ANNUAL REVIEWS

Annual Review of Clinical Psychology

The Questionable Practice of Partialing to Refine Scores on and Inferences About Measures of Psychological Constructs

Rick H. Hoyle,[1] Donald R. Lynam,[2] Joshua D. Miller,[3] and Jolynn Pek[4]

[1]Department of Psychology and Neuroscience, Duke University, Durham, North Carolina, USA; email: rhoyle@duke.edu

[2]Department of Psychological Sciences, Purdue University, West Lafayette, Indiana, USA

[3]Department of Psychology, University of Georgia, Athens, Georgia, USA

[4]Department of Psychology, Ohio State University, Columbus, Ohio, USA

Annu. Rev. Clin. Psychol. 2023. 19:155–76

First published as a Review in Advance on February 7, 2023

The *Annual Review of Clinical Psychology* is online at clinpsy.annualreviews.org

https://doi.org/10.1146/annurev-clinpsy-071720-015436

Keywords

construct validity, residualized variables, statistical control, semipartial effects, questionable measurement practice

Abstract

Partialing is a statistical approach researchers use with the goal of removing extraneous variance from a variable before examining its association with other variables. Controlling for confounds through analysis of covariance or multiple regression analysis and residualizing variables for use in subsequent analyses are common approaches to partialing in clinical research. Despite its intuitive appeal, partialing is fraught with undesirable consequences when predictors are correlated. After describing effects of partialing on variables, we review analytic approaches commonly used in clinical research to make inferences about the nature and effects of partialed variables. We then use two simulations to show how partialing can distort variables and their relations with other variables. Having concluded that, with rare exception, partialing is ill-advised, we offer recommendations for reducing or eliminating problematic uses of partialing. We conclude that the best alternative to partialing is to define and measure constructs so that it is not needed.

Contents

INTRODUCTION 156
GENERAL APPROACHES TO PARTIALING 157
MOTIVATIONS FOR PARTIALING 157
POTENTIAL CONSEQUENCES OF PARTIALING 158
PARTIALING IN PRACTICE 160
EFFECTS OF COLLINEARITY, UNRELIABILITY, AND MAGNITUDE OF ASSOCIATIONS 162
RESIDUALIZATION AND CONSTRUCT OBFUSCATION 165
RECOMMENDATIONS 168
Theory-Informed Statistical Modeling 168
Flexible Analytic Approaches Plus Simple Associations 169
Reducing the Need to Partial Through Better Measurement 170
CONCLUSION 172

INTRODUCTION

Ideally, variance in scores on measures of psychological constructs would be attributable exclusively, or at least primarily, to a well-defined construct users intend to measure. The isolation of constructs through such measures supports meaningful inferences about their nature and associations with other constructs. In practice, scores on measures of psychological constructs inadequately isolate the construct they were intended to measure from other constructs (e.g., Santor et al. 2006, Weidman et al. 2017). Thus, isolation often is attempted through partialing, the use of statistical strategies such as analysis of covariance (ANCOVA) with quasi-experimental designs and multiple regression analysis to remove extraneous variance from a variable for the purpose of isolating its associations with other variables. "Partialing out"—that is, controlling for extraneous sources of variance in scores to better represent the targeted construct—is commonplace in clinical psychological science (e.g., Cândea & Szentagotai-Tăta 2018, Fite et al. 2009). The practice persists despite concerns having been raised repeatedly about unintended and unknown consequences of partialing (e.g., Cohen & Cohen 1983, Lynam et al. 2006, Meehl 1975, Miller & Chapman 2001, Schneider et al. 2015). Rather than clarifying the nature of constructs and mechanisms that underlie associations between variables, partialing often clouds inferences by removing variance that reflects core features of the construct and increasing the influence of both random and systematic sources of measurement error.

In this review, we describe and illustrate the impact of partialing on the meaning of variables and the estimation of associations between them, and we conclude that partialing is in many cases a questionable measurement practice (Flake & Fried 2020). First, we describe two general approaches to partialing in clinical research, motivations for their use, and the unintended ways they can affect the meaning of variables that result from their use. Second, we review common uses of partialing in clinical psychological science. Third, we present findings from two simulations that highlight the complications and distortions partialing can create, especially in the presence of unreliability and patterns of correlation likely to produce suppression. Fourth, we recommend alternative approaches for isolating constructs when measures of them are imprecise; we devote particular attention to the importance of practices that anticipate and attempt to address the concerns that necessitate partialing when developing and refining measures. Our aim is to show that partialing for the purpose of isolating constructs rarely yields findings that are interpretable with reference to the constructs of interest.

GENERAL APPROACHES TO PARTIALING

In clinical science, partialing is most evident in models that simultaneously estimate the association of a set of independent variables (e.g., depression and anxiety) with a single outcome variable (e.g., treatment seeking). We focus on multiple regression analysis with quasi-continuous outcomes estimated using ordinary least squares—the workhorse in clinical science. As detailed below, such models are frequently used to estimate the so-called unique association of each independent variable with an outcome. They might also be used to control for, or eliminate the influence of, variables that are irrelevant to the research question (i.e., covariates) but are plausible alternative explanations for associations involving the variables of interest. In the simplest case, these models take the form

$$y = b_0 + b_1 x_1 + b_2 x_2 + e, \tag{1.}$$

where x_2 is a variable that is partialed from x_1, the focal variable, so that x_1's association with y cannot be attributed to overlap with x_2. The variable e represents random error in y. Modeled in this way, associations involving x_1 are assumed to be attributable to a refined version of x_1 that is free of contamination from x_2.

An alternative form of partialing, which is accomplished outside the prediction context, focuses instead on producing a freestanding, partialed variable assumed to better represent the construct it was intended to measure. Partialing of this type is typically done using simple regression analysis of the form

$$x_1 = b_0 + b_1 x_2 + e, \tag{2.}$$

where, as before, x_2 is a variable that is partialed from x_1, the variable representing the construct of interest. This model partitions variance in x_1 into two components—variance it shares with x_2, reflected in b_1, and variance it does not, reflected in e. The variable e is the partialed form of x_1 (plus random error), which is assumed to reflect a form of the construct x_1 represents that is uncontaminated by x_2. Although partialed variables could, in principle, be produced by removing shared variance with covariates as in the prediction context, they typically are produced by removing variance shared with variables representing similar—often highly similar—constructs.

MOTIVATIONS FOR PARTIALING

Partialing is frequently motivated by one of two concerns, which map onto the two general approaches to partialing reviewed in the section above. One such concern, prediction, is the desire to evaluate the unique association of one or more independent variables in a set with an outcome variable as in Equation 1 (Jaccard et al. 2006). This rationale for partialing may take three forms. One rationale is to evaluate the association of a single focal variable with an outcome variable while accounting for its overlap with a set of independent variables; apart from partialing them from the focal independent variable, the associations of the other independent variables with the outcome variable are not of interest. The evaluation of the incremental validity of a new measure after taking into account one or more established measures is an example. A second rationale is to simultaneously evaluate the associations of a number of independent variables with an outcome variable; all associations are of interest, and each is estimated while controlling for all others in the set. Evaluation of the relative unique contribution of independent variables in a set with the goal of identifying the most important is an example. A third rationale involves a differentiation of independent variables in a set, with some designated as covariates and others as predictors. The goal is to estimate the associations between the predictors and the outcome variable after removing overlap with the nuisance covariates.

A second concern, construct validation, is the attempted refinement of a measured variable to better align with a conceptual definition of the construct it was designed to measure and to differentiate it from similar constructs (Strauss & Smith 2009). As described above, this form of partialing is typically accomplished using Equation 2. An assumption underlying this concern is that the variable as measured is so contaminated by extraneous influences that the construct's association with other variables cannot be determined. The contamination might result from inadvertent measurement of a construct other than the focal construct (e.g., neuroticism in measures of psychological distress; Batty et al. 2016) or the difficulty of designing measures that uniquely capture constructs that are highly similar (e.g., guilt and shame; Averill et al. 2002). In either case, partialing the confounding construct from the construct of interest aims to refine measures of the construct and clarify associations with other constructs.

POTENTIAL CONSEQUENCES OF PARTIALING

The goal of partialing is straightforward: Remove variance in a measured variable attributable to sources other than the intended construct while leaving intact variance attributable to the intended construct. Moreover, as described above, the analytic approaches to partialing are straightforward as well. Yet, for reasons we introduce in this section and detail in the remainder of the article, the likelihood of unintended and undesirable consequences when partialing is high. In this section we provide a conceptual description of how partialing works as a means of illustrating how it can go awry.

Figure 1 depicts potential sources of variance in three hypothetical variables. The variable labeled *A* is a measure of the construct of interest assumed to be contaminated by another construct. The variable labeled *B* is a measure of the construct to be partialed from *A*. And the variable labeled *C* is the partialed variable assumed to be a refined and uncontaminated version of *A*. We have labeled a set of hypothetical sources of variance to illustrate how partialing can produce a variable that is an unsatisfactory representation of the construct of interest.

