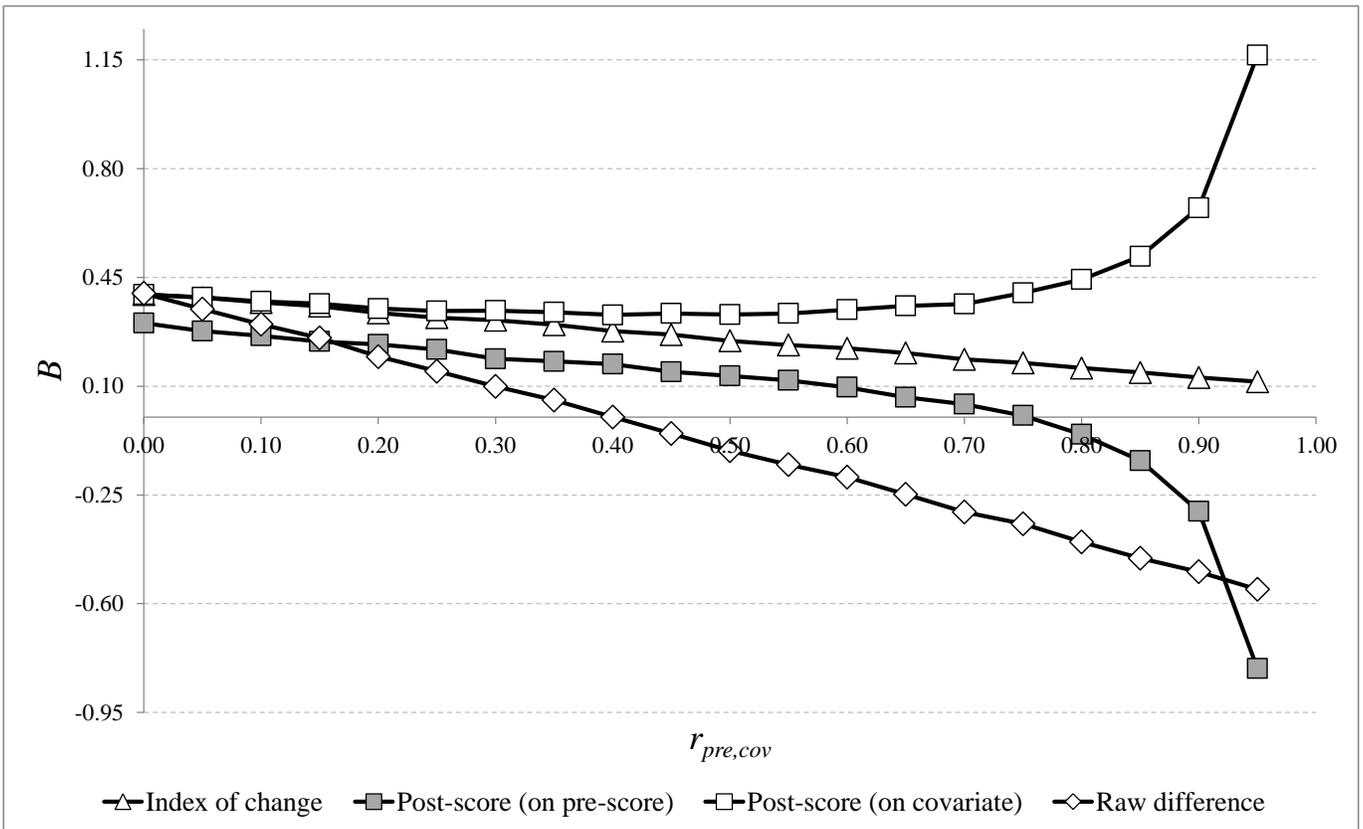