Contextualizing Statistical Suppression Within Pretest-Posttest Designs

Linda Farmus [a] ✉ ⓘ , Jonathan A. Weiss [a] ⓘ , Diana Tajik-Parvinchi [a] ⓘ & Robert A. Cribbie [a] ⓘ

[a]Department of Psychology, York University

Abstract ■ Statistical suppression occurs when adjusting for some third variable enhances or substantially modifies the association between an initial predictor and an outcome. Although many methodologists have discussed this phenomenon, very little work has examined suppression in longitudinal regression models such as the pretest-posttest design. This research addressed this gap with two separate studies. Study One was a literature review that reviewed 80 articles from a variety of fields within psychology. Study Two was an analysis of a large longitudinal clinical dataset via 925 statistical models. Both studies revealed consistent results: in approximately 20% of instances, suppression effects were observed and were attributable to the inclusion of a pretest measure. Results underscore that controlling for pretest measures when assessing change may be of value, as this can help clarify the associations between predictors and posttest outcomes.

Keywords ■ statistical suppression; Lord's paradox; pretest-posttest designs.

✉ lifarm@yorku.ca

Introduction

Statistical suppression occurs when the introduction of a third variable leads to a stronger or directionally opposite association between a predictor variable and an outcome variable. Many methodologists have debated the statistical origins of suppression over the decades (Arah, 2008; Horst et al., 1941; Conger, 1974; Courville & Thompson, 2001; Darlington & Hayes, 2017; Lubin, 1957; MacKinnon, Krull, & Lockwood, 2000; Pandey & Elliott, 2010; Velicer, 1978; Tu, Gunnell, & Gilthorpe, 2008; Tzelgov & Henik, 1991; Tzelgov & Stern, 1978), each with their own unique interpretation of the phenomenon. However, suppression has not been addressed with respect to longitudinal, pretest-posttest data; more specifically, it has not been investigated whether controlling for baseline measures in regression based models that regress posttest scores on predictors may lead to a statistical suppression effect. In other words, it is unclear if pretest measures can serve as suppressors of the relationship between a predictor and a posttest outcome. The main goal of this research is to address this gap in the literature, namely, to elucidate whether controlling for pretest measures may reveal (e.g., enhance) the relationship between a continuous predictor

and a posttest outcome. First, the common methods used to analyze pretest-posttest designs will be outlined. Next, two studies will be presented: 1) a literature review that explored how often statistical control of a pretest variable strengthens or alters the direction of the association between a predictor and posttest outcome, as well as what magnitude of suppression effects are observed; and 2) a reanalysis of a longitudinal clinical data set to explore the prevalence and magnitude of pretest suppressor effects.

Introduction to Pretest-Posttest Designs

Researchers are often interested in the association between a key predictor and change across time, where the predictor is either categorical or continuous. If the primary predictor is categorical, often baseline scores cannot be assumed to be equivalent (excluding groups formed through random assignment). These research designs are often analyzed with either a difference score model that uses the difference between pretest scores and posttest scores as the outcome variable, or a regression based analysis of covariance (ANCOVA) model that includes the pretest as a predictor, statistically removing the variability in posttest scores due to the pretest in order to compare hypothetical populations with equal pretest scores (Van

Breukelen, 2013).

However, given the same data, these models can lead to divergent conclusions, a phenomenon known as Lord's Paradox (Lord, 1967). In the absence of any real change, ANCOVA could conclude an effect of the predictor on change if baseline differences exist between groups, while a difference score approach would conclude no differences among the groups in the amount of change. As general linear models, they may be compared as regression equations with the posttest score as the outcome:

Difference score model: $post_i = b_0 + b_1 group_i + pre_i + e_i$

Regression based model: $post_i = b_0 + b_1 group_i + b_2 pre_i + e_i$

Of primary interest is the coefficient for the effect of the grouping variable, b_1. In the difference score model, b_1 compares the average change in the outcome across the groups. In the regression-based model, b_1 reflects the extent to which group membership predicts posttest scores, after holding pretest scores constant, and is influenced by the magnitude that pretest and group are related. The inclusion of the pretest coefficient, b_2, sets aside variability explained by the pretest scores, leaving only variability due to change (Oakes & Feldman, 2001), and will be equal to one if pretest scores perfectly predict posttest scores. However, the more unreliable the pretest scores (i.e., the more pretest scores include measurement error), the lower this coefficient will be (Gollwitzer, Christ, & Lemmer, 2014). Accordingly, the two models will lead to identical conclusions only if the pretest and grouping variable are unrelated *and* the pretest scores are perfectly reliable ($b_2 = 1$). Otherwise, the results of the models diverge, giving rise to Lord's Paradox. The larger these factors, the more ANCOVA will overestimate the effect of the predictor and diverge from the results of a difference score model. Difference scores can also be plagued with unique challenges, the details of which may be found in Edwards (2001).