Focusing first on *A*, we assume that the majority of the variance in scores is attributable to the focal construct but that, in addition to unreliability, variance in scores is attributable to a single identified contaminating construct and a number of unspecified additional, potentially unknown, constructs. Moving to *B*, we have assumed that the preponderance of variance is a reflection of the construct to be removed from *A* but that, beyond unreliability, some variance is attributable to the focal construct, with an additional share attributable to other constructs that may or may not correspond to that same share in *A*. Finally, compared with *A*, variance in *C* is reduced as a result of removing the influence of the contaminating factor.

A closer examination of *C* highlights some of the concerns in interpreting a partialed variable. Note that the proportion of variance in the partialed variable attributable to the focal construct is less than the proportion in the variable as measured. This can be attributed to three factors. First, we avoid the unrealistic (though unverifiable) assumption that the influence of the contaminating construct has been removed entirely. Second, we assume that the measure of the confounding construct is, to some extent, influenced by the focal construct and that this shared portion of the focal construct is removed through partialing. Finally, we assume that neither unreliability nor the influence of unspecified constructs (i.e., systematic error) has been removed (Westfall &

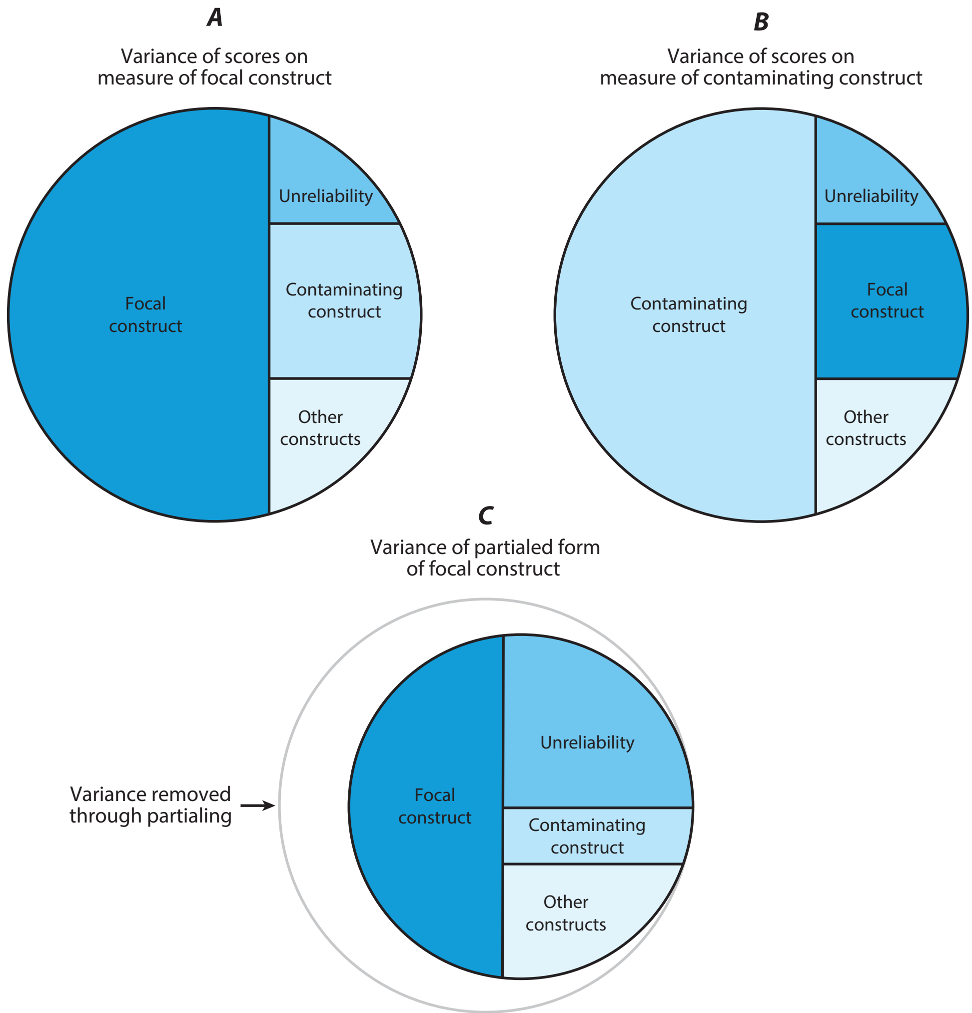


Figure 1

Hypothetical partitioning of variance of scores on measures of a focal construct, *A*; a contaminating construct to be partialed, *B*; and the partialed form of the focal construct, *C*.

Yarkoni 2016, Zinbarg et al. 2010).[1] The result is a variable that is less reliable and less reflective of the focal construct, though the influence of the contaminating construct has been diminished as intended. Importantly, variance that is shared and therefore removed through partialing may be essential to the full expression of the focal construct. Finally, other components outside of the focal and contaminating constructs may begin to exert more influence as they constitute larger portions of the partialed variable. In short, under conditions that are likely typical of measures

[1]The increase in unreliability can be seen in the formula for the reliability of the partialed variable $r_{CC} = \frac{r_{AA} - r_{AB}^2}{1 - r_{AB}^2}$, where *A*, *B*, and *C* correspond to the variables in **Figure 1**. When *A* and *B* are uncorrelated, the right side of the formula simplifies to r_{AA}. The variable r_{CC} is increasingly reduced relative to r_{AA} as the correlation between *A* and *B* increases. The formula assumes no shared systematic error variance.

involved in partialing analyses, the likelihood that a partialed variable is a valid expression of the focal construct free from the influence of contaminating constructs is much lower than assumed.

PARTIALING IN PRACTICE

There are many ways in which partialing through multiple regression analysis is used in clinical research.[2] Some are straightforward and carry little risk of misinterpretation; others pose serious interpretive problems. We review several of the most common uses below and highlight the interpretive issues that may arise in each case.

The most straightforward use of partialing in a predictive context involves controlling for confounds or alternative explanations for a relation between two focal variables, x_1 and y. In this case, the question is whether the observed relation between x_1 and y is reduced or eliminated when other variables are introduced. This approach has been widely used in studies examining the relation between maternal smoking during pregnancy and later disruptive behavior in the offspring. A relation between smoking during pregnancy and conduct problems and/or hyperactivity in offspring has been observed in multiple studies (Ruisch et al. 2018). Such a relation could be confounded by many other variables including low parental socioeconomic status, earlier pregnancy, poorer prenatal care, more stressful life events, maternal psychopathology (e.g., depression, antisocial behavior), and genetics. Analyses attempting to examine this causal relation typically include potential confounding variables in the regression to see if the relation between smoking and conduct problems in offspring can be eliminated (e.g., Maughan et al. 2001). In cases such as these, the question is primarily predictive: Does prenatal maternal smoking predict offspring behavior problems after controlling for potential confounding variables? Nothing is inferred about the construct or measure of maternal smoking.

Partialing is also frequently employed in longitudinal designs, either in analyses of change or in autoregressive paths. In the residual change score approach, earlier scores on a construct (e.g., Time 1 drug use) are partialed from later scores of the construct (e.g., Time 2 drug use). This partialed variable, which represents that part of the Time 2 construct that cannot be predicted from the Time 1 construct, is then predicted using other Time 1 constructs. Observed predictive relations are interpreted as predicting new drug use at Time 2. This approach was employed by Geisner et al. (2018) to examine the role of impulsivity and employment on postcollege alcohol use. In predicting later alcohol use, these authors controlled for previous use and a number of other covariates (i.e., age, gender, race, adaptability, negative affect, marijuana use, and financial stress) before examining the contributions of employment and impulsivity.

Multiple regression is also used in attempts to isolate the contribution of each of several predictors. In these analyses, multiple predictors are entered simultaneously to predict an outcome, and the partial coefficients (e.g., B's, β's, partial and semipartial correlation coefficients) are examined for significance. These analyses are more interpretively problematic than those described above. For example, Walton et al. (2018) examined the unique contributions of the NEO PI-R personality facets to several forms of psychopathology: substance use disorder, anxiety, and fear. The NEO PI-R is a self-report assessment of the five-factor model of personality that consists of five broad domains, each underlain by six facets. Walton et al. regressed each of the three psychopathology

[2]The concerns discussed here are relevant for uses of ANCOVA as well—specifically, cases in which a covariate is correlated with an independent variable as in quasi-experiments (Miller & Chapman 2001, Zinbarg et al. 2010). When ANCOVA is used to analyze data from randomized experiments, its effect is solely on the dependent variable (Darlington & Hayes 2017, pp. 603–11); thus, its effect does not involve partialing in the form this article addresses.

outcomes onto the six facets within each domain to isolate the contribution of each facet to each outcome.