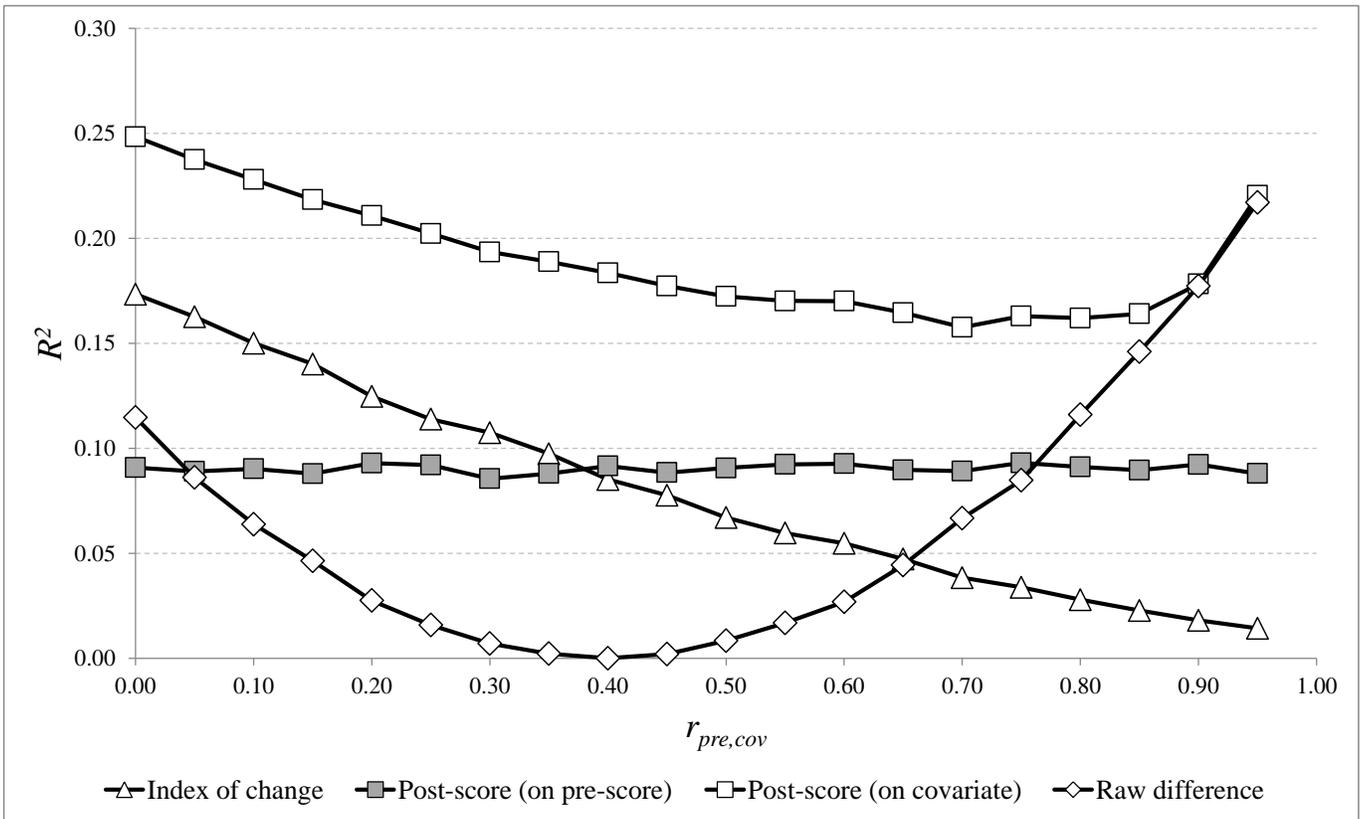