Methodologists have debated Lord's Paradox for decades (e.g., Campbell & Kenny, 1999; Van Breukelen, 2013; Wainer & Brown, 2004; Werts & Linn, 1969, 1971; Wright, 2006), generally interpreting the phenomenon in one of two ways. Some suggest the label *paradox* is not necessary as the conclusions of the models are not incompatible (Bock, 1975; Cox & McCullagh, 1982; Pearl, 2014; Wijayatunga, 2017, among others); the difference-based model predicts change in *group* averages, while ANCOVA makes a prediction about how *individuals* from distinct groups with the same pretest scores change across time. However, others insist that ANCOVA leads to artifactual results for the effect of the key grouping predictor when groups are non-equivalent at baseline, and thus recommend not

to control for pretest when related to the pretest of interest (e.g., quasi-experimental designs; Van Breukelen, 2013; Wainer & Brown, 2004), but recommend its use when predictor and pretest are not correlated, since this increases statistical power (Oakes & Feldman, 2001), or when treatment allocation in experimental studies is based on initial scores (Campbell & Kenny, 1999; Wright, 2006).

Erroneous findings of effects when pretest is controlled may also arise in multiple regression models with continuous predictors of change (Campbell & Kenny, 1999; Castro-Schilo & Grimm, 2018; Eriksson & Häggström, 2014; Gollwitzer et al., 2014; Van Breukelen, 2013). For instance, Eriksson and Häggström (2014) found that if the pretest is measured with error and is related to the continuous predictor, controlling for pretest leads to an overestimation of the effect of the continuous predictor. Even when an effect of a predictor is present, including a pretest covariate can lead to an overestimation of that effect (Castro-Schilo & Grimm, 2018), and the conditions necessary for observing this artifact are common within psychological research (Farmus, Arpin-Cribbie, & Cribbie, 2019). However, Lord's paradox observed with continuous predictors of change relates closely to another statistical phenomenon called suppression, outlined in the following section.

Introduction to Statistical Suppression Effects

The most general conceptualization of suppression is that statistical adjustment for a third (suppressor) variable leads to an increase in the magnitude of association between a predictor and outcome (Arah, 2008; Conger, 1974; MacKinnon et al., 2000). Typically, when an estimate between a predictor and outcome is adjusted for a third variable (e.g., a covariate or potential confounder), the association between the predictor and outcome decreases due to the removal of shared variability among the predictor and third variable—variability that is not accounted for in the raw bivariate association between predictor and outcome (i.e, the partial/semipartial correlation will be less than the raw correlation). In suppression, variability in a predictor (X_1) not shared with the outcome (Y) may serve to weaken their relationship. When this irrelevant variance is in fact shared with a third variable (X_2, the suppressor), statistically controlling for X_2 leads to an increase in the magnitude of the relationship between X_1 and Y (Tzelgov & Henik, 1991). Thus, omitting suppressors can reduce the predictive power of a model and decrease the magnitude of partial regression coefficients (Cohen & Cohen, 1983), thereby increasing the probability of a Type II error (Horst et al., 1941).

To continue with this conceptualization of statistical suppression, a suppressed variable can be characterized as having a squared semipartial correlation that is larger

than its corresponding squared zero-order correlation, $r^2_{(Y,X_1)|X_2} > r^2_{Y,X_1}$, where the variable following the vertical bar (|, i.e., X_2) is conditioned on or partialled out of X_1 (Velicer, 1978). Similarly, we could also conceptualize suppression using the partial correlation; however, here, we focus on the semipartial correlation as an effect size measure in regression, given this measure's popularity. Similarly, an estimated standardized partial regression coefficient being larger in magnitude than its corresponding raw correlation also indicates suppression (i.e., $|\hat{\beta}_{(Y,X_1)|X_2}| > |r_{Y,X1}|$, with $\hat{\beta}$ and r having the same sign; Tzelgov & Stern, 1978). These conditions suggest that a suppressor is characterized primarily by its impact on another variable, rather than its own relation to the outcome.