Finally, partialing is used to residualize scores on a measure of a construct to remove variance shared with one or more extraneous constructs in an effort to purify the construct of interest and to reveal its true nature and effects. This is often done when one or more constructs are thought to differ conceptually but are relatively highly related empirically. In some analyses, this involves entering an extraneous construct at the first step in a regression and using other variables to predict the variance that remains in the construct of interest. In other analyses, the overlapping variance is removed from one or both variables by regressing one variable on the other and saving the residuals as described above. These residualized scores are then correlated with other constructs of interest to highlight differences between them. Such analyses have been used frequently in the study of proactive and reactive aggression (e.g., Miller & Lynam 2006, Raine et al. 2006), shame and guilt (e.g., Bannister et al. 2019, dos Santos et al. 2020, Fee & Tangney 2000), perfectionistic strivings and concerns (e.g., Aldea & Rice 2006, Hill 2014, Hill et al. 2010), and the Dark Triad—the simultaneous examination of psychopathy, narcissism, and Machiavellianism (e.g., Jonason & Tome 2019, Lyons et al. 2017, Sleep et al. 2017). For example, Jonason et al. (2014) examined the relations between several measures of deception and measures of the Dark Triad traits. After presenting the zero-order correlations, the authors reported the partialed relations for each construct (e.g., the association between psychopathy with narcissism and Machiavellianism partialed from it and self-rated deception ability). Although the zero-order relations were generally the same for the Dark Triad traits across indices of deception, the partial relations frequently differed from one another in terms of statistical significance. Interpretive priority was given to the partialed relations. In some cases, only the multivariate, partialed analyses are reported, obscuring the substantive differences that can occur across analytic approaches (e.g., Kowalski et al. 2018).

The uses of partialing that raise conceptual concerns extend beyond the multiple regression framework. In conditional independence network analyses (Costantini et al. 2015, Cramer et al. 2010), which have been used to model complex, causal dynamical systems (e.g., personality, psychopathological symptoms), networks are derived from partial correlations among the network variables while the overlap is controlled for among all variables included in the network. Similarly, partialed variables are the main constructs of interest in the study of personality nuances, which have been defined as "the lowest level at which patterns of responses to questionnaire items continue to have reliable specific variance" (Condon et al. 2020, p. 925). In studies of personality nuances, items are residualized by removing the variance they share with higher-order factors and then correlated with various outcomes. For example, Mõttus et al. (2017) residualized each of the 240 items of the NEO PI-R by removing variance associated with the facet to which the item belonged (calculated without the items) and all other facets. Finally, partialing is also of concern in the use of the bifactor model, which includes a general factor (onto which all items/scales load) and a series of specific factors that are, by necessity, orthogonal to the general factor (e.g., Watts et al. 2019).

These uses of partialing can produce unintended consequences. The first two uses of partialing (i.e., control for confounds, residual change scores/autoregressive paths) are relatively interpretively straightforward with one exception: suppression.[3] Under suppression, the introduction of a variable, x_2, increases the predictive validity of x_1 and/or produces a change in the direction of its effect on y (MacKinnon et al. 2000). In the prenatal smoking example, suppression would occur if the introduction of a potential confounder (e.g., maternal depression) increased rather

[3] Suppression is also problematic for all subsequent uses of partialing.

than decreased the size of the relation between maternal prenatal smoking and offspring behavior problems. In essence, suppression leads to a relation between x_1 and y that was not present at the zero-order level and that cannot be attributed unproblematically to the construct of interest. If the intention is pure prediction (i.e., to account for as much variance as possible in the outcome), then the introduction of a suppressor is beneficial. However, if the intention is to study a causal relation or make inferences about the content of a construct, then this is complicated by the existence of a suppressor.

The other approaches (i.e., unique correlates, residualized variables, network analyses, personality nuances, and bifactor models) are interpretively problematic even in the absence of suppression. Several factors contribute to these problems. First, as noted above, the residual variables contain relatively more error than their raw counterparts, which will lead to instability in the estimates of their associations with other variables (for an example using the Dark Triad, see Vize et al. 2020b). Second, and more problematically, it is difficult to know what these partialed variables actually reflect. With unpartialed variables, the researcher can examine item content to understand what is being assessed; such content is not available for residualized variables. For example, in the abovementioned Walton et al. (2018) study, facets within a domain are all aspects of the same domain and are, therefore, highly intercorrelated; in that study, correlations between facets within a domain ranged from 0.28 to 0.61. Within extraversion, what is positive emotion once the variance it shares with warmth, gregariousness, assertiveness, activity, and excitement seeking is removed? Empirically, this difficulty can be seen in a comparison of the significance tests of zero-order correlations and corresponding regression coefficients. All facets of neuroticism are significantly, positively correlated with substance use disorder, but there are no statistically significant effects for any neuroticism facets in the regression analysis. Within extraversion, warmth and gregariousness are negatively correlated and excitement seeking is positively correlated with substance use disorder. In the regression model, excitement seeking is the only significant predictor, and its semipartial correlation is twice the size of its zero-order correlation. The same interpretive problems plague the Jonason et al. (2014) analyses. What is Machiavellianism when the variance it shares with narcissism and psychopathy is removed? How is one to interpret the vastly different relations to deception seen across the zero-order and semipartial correlations? Similar interpretive concerns apply to network analyses (e.g., Bringmann et al. 2019, Forbes et al. 2019, 2021, Hallquist et al. 2021), personality nuances, and bifactor models (Bonifay et al. 2017, Watts et al. 2019).

EFFECTS OF COLLINEARITY, UNRELIABILITY, AND MAGNITUDE OF ASSOCIATIONS

The rationale for concerns about partialing is long-standing and clear, as evidenced by published results justifying those concerns across a range of topic areas in clinical science. Yet, relatively little is known about the particular analytic circumstances most likely to result in problematic results and interpretation involving partialed variables. To help fill that gap, we present results from two simulation studies that focused on factors likely to affect the degree to which partialing distorts associations and obfuscates the meaning of variables and the constructs they represent. In this section we describe a simulation focused on patterns of association among three variables: a focal predictor, x_1; a variable partialed from it, x_2; and an outcome, y. We varied the magnitude of the association between x_1 and x_2 and their associations with y. For reasons illustrated in **Figure 1**, we expected interpretational problems produced by partialing to be magnified in the presence of unreliability. To evaluate the degree of this expected magnification, we compared partialed effects across levels of the x_1, x_2, and y associations assuming that x_1 and x_2 were measured with either perfect ($\rho_{xx} = 1$) or less than perfect ($\rho_{xx} = 0.75$) reliability. The full design of the simulation study is reflected in the layout of **Figure 2**.

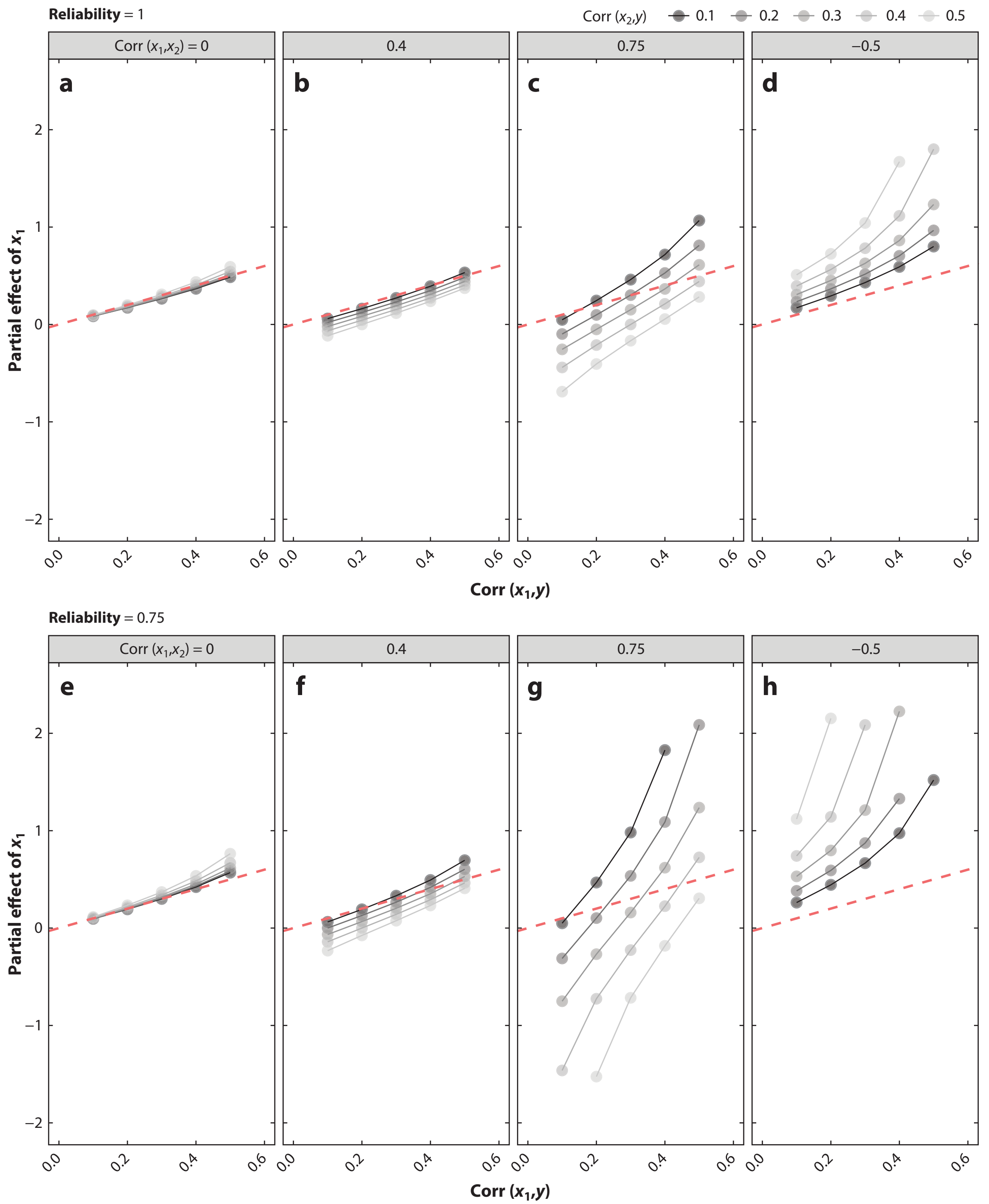
Figure 2*a* shows that, when x_1 and x_2 are uncorrelated and measured without error, the correlations of the raw and partialed predictor with the outcome are virtually identical at all levels of r_{x_2y} considered. Comparing this pattern with the one in **Figure 2*e*** shows a slight tendency toward suppression as the x_2y correlation increases, despite no correlation between x_1 and x_2—an artifact of using covariances in the data-generating model that are converted to correlations for presentation (Cudeck 1989).