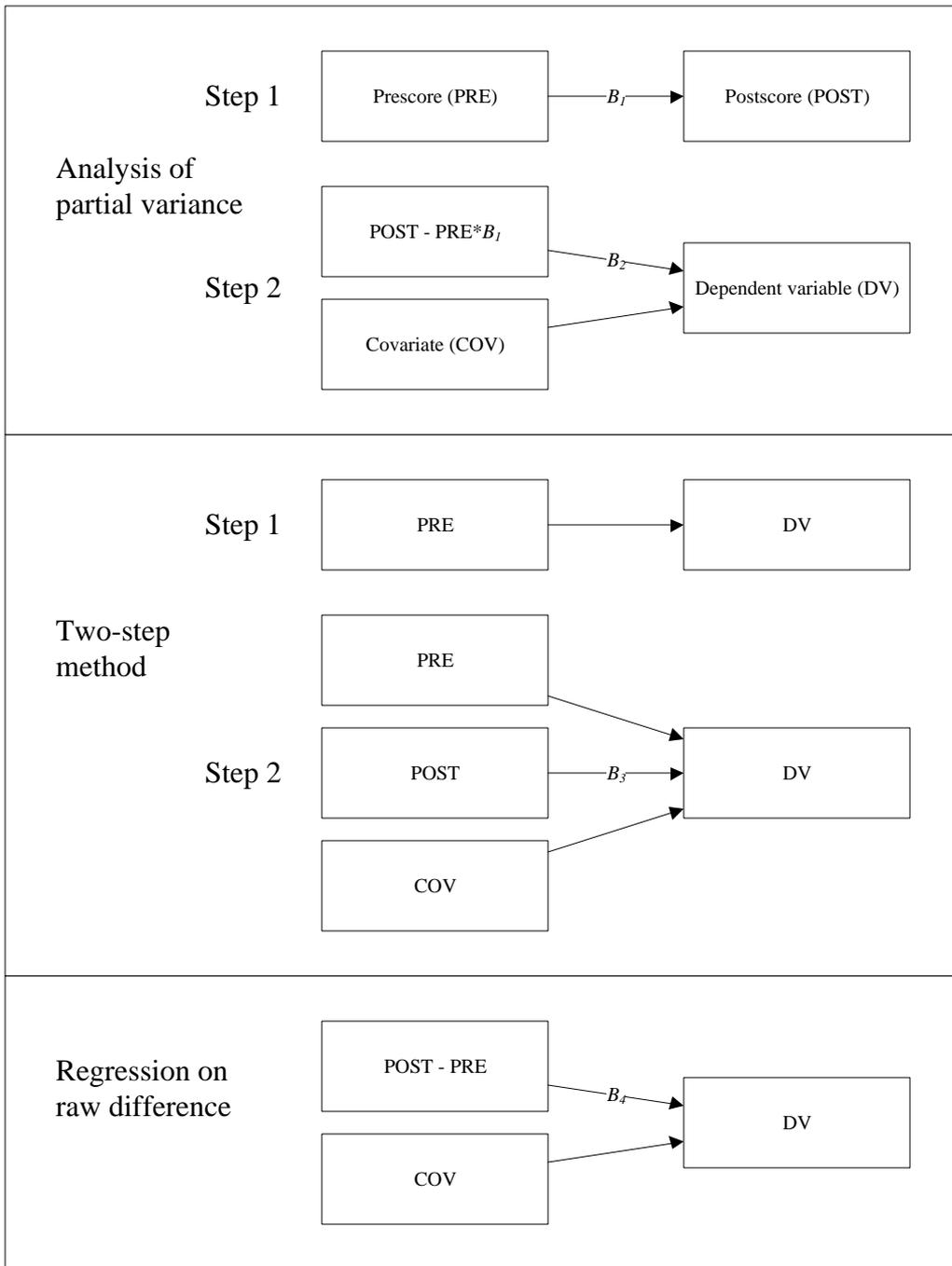