Beyond this basic conceptualization of statistical suppression, there are four types of suppression according to the signs and magnitudes of the bivariate relations among the suppressor, predictor, and outcome. Absolute suppression occurs when the estimated standardized regression coefficient or the semipartial correlation is larger in magnitude than the raw correlation ($|\hat{\beta}_{(Y,X_1)|X_2}| > |r_{Y,X_1}|$ or $|r_{(Y,X_1)|X_2}| > |r_{Y,X_1}|$), but the suppressor may be related to the outcome (Tzelgov & Henik, 1991). This is the form of suppression described above. A stricter form of absolute suppression, called classical suppression, occurs when the suppressor is unrelated to the outcome ($r_{Y,X_2} \approx 0$). Negative suppression (Darlington, 1968; Lubin, 1957) occurs when a predictors' sign reverses after a third variable is statistically controlled ($\hat{\beta}_{(Y,X_1)|X_2}$ or $r_{(Y,X_1)|X_2}$ is opposite in sign to r_{Y,X_1}). Here, the suppressor is related to both the predictor and the outcome, and all variables have positive bivariate correlations or negative bivariate correlations. Mutual or reciprocal suppression (Conger, 1974) occurs when the estimated absolute standardized regression coefficients (or semipartial correlations) for two predictors are both larger in magnitude than (but have the same sign as) their respective bivariate correlations with the outcome ($|\hat{\beta}_{(Y,X_1)|X_2}| > |r_{Y,X_1}|$ and $|\hat{\beta}_{(Y,X_2)|X_1}| > |r_{Y,X_2}|$ or $|r_{(Y,X_1)|X_2}| > |r_{Y,X_1}|$ and $|r_{(Y,X_2)|X_1}| > |r_{Y,X_2}|$). Here, each predictor suppresses the other; the two predictors are negatively correlated with one another, but each are positively related to the outcome (or positively related to one another and negatively related to the outcome).

Suppression in Psychology Research

Although statistical suppression has been addressed and detected by methodologists in many disciplines within psychology, including personality (Hicks & Patrick, 2006; Watson, Clark, Chmielewski, & Kotov, 2013), clinical (Gaylord-Harden, Cunningham, Holmbeck, & Grant, 2010), experimental (Brown & Coyne, 2017), and forensic settings (Blonigen et al., 2010), the phenomenon is generally undetected,

underreported, and not well understood within the behavioral sciences (Gutierrez-Martinez & Cribbie, 2019; Pandey & Elliott, 2010) because researchers tend to examine bivariate associations among variables of interest first to explore and to assess potential relationships. When a researcher anticipates but does not find a relationship at the bivariate level, they may simply discard those variables before exploring further. Likewise, when a set of candidate variables are tested in an exploratory research context, those variables not substantially related to an outcome are often ignored. Even if led by a theory on how a group of variables should interact, non-significance often leads investigators to turn their attention away from predictors (Koeske, 1998). However, a zero-order correlation may be weak in the presence of a true relationship for several reasons, including low reliability, invalid measures, a non-linear relationship, or an obscured moderation effect. Another possibility is that the predictor's contribution is only made clear when another variable (i.e., a suppressor) is statistically controlled.

### Suppression in Pretest-Posttest Designs

As outlined above, literature on Lord's Paradox suggests that controlling for pretest when it relates to the predictor may lead to an overestimation of the effects in both categorical or continuous predictors. Thus, this tenet would generalize to all multiple regression models that regress a posttest outcome on a continuous predictor and pretest scores. However, if controlling for pretest increases the association between a predictor and change, is this an artifact, or is it simply statistical suppression at play?

Through simulation, a suppression situation has been depicted in Figure 1. In the left graph, the bivariate scatterplot shows that an association between a continuous predictor and a posttest outcome is absent (i.e., $r = .014$). After controlling for pretest scores, there is a strong and positive association between the continuous predictor (CP) and the posttest (i.e., $r_{(post,CP)|pre} = .55$ and $\beta_{(post,CP)|pre} = 0.77$). In blue are those high on levels of the pretest, in green are those who are mid-level on the pretest, and in red are those low on pretest levels. Thus, looking at individuals with the same pretest levels, there is a strong association between the predictor and the outcome. This research is premised on the hypothesis that these types of situations are not rare, and that evidence will be found in current psychology studies.

The current research will explore the role of statistical suppression in regression-based pretest-posttest models with continuous predictors. Tu et al. (2008) have argued that suppression effects and Lord's paradox are different variations of "reversal paradox" since both pertain to reversal, diminishing, or strengthening of an associa-