The more interesting panels in **Figure 2** are those in which the focal variable and the variable partialed from it are correlated. Looking across the second, third, and fourth panels (**Figure 2*b*–*d*,*f*–*h***) and comparing between rows shows that although the patterns are the same, the steepness and spread of the raw versus partialed effects of x_1 on y are magnified when x_1 and x_2 are measured with typical error. The second and third panels in each row (**Figure 2*b*,*c*,*f*,*g***) represent the relatively common circumstance in which x_1 and x_2 are positively correlated with each other and both are correlated with y. In the absence of measurement error, when the correlation between x_1 and x_2 is moderate (**Figure 2*b***), the effect of partialing is to reduce the association between the focal variable and the outcome, increasingly so as the association between the variable to be partialed and the outcome increases. In the presence of measurement error with the same pattern of correlations, the effect of partialing is both more pronounced and more complex, sometimes reducing and sometimes increasing the association between x_1 and y. When the correlation between x_1 and x_2 is strong ($r = 0.75$), a magnitude not unusual when partialed forms of variables are produced through residualization (e.g., guilt and shame, proactive and reactive aggression), the differences between the association of raw and partialed forms of the focal variable and the outcome are pronounced. When x_1 and x_2 are measured without error (**Figure 2*c***), the x_1y correlation is increasingly reduced as the x_2y correlation increases when the x_1y correlation is low. However, when the x_1y correlation is relatively high ($r = 0.50$), partialing is as likely to produce suppression as attenuation. As is evident across the figure, this pattern is pronounced in the presence of measurement error. When x_1 is partialed, the x_1y association is highly underestimated when the x_2y association is weak but highly overestimated when the x_2y association is strong.

The rightmost panels in each row (**Figure 2*d*,*h***) correspond to a pattern of associations between the variables likely to produce suppression: x_1 and x_2 are negatively associated with each other ($r = -0.50$) but positively associated with y. The expected suppression is evident in both panels, with all points falling above the reference line. The extent of suppression increases as the correlations of x_1 and x_2 with y increase. In the presence of measurement error, relatively modest correlations between x_1 and y and between x_2 and y produced substantial suppression effects suggesting a much stronger association between the partialed focal variable and the outcome than between the raw focal variable and the outcome.[4]

The simulation results show the patterns of association that are most likely to yield discrepant estimates of association between the outcome and the focal variable in raw and partialed forms. The discrepancies are particularly concerning when the correlation between the focal variable and the variable to be partialed from it is strong and positive or negative. In such cases, the researcher is left with the challenge of correctly interpreting what has been removed from the focal variable

[4]Also evident in the panel showing the pattern in the presence of measurement error (**Figure 2*h***) are missing points at higher levels of the x_1y and x_2y associations. To incorporate unreliability and the assumed regression model into the simulation, we imposed constraints on the data-generating model. With those constraints, there was no solution for the graphed values when $r_{x_1x_2}$ was high. The constraints and other features of the simulation are described in the documents available at **https://doi.org/10.17605/OSF.IO/AGU4R**.



(Caption appears on following page)

Figure 2 (*Figure appears on preceding page*)

Results of simulation showing the influence of the correlation between a predictor (x_1) and a partialed variable (x_2), the correlations of x_1 and x_2 with an outcome (y), and unreliability of x_1 and x_2 on estimates of the partialed effect of x_1 on y. Graphs in the top row (panels *a–d*) assume $\rho_{xx} = 1$ in the population; those in the bottom row (panels *e–h*) assume $\rho_{xx} = 0.75$. Within each row, panels differ in the correlation between x_1 and x_2 in the population. Lines plotted within panels correspond to five levels of association between x_2 and y in the population. The x axes reflect increasing correlation between raw x_1 and y. The y axis is the association between partialed x_1 and y. In each panel, the dashed red reference line corresponds to equal values of the raw and partialed r's.

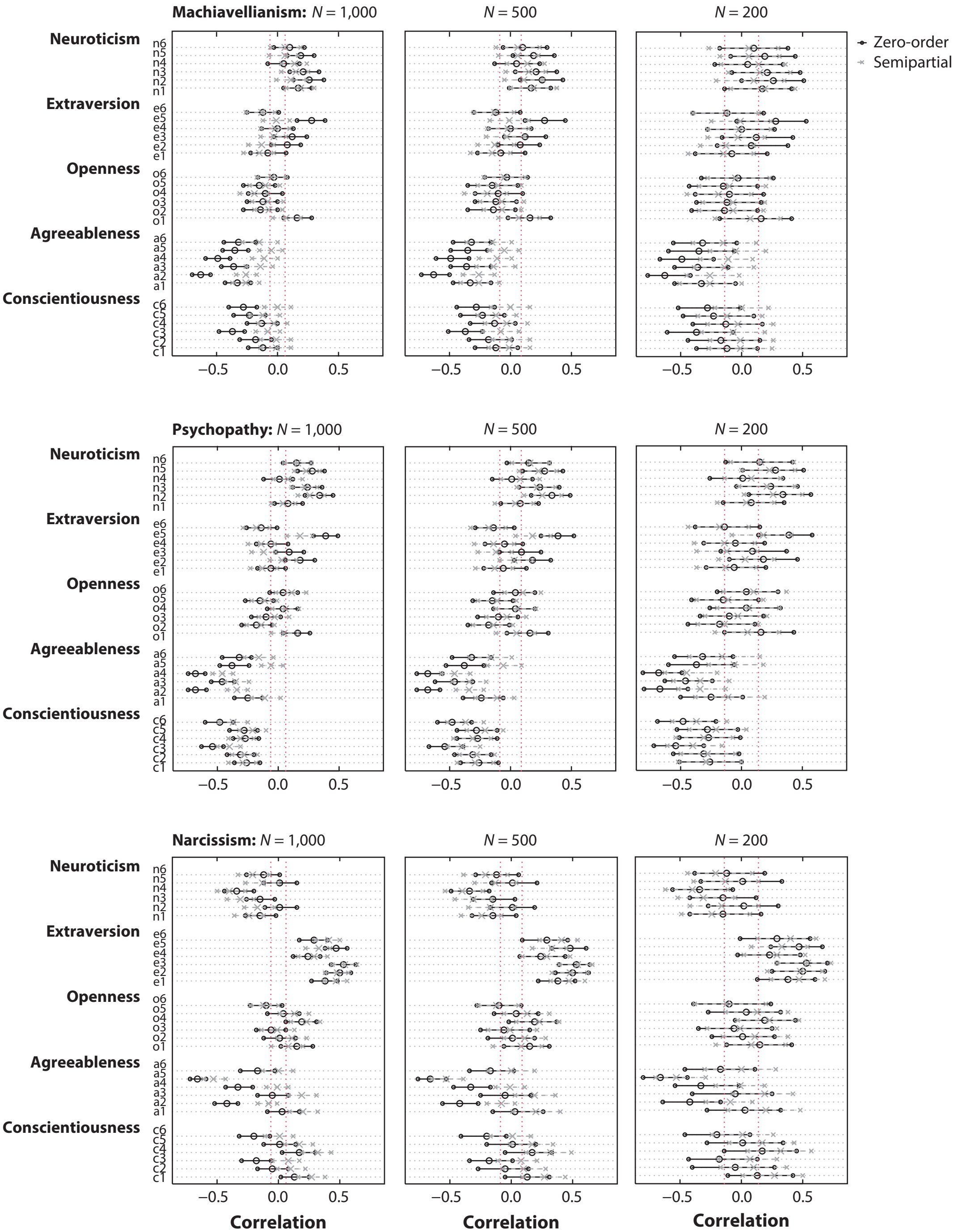
that altered its association with the outcome. The interpretation is particularly challenging when the association between the focal variable and the variable to be partialed from it is negative, as this pattern may reveal an apparent association between the partialed form of the focal variable and the outcome that was not evident for the raw form of the focal variable. These challenges are further complicated by the magnifying effect of measurement error on the discrepancies between the raw and partialed associations.