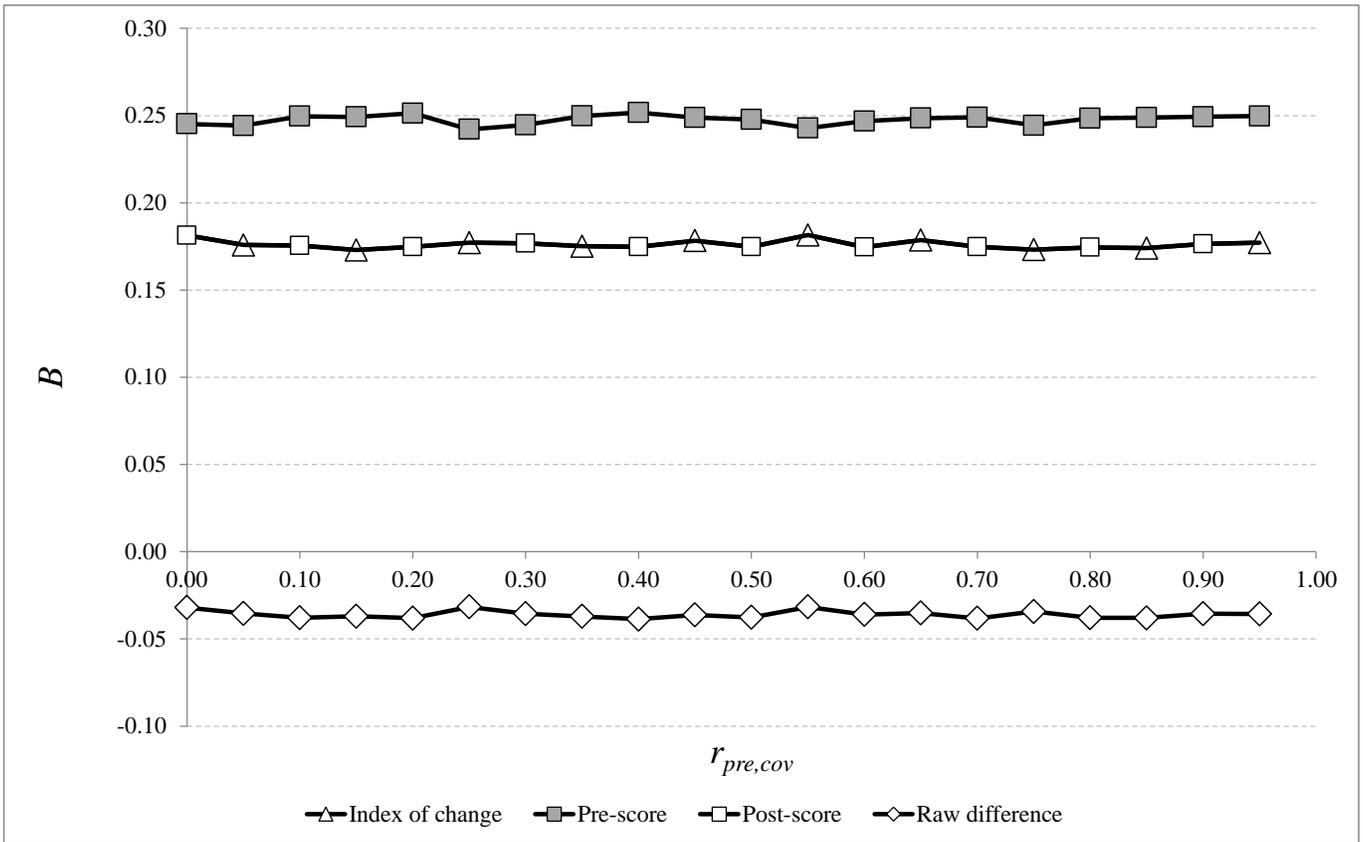