Figure 1 ■ Scatterplots depicting associations before and after controlling for pretest illustrating a relation between a continuous predictor and a posttest outcome, unadjusted for pretest (left), and after adjustment for pretest (right). Squares represent those high on pretest, triangles represent those with moderate levels of pretest, and circles represent those with low levels of pretest.
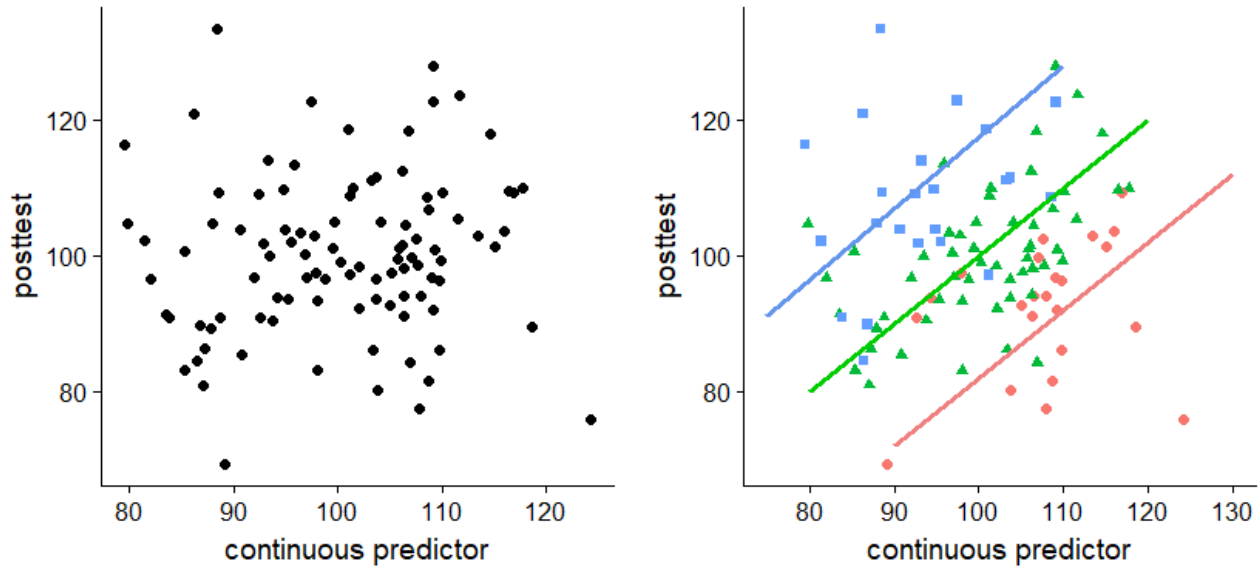
tion after statistically adjusting for a third variable. Our interest here is in the nature of the suppression effects with respect to prevalence, type, and magnitude. The primary research questions addressed through two studies are whether the inclusion of pretest measures within regression-based models is justified on the basis that it leads to statistical suppression effects. In Study One, a literature review was conducted to explore instances in which the pretest acted as a suppressor within social and behavioral science research, quantifying prevalence, the types of suppression that arise, whether researchers acknowledge that suppression has occurred, what fields of psychology experience suppression effects most often, and the magnitude of the suppression effects. In Study Two, a data set is reanalyzed to assess for frequency, type, and magnitude of suppression effects.

Although the body of literature on statistical suppression has demonstrated its effects and mechanisms, we know of no other study that has specifically explored the potential for pretest measures to act as suppressor, tracking its prevalence, nor the types of suppression that tend to occur when pretest does suppress the association between a predictor and posttest outcome. Thus, the novelty here is in exploring these characteristics. A tertiary interest pertains to whether the relationship between the predictor and suppressor is influential on the magnitude of the

suppression effects, as would be suggested by literature on Lord's paradox with continuous predictors.

Study One: Literature Review

A literature review was conducted using psychology journals published between 2008 and 2017. Articles that adopted multiple regression models with a posttest measure as the outcome and the corresponding pretest measure and another continuous variable as predictors were included in the study. Only two predictors were allowed in this situation because if the continuous predictor's association with the posttest increased with the addition of the pretest (or vice versa), then the pretest (or continuous predictor) must be a suppressor. With more than one predictor, in addition to the pretest, it becomes difficult to identify which predictor is acting as a suppressor. Thus, if we observe an increase in the association between a continuous predictor and a posttest outcome, and the pretest is the only other predictor in the model, we can be certain that the pretest is the suppressor variable.

Google Scholar was used to search for articles that included the terms *psychology, regression, correlation*, and variations for the labels *pretest* and *posttest* (i.e., *T1/T2, Time 1/Time 2, baseline/follow-up*). The abstracts were reviewed to determine if the studies were of a longitudinal nature, and if so, the full article was scanned for a regres-

sion model including a pretest covariate, a continuous predictor, raw correlations, and standardized partial regression coefficients. It was also recorded if partial or semipartial correlations were reported instead of the standardized partial coefficient. In order to ensure that we had a sufficient sample, we continued until 80 articles were found that met our full inclusion criteria and could be assessed for suppression. Given the strict inclusion criteria, finding 80 articles was extremely challenging; over 1000 articles were reviewed in order to find 80 that met the required criteria.