RESIDUALIZATION AND CONSTRUCT OBFUSCATION

Although the first simulation is informative, it lacks the important element of knowledge about the constructs on which the correlations are based. Our second simulation adds this element along with a consideration of how sample size may affect inferences about estimates of associations involving partialed variables. Specifically, we considered the case of the Dark Triad: the correlated constructs of Machiavellianism, narcissism, and psychopathy. Rather than simulated population data, we treated as population data responses from a combined sample of respondents ($N = 2{,}432$) from four studies that included measures of the Dark Triad constructs and the five-factor model (Collison et al. 2018, Miller et al. 2017, Rose et al. 2023, Vize et al. 2020a). Details about the sample and measures are available at **https://doi.org/10.17605/OSF.IO/AGU4R**.

Unlike in the first simulation, we did not manipulate the pattern of correlations between the variables of interest. Rather, we estimated correlations from the data based on samples of different sizes from the population data set, across which the correlations in the population between Dark Triad variables were $\rho_{\mathrm{Mn}} = 0.41$, $\rho_{\mathrm{Mp}} = 0.60$, and $\rho_{\mathrm{np}} = 0.44$. Specifically, we drew 10,000 random samples (with replacement) each of size 1,000, 500, and 200. For each of the resultant samples we obtained two correlations between each of the Dark Triad constructs and the 30 facets of the five-factor domains. Zero-order correlations are the associations of raw scores on Machiavellianism, narcissism, and psychopathy with personality facet scores. Semipartial correlations are the associations of each of the Dark Triad variables after the other two have been partialed from it with the facet scores. Partialed forms of the Dark Triad variables are residuals from regressing each variable on the other two. In addition to the median of the 540 correlations, we examined observed ranges across the 10,000 data sets of each size. These results are presented in visual format in **Figure 3**. Of specific interest in the figure are (*a*) the gap between the median zero-order and semipartial r's and (*b*) the degree of overlap in the ranges of estimates around the medians.

Before commenting on specific patterns for each Dark Triad variable and set of personality facets, we note a general finding related to sample size. As is to be expected, sampling variability is lower with larger samples. This can be seen in **Figure 3** in both the increasing width of the ranges moving left to right in each row of panels and the increasing width of the region of rejection for statistical tests. The width of the range and region of rejection change in tandem as a function of sample size, though the median estimates are more likely to fall outside the region of rejection for the larger N's. This characteristic of the general pattern can be seen by focusing specifically on r's between the personality facet n6 (i.e., sixth facet of n̲euroticism) and Machiavellianism. When $N = 1{,}000$ or 500, the median estimate of the zero-order r is positive and significant; the median



(*Caption appears on following page*)

Figure 3 (*Figure appears on preceding page*)

Results of simulation showing the differences between zero-order and semipartial correlations for each Dark Triad variable across personality facets at three sample sizes. The horizontal axis in each panel is magnitude and direction of correlation coefficient. The vertical dashed red reference lines designate the region of statistical nonsignificance in a test against zero ($p < 0.05$). For each facet, the median zero-order r and the semipartial r are shown. The ranges are indicated by a black solid line for the zero-order r and a gray dashed line for the semipartial r.

estimate falls within the region of rejection for $N = 200$ and, therefore, is not significantly different from zero. The median estimate of the semipartial r is not significantly different from zero at any sample size. The ranges for the two r's are highly overlapping, which contrasts with the ranges for n5 and Machiavellianism; when $N = 1{,}000$, the overlap of the two ranges is only about 50%.

Turning to specific patterns that illustrate the interpretive challenges associated with correlations involving partialed variables, first note that there is no clear pattern to the differences between the zero-order and partialed correlations. To more formally examine the pattern, we correlated the profiles of zero-order and semipartial correlations across the 30 personality facets for each of the Dark Triad variables using a double-entry correlation. The results are presented in **Table 1**. Because of the large number of samples, the sampling distributions are stable across sample sizes; the mean intraclass correlation coefficients (ICCs) are the same within rounding. Although the patterns are similar, on average, for narcissism and psychopathy, such is not the case for Machiavellianism. Moreover, when considering the range of observed ICCs across the 10,000 samples and focusing specifically on the lower limit of the ranges, profiles could be only modestly similar in any particular study at a sample size typical of clinical research.

Returning to **Figure 3** and focusing on the panel in which $N = 1{,}000$ for Machiavellianism, notice that for some facets, the ranges of observed correlations do not overlap (e.g., e5, a2, c6). Moreover, in a subset of these cases, the statistical test against zero is always rejected for the zero-order r but highly unlikely to be rejected for the semipartial r (e.g., a5, c6). Consistent with the ICC results, the differences between the two coefficients across facets are smaller for narcissism and psychopathy, but strong differences are nonetheless evident. Within the $N = 1{,}000$ panels for narcissism and psychopathy, the full ranges of the observed correlations for a4 and a5 do not overlap. The impact of partialing on the magnitude of these correlations is consistent and strong. The likelihood of observing suppression can be seen in the pattern of differences for the agreeableness and conscientiousness facets. For example, for narcissism in the $N = 1{,}000$ condition, the median zero-order r for c3 is significant and negative, but the median semipartial r is significant and positive. For c3, both coefficients are positive and significant, but the median for the semipartial r is above

Table 1 Mean and range of intraclass correlation coefficients between median zero-order and semipartial correlations for each Dark Triad variable with 30 personality facets for 10,000 samples of three sizes

	Machiavellianism	Neuroticism	Psychopathy
$N = 200$			
Mean	0.52	0.83	0.81
Range	0.09–0.82	0.52–0.98	0.47–0.95
$N = 500$			
Mean	0.51	0.83	0.82
Range	0.20–0.73	0.63–0.94	0.65–0.91
$N = 1{,}000$			
Mean	0.50	0.84	0.82
Range	0.31–0.67	0.71–0.92	0.73–0.90

the upper end of the range for the zero-order r. For Machiavellianism, the sign of the coefficients flips from positive for the zero-order r to negative for the semipartial r (median values significant in both cases). A more common pattern is for the zero-order r to be significant but the corresponding semipartial r to be nonsignificant (e.g., for narcissism—a2, a4, a6, c6). In short, nearly every form of difference between the zero-order and semipartial r's is evident across the personality facets.

These problematic patterns are more likely and less predictable when sample size is closer to what is typical in clinical research. Referring to the rightmost panels in each set, given the wide and overlapping ranges of the zero-order and semipartial r's when $N = 200$, the difference could, for a specific study, be small or large, reflecting a semipartial r that is weaker or stronger than the zero-order r, with signs that are the same or different. When sample size is modest, interpretational problems that arise from partialing are exacerbated by the unlikelihood of replication across similarly powered studies. Replication of significant semipartial effects across independent samples is necessary to ensure that they are not Type I errors.

RECOMMENDATIONS

We have provided evidence that partialing as typically practiced in clinical psychology research (*a*) often does not serve the purpose for which it was intended and (*b*) often does not produce results that can be interpreted with reference to the constructs of interest. Yet, partialing typically is motivated by understandable concerns about the clarity of constructs as measured and inferences about their associations with other constructs. The challenge, then, is how to better manage the conceptual and inferential challenges that lead researchers to use the partialing methods we have shown to be problematic. Although no single technique or strategy can fully address those challenges, it is important that potential solutions do not, as is the case with partialing, create additional conceptual and interpretational problems. To that end, we propose greater precision and rigor through the use of alternative conceptual, analytical, and measurement strategies.

Theory-Informed Statistical Modeling

In practice, covariates and nuisance variables are often included in statistical models with neither adequate justification nor a conceptual understanding of how the focal variable and variables to be partialed from it relate to each other and to outcome variables they might be used to predict. This concern was referred to more than half a century ago by Meehl (1971) and has been echoed by others in the years since (e.g., Hünermund & Louw 2005, Jaccard et al. 2006, Kim et al. 2021, Lawson & Robins 2021, Lynam et al. 2006, Miller & Chapman 2001, Schneider et al. 2015, Spector & Brannick 2011, Wright 2021). Effectively addressing this concern requires moving beyond atheoretical partialing (Jaccard et al. 2006) to specifying thoughtfully and explicitly the nature of the association between a focal variable and variables to be partialed from it (Lawson & Robins 2021, Spector & Brannick 2011). Specified as such, it may become evident that multiple regression analysis or one of the multivariate models that rely on partialing discussed earlier would not be appropriate for addressing overlap between variables as they relate to a process or mechanism of interest.