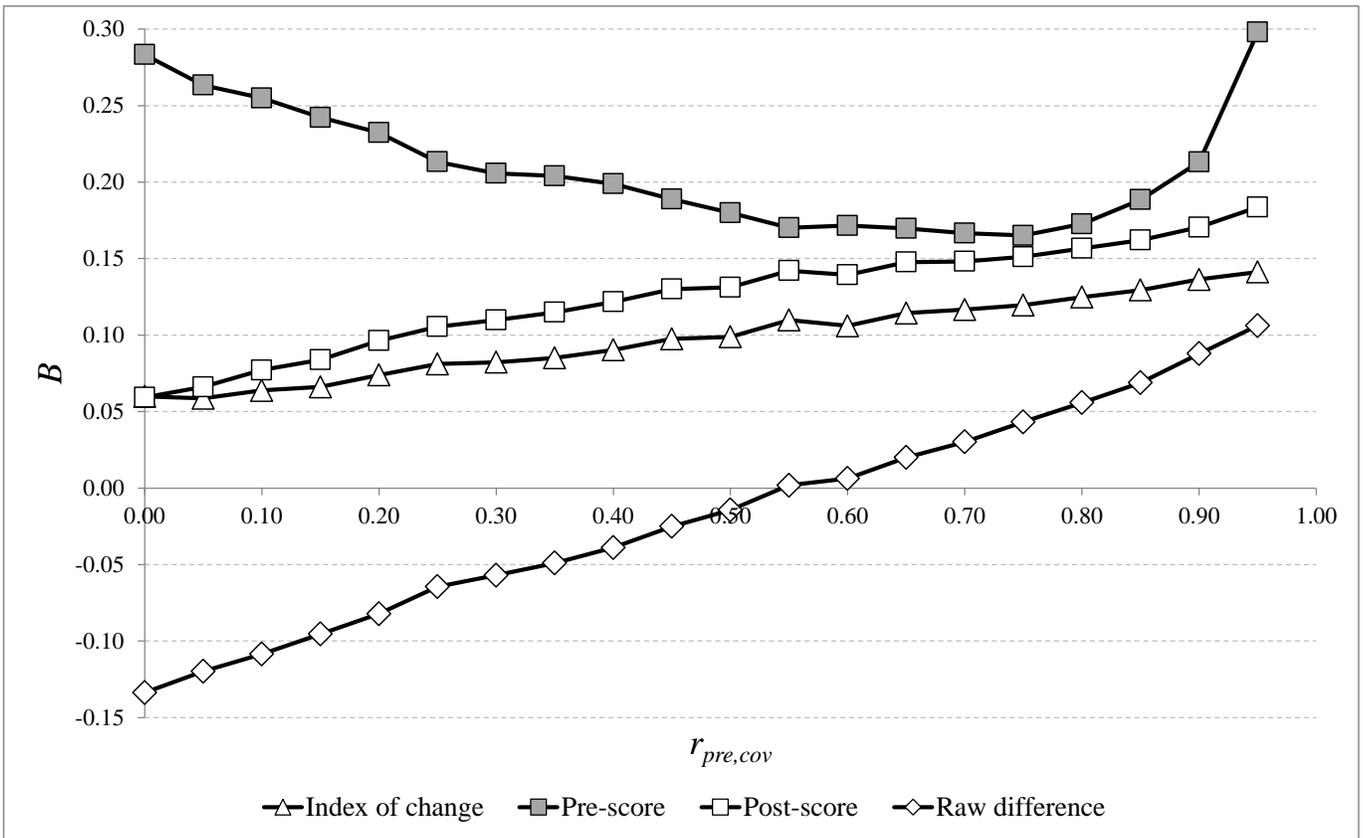
Figure 1