For the purposes of the literature review, our statistical criteria for determining if suppression was present was $|\hat{\beta}_{(post,CP)|pre}| > |r_{post,CP}|$ (both having the same sign), $|\hat{r}_{(post,CP)|pre}| > |r_{post,CP}|$ (both having the same sign), or if $\hat{\beta}_{(post,CP)|pre}$ or $\hat{r}_{(post,CP)|pre}$ were opposite in sign to $r_{post,CP}$. Furthermore, if there was enough information to ascertain the type of suppression (i.e., in addition to the article reporting $r_{post,CP}$ and $\hat{\beta}_{(post,CP)|pre}$ or $\hat{r}_{(post,CP)|pre}$, $r_{post,pre}$ and either $\hat{\beta}_{(post,pre)|CP}$ or $\hat{r}_{(post,pre)|CP}$ were also reported), we defined the types as follows: absolute ($|\hat{\beta}_{(post,CP)|pre}|$ or $|\hat{r}_{(post,CP)|pre}| > |r_{post,CP}|$ and $\left|\hat{\beta}_{(post,pre)|CP}\right|$ or $|\hat{r}_{(post,pre)|CP}| \leq |r_{post,pre}|$); classical ($|\hat{\beta}_{(post,CP)|pre}|$ or $|\hat{r}_{(post,CP)|pre}| > |r_{post,CP}|$ and $r_{post,pre} = 0$ or near 0); negative ($\hat{\beta}_{(post,CP)|pre}$ or $\hat{r}_{(post,CP)|pre}$ opposite in sign to $r_{post,CP}$; whether the pretest increased in magnitude or not did not change the classification as negative suppression); and mutual ($|\hat{\beta}_{(post,CP)|pre}|$ or $|\hat{r}_{(post,CP)|pre}| > |r_{post,CP}|$ and ($\left|\hat{\beta}_{(post,pre)|CP}\right|$ or $|\hat{r}_{(post,pre)|CP}| > |r_{post,pre}|$). As indicated above, for absolute, classical or mutual suppression the partial relationships must have the same sign as the raw relationship. Squared semipartial correlations were also accepted (however, being absolute values, these preclude the possibility of ascertaining negative suppression).

The prevalence of the suppression effects was recorded to determine how often the inclusion of the pretest covariate led to a suppression effect. The magnitude of suppression was computed as the difference in magnitude between the partial coefficient for the predictor and the correlation ($|\beta_{(post,CP)|pre}| - |r_{post,CP}|$). Further, it was recorded how often authors explicitly mentioned suppression effects or acknowledged that the phenomenon observed is unusual.

Literature Review Results

Descriptives. Eighty articles over 52 different journals were identified using the search and inclusion criteria. Of these articles, the primary topic most often fit within the context of either developmental ($n = 25$ or 31.25%), clinical ($n = 20$ or 25%), or social/personality ($n = 17$ or 21.25%) psychology domains. Other articles included research within educational psychology ($n = 5$ or 6.25%), applied psychology ($n = 9$ or 11.25%), or cognitive psychology ($n = 4$ or 5%). The finding that the majority of articles meeting inclusion criteria were extracted from either clinical or developmental journals is somewhat expected given that research in these fields may more often focus on questions of a longitudinal nature.

Incidences of Statistical Suppression. Of the 80 articles coded, 18% ($n = 14$) showed evidence that the pretest variable was acting as a suppressor on the relationship between the continuous predictor and the posttest outcome. Four of these articles reported two separate instances of suppression, and thus there were 18 total instances of suppression recorded. The increase in magnitude was at least 0.10 in seven of these instances (39%), and 9% of all articles. In the case of negative suppression, if the raw correlation was $r = .04$ and $\beta = -0.08$ (as an example), then the magnitude of suppression was considered greater than 0.10 since the change is 0.12 (.04 – [-0.08]).

Of the 18 instances of suppression, 13 (72%) provided enough information to assess the type of suppression, with six instances of mutual suppression (33%), four of absolute suppression (22%), and three of negative suppression (17%; see Table 1). None of the studies included squared semipartial correlations, and so there was no chance for misclassification of instances of negative suppression as absolute suppression.

Thus, the literature review provides evidence that in about 20% of instances, inclusion of the pretest leads to suppression of the relationship between a continuous predictor and a posttest outcome, and in about 10% of instances the association is strengthened/changed by at least .10 (in standardized/correlation units).

Study Two: Data Reanalysis

The results of the literature review may indicate instances of the pretest variable as a suppressor. However, we also chose to examine the same research question using recently collected longitudinal data. The dataset was derived from a 10-week randomized controlled trial using cognitive behavioural therapy to improve emotion regulation among autistic children (Weiss et al., 2018). Fifty-eight parent-child dyads participated. The children were between 8 and 12 years of age ($M = 9.69$, $SD = 1.26$) and mostly male (90.9%). Parents (83.6% female) were between 35 and 52 years of age ($M = 43.46$, $SD = 4.09$).