Most fundamentally and as illustrated in the first simulation presented above, the association between a focal variable and control variables is not the only association of relevance in the context of prediction models. The extent to which each is associated with an outcome variable influences the effect of partialing on statistical results. Thus, at the most basic level, moving beyond atheoretical partialing requires a theoretical account of how the focal and control variables are related as well as an account of how the control variables are related to the outcome. That is to say, the researcher needs a theory, or at least a defensible argument, about what part of the focal variable is associated with what part of the outcome variable when partialing is used. Although the account could be incorrect, it at least attempts an explanation for anticipated differences between zero-order and semipartialed associations that precedes knowledge of the results, thereby avoiding a questionable measurement practice (Flake & Fried 2020). It also guards against the mindless inclusion of covariates, often for no reason other than that it is what others have done (Stark & Saltelli 2018) and with no clear explanation of how they influence the statistical association of interest.

In the absence of relevant theory, it is incumbent on researchers who routinely use partialed variables to examine differences between the associations of raw and partialed forms of variables with well-understood variables that may or may not have an obvious theoretical connection to the variable of interest. Our use of the personality facets from the five-factor model to compare raw and partialed forms of the Dark Triad variables in the second simulation study is an example (see also Sleep et al. 2017). A similarly informative comparison can be done at the item level. Vize et al.'s (2020b) use of item-level personality scores to examine differences between raw and residualized Dark Triad variables is an example. Strategic comparisons can provide a basis for theorizing about associations involving the construct of interest, which can provide a basis for improved modeling of semipartial effects in prediction models.

Flexible Analytic Approaches Plus Simple Associations

Multiple regression analysis and, for quasi-experimental studies, ANCOVA are the most frequently used analytic approaches for estimating the effect of a focal variable on an outcome while accounting for competing or extraneous influences. Yet these analytic approaches are limited in the types of associations between focal, control, and outcome variables they can model. Moreover, they do not, by default, provide zero-order correlations, which are essential to understanding how partialing affects associations between the variables. Of the seven forms of association between a focal variable, a control variable, and an outcome variable described by Jaccard et al. (2006), only two are correctly modeled using multiple regression analysis.[5] Even for those straightforward models, information beyond what is typically provided in statistical output from these analyses (i.e., zero-order correlations) is needed to interpret the impact of accounting for effects of the control variable on the estimates of the associations between the focal and outcome variables.

When theory or logic points to a causal association between a focal variable and a control variable, or an indirect effect of either the focal or control variable on the outcome variable, an analytic approach more flexible than traditional approaches is required. Structural equation modeling (SEM), of which multiple regression and ANCOVA are special cases, offers the necessary flexibility to model direct associations between focal and control variables and indirect effects of either on outcome variables. SEM has the added benefit of providing a means of addressing unreliability of measurement, thereby eliminating the complicating effects of that source of variance on partialed effects (Wang & Eastwick 2020, Westfall & Yarkoni 2016). SEM offers flexibility in the specification of multiple relations between variables in a model, and, as such, it is a natural fit for analyses based on theory-informed models as described above.

[5] Denoting the focal variable as X, the control variable as C, and the outcome variable as Y, the seven causal models described by Jaccard et al. (2006) are (*a*) X and C are corrected causes of Y; (*b*) C causes X, which causes Y; (*c*) X causes C, which causes Y; (*d*) X causes Y directly and indirectly through C; (*e*) X and C are correlated, but only X causes Y; (*f*) C causes X and Y; and (*g*) C causes Y directly and indirectly through X. Multiple regression analysis is appropriate for Models *a* and *e*, though the test of the focal regression weight in Model *e* is underpowered because the associated standard error is inflated.

Whether multiple regression analysis, ANCOVA, or a more flexible analytic approach such as SEM is used, zero-order correlations should always be reported alongside any effects involving partialed forms of variables (Funder & Ozer 2019, Vize et al. 2018). Because scores underlying zero-order correlations are based on item responses, the nature of the covarying constructs is transparent if not always clear. Zero-order correlations can be compared within and between studies, allowing statements about the relative magnitude and consistency of associations. They are based on full versions of the constructs as measured, including overlapping but also essential features of the constructs of interest. Echoing a point made above and illustrated in the first simulation, understanding what partialing is doing requires going beyond comparing zero-order and semipartial associations between a focal variable and an outcome variable. Information about the zero-order associations between focal and control variables as well as between control and outcome variables is necessary to fully appreciate how partialing changes an effect of interest and to offer a defensible interpretation of the partialed effect. Although reporting zero-order correlations between raw scores is somewhat common with multiple regression analysis, it should be required. Moreover, to the extent possible, partialed effects should be interpreted, both statistically and conceptually, with explicit reference to those correlations, and differences across the two should be acknowledged. Importantly, however, researchers must guard against undisclosed researcher degrees of freedom in terms of statistical inference (Simmons et al. 2011), stating a priori (ideally in a preregistration) whether estimates and tests will be interpreted for the zero-order or the semipartialed effects, which often will disagree.

Reducing the Need to Partial Through Better Measurement

Our recommendations to this point have assumed the use of measures that capture more than the construct of interest, complicating the interpretation of observed associations between raw scores on the measures and other variables. Partialing is necessitated by such imprecise measurement, which stems, in part, from underdeveloped or problematic conceptualizations of constructs. As such, the most efficient path to a reduction in partialing in clinical science is better conceptualization and measurement. We acknowledge that the nature of the constructs clinical researchers measure may make it difficult to design measures that are not contaminated by unintended influences (e.g., symptom clusters). Nevertheless, we assume that measurement could be better and that the issues problematically addressed by partialing may help guide development of those better measurement practices.

Essential to the development of sharper conceptualizations and better measures is the process of construct validation (e.g., Clark & Watson 1995, Cronbach & Meehl 1955, Loevinger 1957, Strauss & Smith 2009). The process begins with an explicit and detailed articulation of a construct, which includes stating its expected associations, in terms of sign and strength, with other constructs. As part of an ongoing, dynamic process, one then evaluates the degree to which a potential new measure of the target construct relates to other constructs as expected. Results of the evaluation may lead to refinements of the candidate measure (dropping, adding, or rewording candidate items) or adjustments to the conceptualization of the construct. Thus, conceptualization of the construct changes to some degree, and the candidate measure is updated accordingly. In this way, scores produced by a measure are not broadly assumed to be construct valid because construct validation is an ongoing process. Similarly, measurement development should always be an iterative process in which obtained results are used to strengthen and hone a measure (Clark & Watson 1995).

A critical concern in measure development, one that anticipates the problems we have highlighted, is minimizing overlap between the construct and its candidate measure with "sibling constructs" (Lawson & Robins 2021). Effectively addressing construct overlap during measure development is the most efficient means of minimizing the need for partialing or, when overlap is unavoidable, facilitating clearer interpretation of partialed associations. In this regard, it is important to specifically consider the constructs most covarying/comorbid with the target construct, as those will be the most likely ones used in the partialing practices described above (Smith & McCarthy 1995). In general, overlap between measures of different constructs can be reduced by identifying the conceptual core of a construct and writing items targeted at it while minimizing content that captures antecedents, outcomes, or correlates. Yet, despite efforts to sharply define a construct and write items intended to reflect only core features of it, researchers may nonetheless find substantial empirical overlap between target and sibling constructs. For example, efforts to develop a new measure of depression with minimal overlap with existing measures of anxiety may prove difficult because some overlap beyond their shared features may lie with existing measures of anxiety. In this way, researchers may need to create measures of both their target construct and sibling constructs if they are to have better isolation of constructs in their own research.

Putting forward a detailed conceptualization of a construct, as well as its nomological network, will aid in avoiding the pitfall of including items that reflect sibling constructs instead of solely items focused on the core of the construct. Yet, articulation of a detailed conceptualization and items that map onto it does not guarantee that candidate items will not inadvertently reflect sibling constructs. The detection of such items for elimination or refinement requires careful empirical observation. Informative observation is enabled by the inclusion of measures of sibling constructs at the stage of item development and selection. By correlating candidate items and scores on candidate item sets with sibling constructs, the researcher can detect items that best isolate the focal construct from sibling constructs with the goal of an item set that, when scored as a composite, minimizes overlap with sibling constructs. This back and forth between item selection and evaluation of discriminant validity is best done at the point of measure development, though it may need further consideration as a measure is used in research or as other sibling constructs are proposed.[6] The challenge in this work is not forsaking fidelity to the conceptualization of the focal construct in the interest of discriminant validity. Given psychology's "crud factor" (Meehl 1990; see also Orben & Lakens 2020), however, researchers must recognize that it will be next to impossible to create perfectly clean and/or pure measures with no discriminant validity problems.