Figure 2

Figure 3
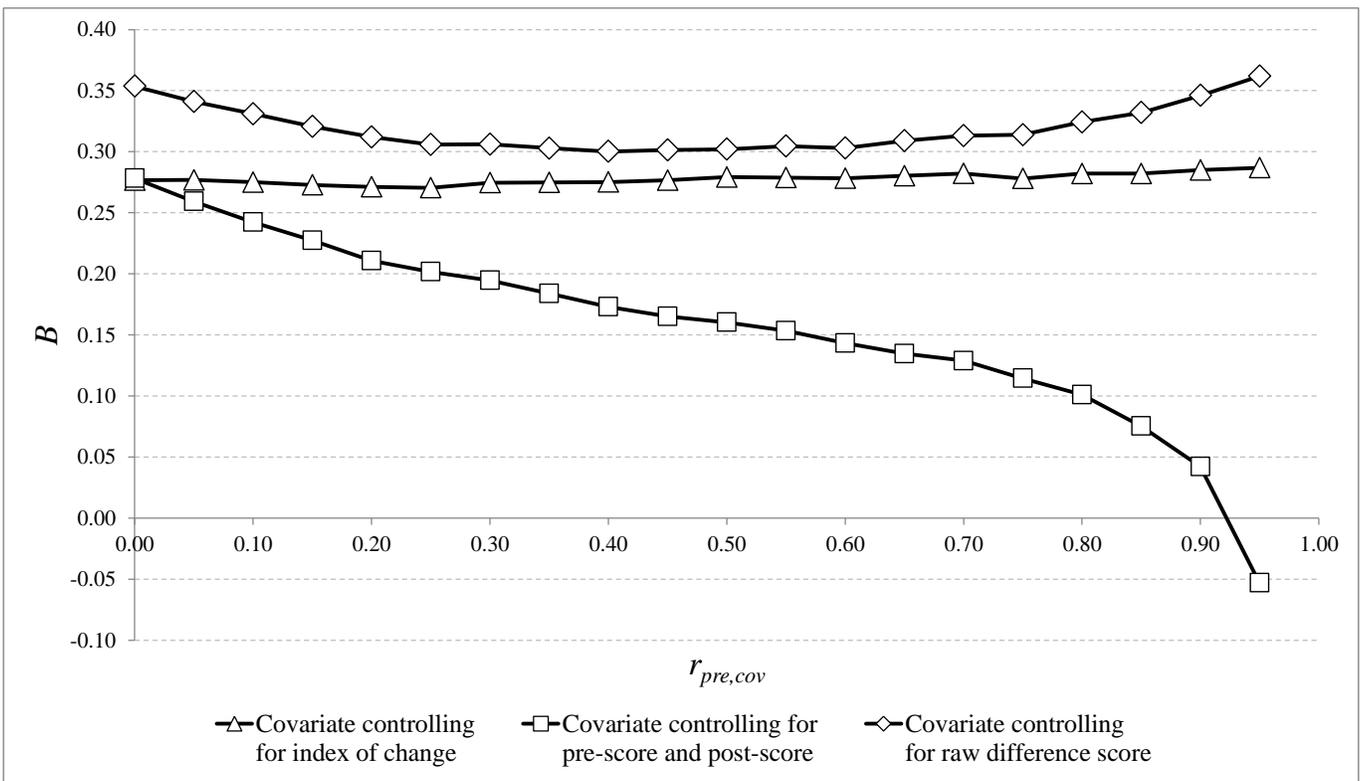
Figure 4

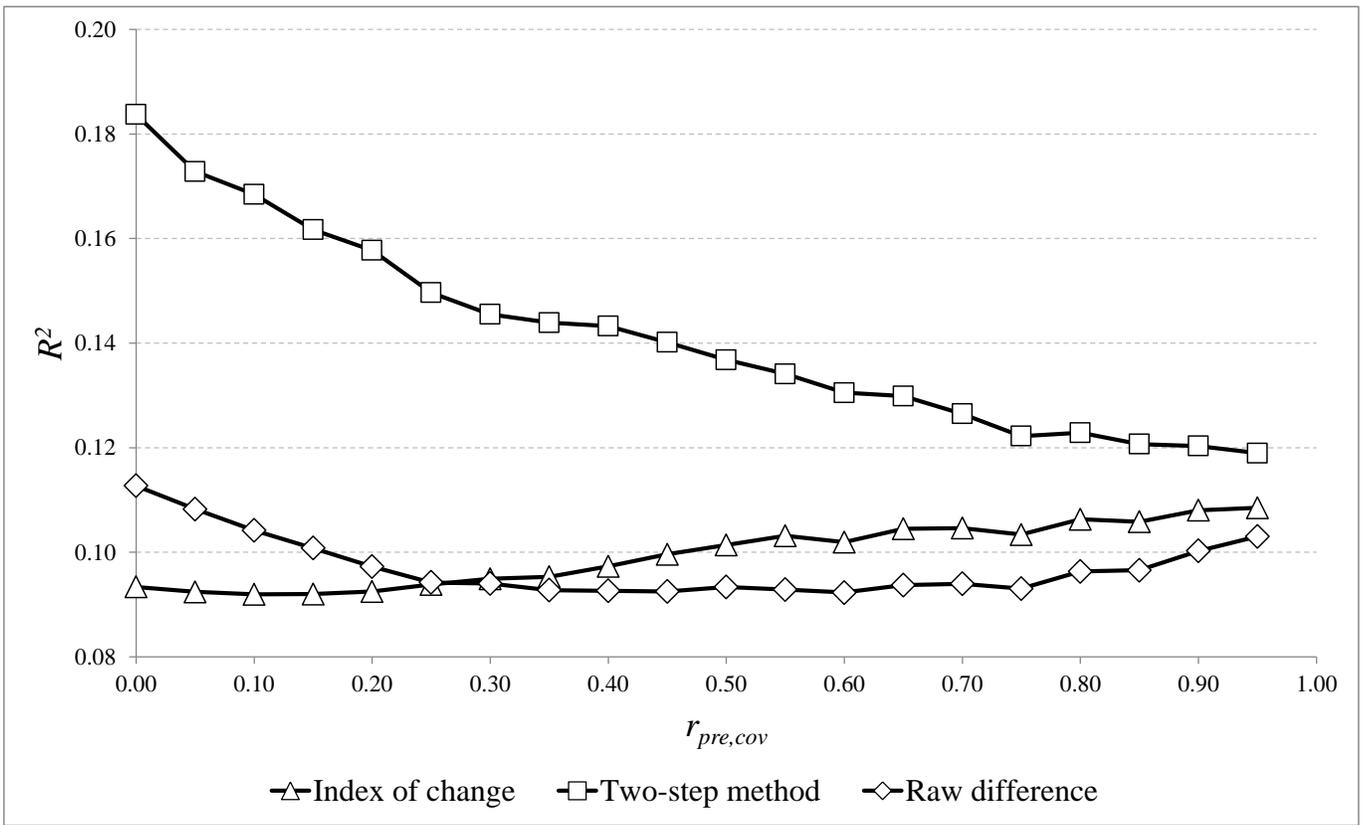Figure 5

Figure 6

Figure 7

Figure 8

Table 1

Examples of pre-scores, post-scores, and the resulting differential; covariates; and dependent variables in three research contexts.

Research context	Risk information seeking and processing	Third-person perception	Public opinion of war and peace
Example citation	Griffin et al. (1998)	Ho et al. (2012)	McAlister (2001)
Pre-score	Perceived current knowledge	One's own susceptibility to media influences	Past moral disengagement
Post-score	Desired knowledge	Others' susceptibility to media influences	Current moral disengagement
Differential	Information insufficiency	Third-person perception	Change in moral disengagement
Covariate	Affective response to a perceived risk	Intrinsic religiosity	Exposure to message supporting moral engagement
Dependent variable	Risk information seeking intention	Support of censorship	Opposition to war

Note. The covariate can be any variable(s) in a statistical model that either predict(s) the differential or moderate(s) its effect on the dependent variable. Figures 1 and 4 depict these two configurations, respectively.

Table 2

Simulation reference distribution and covariance matrix

	Distribution				Correlations		
	M	*SD*	skewness	kurtosis	1	2	3
1. Pre-score	0.00	1.00	0.00	0.00	--		
2. Post-score	0.00	1.00	0.00	0.00	.30	--	
3. Covariate	0.00	1.00	0.00	0.00	$r_{pre,cov}$	.40	--
4. Dependent variable	0.00	1.00	0.00	0.00	.30	.25	.30

Note. Values of $r_{pre,cov}$ ranged from 0 to .95 in increments of .05. The remaining coefficients reference those of Kahlor (2007) in order to achieve a degree of realism in the simulation.