Several outcomes and predictors were explored, including clinical, developmental, and parent coregulation measures. A series of multiple regression models were conducted in which posttest outcomes were regressed on

Table 1 ■ The Number of Studies per Area with Evidence of Absolute, Negative or Mutual Suppression.

Journal	Studies with unidentifiable suppression	Studies with identifiable suppression	Type of Suppression		
			Absolute	Negative	Mutual
clinical	0	3	1	1	1
developmental	4	4	1	1	2
social/personality	0	4	2	1	1
applied	0	1	0	0	1
cognitive/neuropsychology	0	0	0	0	0
educational	1	1	0	0	1

their corresponding pretest measure and one other predictor to determine whether the inclusion of the pretest resulted in an increased (or reversed) association between the predictor and the posttest outcome.

Following the results of Study One, we expect a similar proportion of suppression effects ($\approx$ 20%) arising in the reanalysis study. Based on previous literature that regression artifacts tend to arise as a function of the correlation between the pretest and the predictor, graphs will depict whether the magnitude of the suppression effect is stronger based on the correlation between the pretest and the predictor.

Data Reanalysis Results

A list of pretest-posttest and predictor variables were selected by the authors of the primary study which were of theoretical interest. The list was comprised of 25 pairs of outcomes measured at pretest and posttest and 38 total predictors of change, including developmental (e.g., verbal reasoning ability measured by the Vocabulary subtest of the Full Scale-2 [FSIQ-2] from the Wechsler Abbreviated Scale of Intelligence, Second Edition; WASI-II; Wechsler, 2011); clinical (e.g., inhibition and coping subscale scores from the *Child Emotion Management Scales* (CEMS; Zeman, Cassano, Suveg, & Shipman, 2010), internalizing and externalizing subscale scores from the *Behavior Assessment System for Child, Second Edition – Parent Rating Scales* (BASC-2 PRS; Reynolds & Kamphaus, 2004), social cognition and social communication subscale scores from the *Social Responsive Scale – Second Edition School-Age Form* (SRS-2; Constantino & Gruber, 2012), and parent psychopathology (i.e., subscale scores from the Depression, Anxiety, and Stress Scale (DASS; Lovibond & Lovibond, 1995) factors, as well as Lability/Negativity and Emotion Regulation subscale scores from the *Emotion Regulation Checklist* (ERC; Shields & Cicchetti, 1997). The full list of variables is provided in supplementary material housed at https://osf.io/sc9tr/.

The posttest outcomes were regressed on their respective pretest measure plus a single additional pretest predictor (pretest measures also became the main predictors in other models examining other pretest-posttest pairs). Thus, 925 multiple regression models were conducted (25×37), of which 22.5% ($n = 208$) indicated a suppression effect. These results on the incidence of suppression closely mirror those found in the literature review study above. Of these 208 models, 46% ($n = 96$) indicated negative suppression, 20% ($n = 42$) indicated absolute suppression, and 34% ($n = 70$) indicated mutual suppression. The pretest and posttest were, as expected, always (moderately to strongly) related (range between $r = .36$ and $r = .90$), and hence none of the models met the criteria for classical suppression (which requires the suppressor and outcome to be unrelated). In terms of the magnitude of the suppression effect, 24% ($n = 50$) of $\beta_{(post,CP)|pre}$ were larger in magnitude than $r_{post,CP}$ by at least 0.10 and 33% ($n = 68$) were at least 0.05 larger.

In order to provide some context regarding the nature of the detected statistical suppression effects, three specific examples of suppression effects from these analyses were selected. The first instance involves posttest inhibition scores from the CEMS scale regressed on pretest DASS Parent Stress scores, where initially there was a weak bivariate association of $r = -.09$. When pretest CEMS Inhibition was added, the association between posttest CEMS Inhibition and DASS Parent Stress increased to $\beta = -0.21$ and $\hat{r}_{(postCEM,DAS)|preCEM} = -.21$. That is, among youth with autism with the same levels of pretest CEM Inhibition, greater pretest DASS Parent Stress levels predict lower posttest CEMS Inhibition levels (Figure 2). A second example that highlights an instance of absolute statistical suppression is when posttest emotion regulation (ER subscale of ERC) is regressed on verbal reasoning (VR) and pretest ER. The correlation between posttest ER and pretest VR is $r = .04$, but the semipartial correlation increases to $\hat{r}_{(post-ER,VR)|pre-ER} = .18$ and its standard-