The process of measure development we have outlined is complicated by the fact that many clinical constructs are multidimensional in nature. The field would benefit from the creation and use of measures that acknowledge this reality and allow for a building block approach that allows for cleaner and easier decomposition of constructs. If a researcher intends to assess a multidimensional construct, homogeneous subscales assessing the dimensions should be created first; then they may be combined to form the broader construct (Smith et al. 2009). Alternatively, existing homogeneous measures of basic traits could be evaluated as potential building blocks for complex clinical diagnoses or syndromes (for a review of successful efforts to build measures of personality disorders using personality facets, see Bagby & Widiger 2018). The result would be a better

[6]As this discussion makes clear, understanding what a set of items is measuring, or was intended to measure, involves knowledge of item content. Items are evaluated and refined during the process of construct validation (Loevinger 1957). Knowledge about the construct reflected in item content is lost when partialed forms of scale scores are used in analyses. Moreover, evidence of structural validity (e.g., factor analysis) cannot be determined. Thus, neither content nor structural validity can be evaluated for partialed variables. For partialed variables, predictive validity is the only form of construct validation possible. If the partialed scores fail an evaluation of predictive validity, the reason for the failure cannot be determined because the scores do not reflect item content.

understanding of the features that overlap in clinical constructs and disorders (e.g., negative affectivity for depression and anxiety) as well as those that differentiate (e.g., anxiety–physiological arousal; depression–low positive affect; Clark & Watson 1991). A deeper understanding of multidimensional constructs reflected in better measures of them would facilitate inferences and interpretation when partialing cannot be avoided.

CONCLUSION

Progress in clinical science requires rigorous measurement of well-defined constructs that support unambiguous inferences from estimates and tests of associations between constructs. When this requirement is met, statistical approaches aimed at refining variables to more closely match the construct they were intended to reflect are not needed. Moreover, observed associations are more likely to be replicable and interpretable than comparable associations involving variables refined through statistical adjustments to scores. When a construct is either not well defined or well defined but poorly measured, clinical researchers are required to engage in the questionable measurement practice of partialing as a means of refining scores and isolating the construct of interest. Our review has highlighted the many ways partialing distorts rather than clarifies the nature and correlates of constructs. These range from producing to eliminating to reducing to increasing to changing the direction of association between a variable based on a problematic measure and correlates or outcomes. The distortion is magnified in the inevitable presence of unreliability. The result is estimates and tests of association that are likely to vary widely across studies, especially when sample sizes are modest, introducing rather than eliminating uncertainty about the nature of the construct and its relations with other constructs. For this reason, we recommend deeper thinking about the form of association between constructs and the use of more flexible statistical approaches to modeling them when partialing cannot be avoided. At a more fundamental level, new measures should be developed with reference to well-defined constructs and with due attention to the need to differentiate between similar constructs. Importantly, as theoretical models and empirical findings in a literature accumulate, definitions of constructs should be sharpened and measures better focused to reflect improved understanding. The result will be a more rigorous, robust, and influential clinical science.

SUMMARY POINTS

1. Common uses of partialing are a questionable measurement practice because they attempt to generate meaningful and replicable findings from measures known to be contaminated by constructs they were not intended to measure.
2. Partialing is evident and often intentional in multiple regression analysis and certain uses of analysis of covariance to control for nuisance variables; however, it is also present and often unrecognized in an array of multivariate models increasingly common in clinical science (e.g., network analyses, bifactor models). Interpretation of effects detected using these models rarely acknowledges the potential distorting effects of partialing.
3. Partialing increases the relative proportion of error and nuisance construct variance and has a negligible effect when there is no overlap between the focal and contaminating constructs.
4. Our simulations show that partialing changes variables, often in ways that are dramatic and not easily understood. It is difficult to know what is left in a variable once variance

shared with other variables is removed. Partialed variables are interpretively opaque relative to their raw counterparts.

5. Measurement error increases the uncertainty created by partialing. At levels of unreliability typical of measures in clinical science, the effect of partialing is pronounced and complex, sometimes reducing and sometimes increasing the association between a focal construct and outcome.
6. Atheoretical partialing is particularly problematic because it assumes a pattern of associations between the focal, control, and outcome variables in predictive models that may not accurately reflect the mechanisms that underlie their associations. It also contributes to the common failure to distinguish between explanations for zero-order and semipartial effects.
7. Commonly used statistical approaches such as multiple regression analysis and analysis of covariance are limited in their ability to accurately reflect theoretically informed mechanisms that may underlie the associations between focal, control, and outcome variables. For this reason, more flexible analytic approaches, preferably those that allow for modeling unreliability, should be used to model partialed effects.
8. Better conceptualization and measurement of constructs would reduce the need for partialing, thereby improving the replicability of effects and their interpretation.

FUTURE ISSUES

1. Many constructs in clinical science are inherently complex and likely will defy the most rigorous attempts at clean measurement. Novel statistical approaches and strategies for isolating constructs are needed that do not create the problems we have identified with partialing as currently practiced.
2. The measurement of complex and multidimensional constructs could be improved by focusing on careful measurement of unidimensional constituent features, which could be weighted and combined to create measures of these constructs that are similar or different from other constructs in ways that are easily identified.
3. Scale development efforts that emphasize discriminant validity at the candidate item and scale phase of development anticipate cross-construct contamination and are recommended. However, such efforts should take care to avoid losing focus on the conceptual core of the construct of interest in an effort to minimize its overlap with other constructs.
4. Better measures need to be constructed that allow broad constructs to be decomposed into more homogeneous components. This should allow for clearer understanding of what is removed and what remains following partialing.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

This research was supported by National Institute on Drug Abuse grant P30DA023026 to R.H.H.

LITERATURE CITED

Aldea MA, Rice KG. 2006. The role of emotional dysregulation in perfectionism and psychological distress. *J. Counsel. Psychol.* 53:498–510

Averill PM, Diefenbach GJ, Stanley MA, Breckenridge JK, Lusby B. 2002. Assessment of shame and guilt in a psychiatric sample: a comparison of two measures. *Personal. Individ. Diff.* 32:1365–76

Bagby RM, Widiger TA. 2018. Five Factor Model personality disorder scales: an introduction to a special section on assessment of maladaptive variants of the Five Factor Model. *Psychol. Assess.* 30:1–9

Bannister JA, Colvonen PJ, Angkaw AC, Norman SB. 2019. Differential relationships of guilt and shame on posttraumatic stress disorder among veterans. *Psychol. Trauma* 11:35–42

Batty GD, McIntosh AM, Russ TC, Deary IJ, Gale CR. 2016. Psychological distress, neuroticism, and cause-specific mortality: early prospective evidence from UK Biobank. *J. Epidemiol. Commun. Health* 70:1136–39

Bonifay W, Lane SP, Reise SP. 2017. Three concerns with applying a bifactor model as a structure of psychopathology. *Clin. Psychol. Sci.* 5:184–86

Bringmann LF, Elmer T, Epskamp S, Krause RW, Schoch D, et al. 2019. What do centrality measures measure in psychological networks? *J. Abnorm. Psychol.* 128:892–903

Cândea DM, Szentagotai-Tătar A. 2018. Shame-proneness, guilt-proneness and anxiety symptoms: a meta-analysis. *J. Anxiety Disord.* 58:78–106

Clark LA, Watson D. 1991. Tripartite model of anxiety and depression: psychometric evidence and taxonomic implications. *J. Abnorm. Psychol.* 100:316–36

Clark LA, Watson D. 1995. Constructing validity: basic issues in objective scale development. *Psychol. Assess.* 7:309–19

Cohen J, Cohen P. 1983. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences.* Hillsdale, NJ: Erlbaum. 2nd ed.

Collison KL, Vize CE, Miller JD, Lynam DR. 2018. Development and preliminary validation of a Five Factor Model measure of Machiavellianism. *Psychol. Assess.* 30:1401–7

Condon DM, Wood D, Mõttus R, Booth T, Costantini G, et al. 2020. Bottom up construction of a personality taxonomy. *Eur. J. Psychol. Assess.* 36:923–34

Costantini G, Epskamp S, Borsboom D, Perugini M, Mõttus R, et al. 2015. State of the aRt personality research: a tutorial on network analysis of personality data in R. *J. Res. Personal.* 54:13–29

Cramer AOJ, Waldorp LJ, van der Maas HIJ, Borsboom D. 2010. Comorbidity: a network perspective. *Behav. Brain Sci.* 33:137–93

Cronbach LJ, Meehl PE. 1955. Construct validity in psychological tests. *Psychol. Bull.* 52:281–302

Cudeck R. 1989. Analysis of correlation matrices using covariance structure models. *Psychol. Bull.* 105:317–27

Darlington RB, Hayes AF. 2017. *Regression Analysis and Linear Models*. New York: Guilford

dos Santos MA, de Freitas e Castro JM, de Freitas Lino Pinto Cardoso CS. 2020. The moral emotions of guilt and shame in children: relationship with parenting and temperament. *J. Child Fam. Stud.* 29:2759–69

Fee RL, Tangney JP. 2000. Procrastination: a means of avoiding shame or guilt? *J. Soc. Behav. Personal.* 15:167–84

Fite PJ, Stoppelbein L, Greening L. 2009. Proactive and reactive aggression in a child psychiatric inpatient population. *J. Clin. Child Adolesc. Psychol.* 38:199–205