Figure 2 ■ Scatterplots depicting the association between posttest inhibition and pretest stress. On the left is the bivariate association between posttest CEM inhibition and DASS stress. On the right is the association between posttest CEM inhibition and DASS stress after introducing pretest CEM inhibition. Red circles are participants low on levels of pretest CEM inhibition, green triangles are participants with moderate levels of pretest CEM inhibition, and blue squares are those high on pretest CEM inhibition.
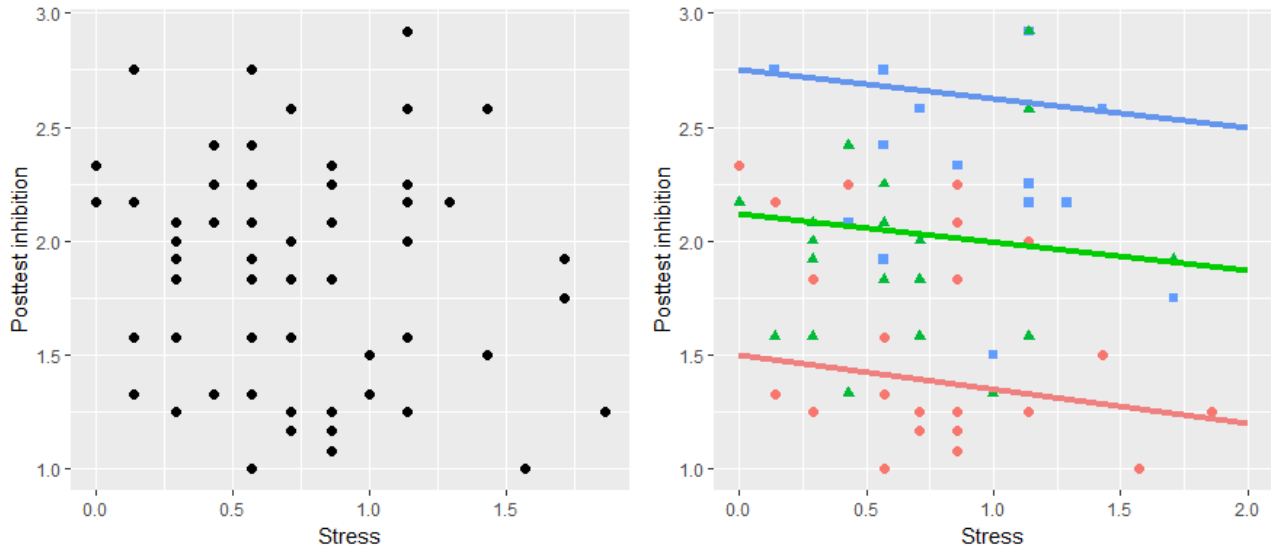
ized coefficient to $\beta = 0.19$ when the pretest ER is held constant. That is, among participants with the same levels of pretest ER, higher pretest VR scores predict higher posttest ER scores (Figure 3). The graphs depict slopes for the pretest when it is cut into low, medium and high levels of the measure. Lastly, posttest scores on the *Emotion Regulation and Social Skills Questionnaire* (ERSSQ; Beaumont & Sofronoff, 2008) were regressed on pretest severity scores from the Anxiety Disorders Interview Schedule: Parent Interview - Fourth Edition (ADIS-P IV; Silverman & Albano, 1996). The initial raw correlation between posttest ERSSQ and pretest ADIS-P severity was $r = -.12$ (Figure 4). With the addition of pretest ERSSQ in the model, the coefficients for ADIS-P severity increased to $\beta = 0.22$ and $\hat{r}_{(post-ERSSQ,ADIS)|pre-ERSSQ} = .20$. Here, we see a reversal in signs as well as a larger magnitude in absolute value for the association between emotion regulation and symptom severity, after accounting for pretest emotion regulation. These are just three examples demonstrating how the inclusion of pretest changes the magnitude and/or sign of the association between a predictor and a posttest outcome.

A tertiary purpose of this study was to look at whether the observed suppression effects are a function of the magnitude of association between the pretest measure and the predictor. For instances where the signs were the same (i.e., absolute and mutual), we calculated the magnitude of the suppression effect by subtracting the absolute value of the correlation between the predictor and posttest from the standardized partial regression coefficient for the predictor controlling for the pretest ($|\beta_{(post,CP)|pre} - r_{post,CP}|$). For instances where the sign of the predictor changed (i.e., negative suppression), we calculated suppression as the total degree of unit change. As expected, we found that the magnitude of suppression effects tended to be stronger when there was a stronger association between the predictor and the outcome for all three types of suppression recorded (Figure 5). Generally, absolute suppression occurred with weaker correlations among the predictor and the pretest, whereas mutual and negative suppression occurred with stronger relations among the predictor and the predictor . However, as noted already, our definition of negative suppression may be too liberal in that all sign flips for predictors were labelled as negative suppression, even if the pretest coefficient was stronger than its bivariate correlation with post (which could also be labelled as mutual suppression).