Flake JK, Fried EI. 2020. Measurement schmeasurement: questionable measurement practices and how to avoid them. *Adv. Methods Pract. Psychol. Sci.* 3:456–65

Forbes MK, Wright AGC, Markon KE, Krueger RF. 2019. The network approach to psychopathology: promise versus reality. *World Psychiatry* 18:272–73

Forbes MK, Wright AGC, Markon KE, Krueger RF. 2021. On unreplicable inferences in psychopathology symptom networks and the importance of unreliable parameter estimates. *Multivar. Behav. Res.* 56:368–76

Funder DC, Ozer DJ. 2019. Evaluating effect size in psychological research: sense and nonsense. *Adv. Methods Pract. Psychol. Sci.* 2:156–68

Geisner IM, Koopmann J, Bamberger P, Wang M, Larimer ME, et al. 2018. When the party continues: impulsivity and the effect of employment on young adults' post-college alcohol use. *Addict. Behav.* 77:114–20

Hallquist MN, Wright AGC, Molenaar PC. 2021. Problems with centrality measures in psychopathology symptom networks: why network psychometrics cannot escape psychometric theory. *Multivar. Behav. Res.* 56:199–223

Hill AP. 2014. Perfectionistic strivings and the perils of partialling. *Int. J. Sport Exerc. Psychol.* 12:302–15

Hill RW, Huelsman TJ, Araujo G. 2010. Perfectionistic concerns suppress associations between perfectionistic strivings and positive life outcomes. *Personal. Individ. Differ.* 48:584–89

Hünermund P, Louw B. 2005. On the nuisance of control variables in regression analysis. arXiv:2005.10314v3 [econ.EM]

Jaccard J, Guilamo-Ramos V, Johansson M, Bouris A. 2006. Multiple regression analyses in clinical child and adolescent psychology. *J. Clin. Child Adolesc. Psychol.* 35:456–79

Jonason PK, Lyons M, Baughman HM, Vernon PA. 2014. What a tangled web we weave: the Dark Triad and deception. *Personal. Individ. Differ.* 70:117–19

Jonason PK, Tome J. 2019. How happiness expectations relate to the Dark Triad traits. *J. Soc. Psychol.* 159:371–82

Kim E, Anderson CA, Gentile DA. 2021. 7 ± 2 deadly sins of video game violence research. In *Masters of Media: Controversies and Solutions*, Vol. 1, ed. VC Strasburger, pp. 19–41. Lanham, MD: Rowman & Littlefield

Kowalski CM, Kwiatkowska K, Kwiatkowska MM, Ponikiewska K, Rogoza R, Schermer JA. 2018. The Dark Triad traits and intelligence: Machiavellians are bright, and narcissists and psychopaths are ordinary. *Personal. Individ. Differ.* 135:1–6

Lawson KM, Robins RW. 2021. Sibling constructs: what are they, why do they matter, and how should you handle them? *Personal. Soc. Psychol. Rev.* 25:344–66

Loevinger J. 1957. Objective tests as instruments of psychological theory. *Psychol. Rep.* 3:635–94

Lynam DR, Hoyle RH, Newman JP. 2006. The perils of partialling: cautionary tales from aggression and psychopathy. *Assessment* 13:328–41

Lyons M, Croft A, Fairhurst S, Varley K, Wilson C. 2017. Seeing through crocodile tears? Sex-specific associations between the Dark Triad traits and lie detection accuracy. *Personal. Individ. Differ.* 113:1–4

MacKinnon DP, Krull JL, Lockwood CM. 2000. Equivalence of the mediation, confounding and suppression effect. *Prev. Sci.* 1:173–81

Maughan B, Taylor C, Taylor A, Butler N, Bynner J. 2001. Pregnancy smoking and childhood conduct problems: a causal association? *J. Child Psychol. Psychiatry* 42:1021–28

Meehl PE. 1971. High school yearbooks: a reply to Schwartz. *J. Abnorm. Psychol.* 77:143–48

Meehl PE. 1975. Genes and the unchangeable core. *VOICES: Art Sci. Psychother.* 10:25–35

Meehl PE. 1990. Why summaries of research on psychological theories are often uninterpretable. *Psychol. Rep.* 66:195–244

Miller GA, Chapman JP. 2001. Misunderstanding analysis of covariance. *J. Abnorm. Psychol.* 110:40–48

Miller JD, Hyatt CS, Maples-Keller JL, Carter NT, Lynam DR. 2017. Psychopathy and Machiavellianism: a distinction without a difference? *J. Personal.* 85:439–53

Miller JD, Lynam DR. 2006. Reactive and proactive aggression: similarities and differences. *Personal. Individ. Differ.* 41:1469–80

Mõttus R, Kandler C, Bleidorn W, Riemann R, McCrae RR. 2017. Personality traits below facets: the consensual validity, longitudinal stability, heritability, and utility of personality nuances. *J. Personal. Soc. Psychol.* 112:474–90

Orben A, Lakens D. 2020. Crud (re)defined. *Adv. Methods Pract. Psychol. Sci.* 3:238–47

Raine A, Dodge K, Loeber R, Gatzke-Kopp L, Lynam D, et al. 2006. The Reactive-Proactive Aggression Questionnaire: differential correlates of reactive and proactive aggression in adolescent boys. *Aggress. Behav.* 32:159–71

Rose L, Miller JD, Lynam DR. 2023. Validation of the Five-Factor Model Antagonistic Triad Measure. *Assessment* 30:782–97

Ruisch IH, Dietrich A, Glennon JC, Buitelaar JK, Hoekstra PJ. 2018. Maternal substance use during pregnancy and offspring conduct problems: a meta-analysis. *Neurosci. Biobehav. Rev.* 84:325–36

Santor DA, Gregus M, Welch A. 2006. Eight decades of measurement in depression. *Measurement* 4:135–55

Schneider BA, Avivi-Reich M, Mindaugas Mozuraitis M. 2015. A cautionary note on the use of the analysis of covariance (ANCOVA) in classification designs with and without within-subject factors. *Front. Psychol.* 6:474

Simmons JP, Nelson LD, Simonsohn U. 2011. False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol. Sci.* 22:1359–66

Sleep CE, Lynam DR, Hyatt CS, Miller JD. 2017. Perils of partialing redux: the case of the Dark Triad. *J. Abnorm. Psychol.* 126:939–50

Smith GT, McCarthy DM. 1995. Methodological considerations in the refinement of clinical assessment instruments. *Psychol. Assess.* 7:300–8

Smith GT, McCarthy DM, Zapolski TCB. 2009. On the value of homogeneous constructs for construct validation, theory testing, and the description of psychopathology. *Psychol. Assess.* 21:272–84

Spector PE, Brannick MT. 2011. Methodological urban legends: the misuse of statistical control variables. *Organ. Res. Methods* 14:287–305

Stark PB, Saltelli A. 2018. Cargo-cult statistics and scientific crisis. *Significance* 15:40–43

Strauss ME, Smith GT. 2009. Construct validity: advances in theory and methodology. *Annu. Rev. Clin. Psychol.* 5:1–25

Vize CE, Collison KL, Miller JD, Lynam DR. 2018. Examining the effects of controlling for shared variance among the Dark Triad using meta-analytic structural equation modeling. *Eur. J. Personal.* 32:46–61

Vize CE, Collison KL, Miller JD, Lynam DR. 2020a. The "core" of the Dark Triad: a test of competing hypotheses. *Personal. Disord.* 11:91–99

Vize CE, Collison KL, Miller JD, Lynam DR. 2020b. Using item-level analyses to better understand the consequences of partialing procedures: an example using the Dark Triad. *J. Personal.* 88:719–34

Walton KE, Pantoja G, McDermut W. 2018. Associations between lower order facets of personality and dimensions of mental disorder. *J. Psychopathol. Behav. Assess.* 40:465–75

Wang YA, Eastwick PW. 2020. Solutions to the problems of incremental validity testing in relationship science. *Pers. Relatsh.* 27:156–75

Watts AL, Poore HE, Waldman ID. 2019. Riskier tests of the validity of the bifactor model of psychopathology. *Clin. Psychol. Sci.* 7:1285–303

Weidman AC, Steckler CM, Tracy JL. 2017. The jingle and jangle of emotion assessment: imprecise measurement, casual scale usage, and conceptual fuzziness in emotion research. *Emotion* 17:267–95

Westfall J, Yarkoni T. 2016. Statistically controlling for confounding constructs is harder than you think. *PLOS ONE* 11:e0152719

Wright PJ. 2021. Pornographic socialization as "selective-exposure": Let it go, let it go II. *Arch. Sex. Behav.* 50:393–99

Zinbarg RE, Suzuki S, Uliaszek AA, Lewis AR. 2010. Biased parameter estimates and inflated Type I error rates in analysis of covariance (and analysis of partial variance) arising from unreliability: alternatives and remedial strategies. *J. Abnorm. Psychol.* 119:307–19