Discussion

Researchers assessing pretest-posttest change often need to decide whether to control for pretest variables or use raw change scores as the outcome variable. This research

Figure 3 ■ Scatterplots depicting the association between posttest emotion regulation and pretest verbal reasoning. On the left is the bivariate association between posttest emotion regulation (ER) and verbal reasoning. On the right is the association between posttest ER and pretest verbal reasoning after introducing pretest ER. Red circles are participants low on pretest ER, green triangles are participants with moderate levels of pretest ER, and blue squares are those high on pretest ER.
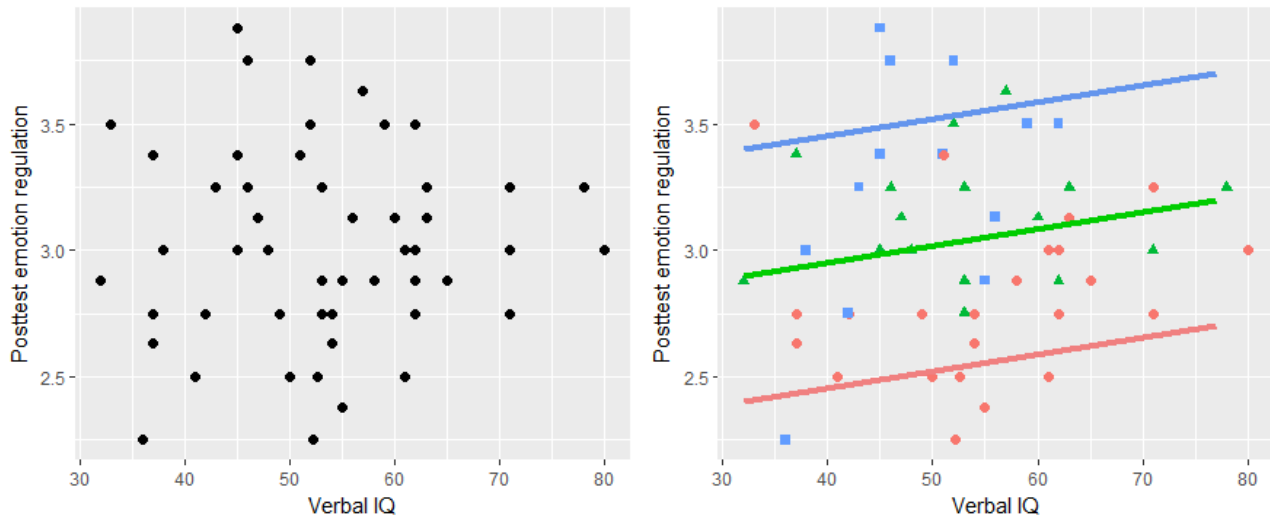
sought evidence that pretest measures may act as suppressor variables, both within published psychology literature and in data analyzed independently. The results of the literature review (Study One) suggest that suppression is not a rare phenomenon within pretest-posttest designs. Whenever pretest measures of psychological constructs are utilized, there is the potential to observe statistical suppression across a wide variety of research contexts and disciplines within psychology. This occurred at a rate of approximately 20%, with 9% of articles showing an increase of at least .10 in absolute magnitude. The most common type of suppression was mutual, whereby both the predictor's association with posttest and the pretest's association with posttest was strengthened with the inclusion of the other variable, leading to each accounting for error in the other. A third of the articles did not include enough information to assess the type of suppression, which underscores the necessity for researchers and journals to require full reporting of bivariate and conditional relationships among variables.

Study Two involved reanalysing a longitudinal dataset based on a treatment cycle of CBT for children with autism between the ages of eight and twelve. The data included a variety of clinical, demographic and parental variables measures before and after treatment, and provided the opportunity to examine the role of pretest measures in elucidating important relations among predictors and out-

comes following therapy. Studies of this nature may be analysed with a difference score model or a regression-based model, the latter of which was the focus of this study. Consistent with the results of the literature review, the prevalence of the pretest suppressing the association between a predictor and outcome was about 20%. However, unlike the results of Study One, most of the instances were negative suppression (46%). Furthermore, 24% of the suppression effects were characterized by a magnitude of at least 0.10. This finding may be somewhat expected, given that pretest measures are often highly correlated with posttest measures, and their inclusion in a regression will naturally lead to a large chunk of the variability 'pie' being consumed, leaving less variability to be accounted for by the suppression effect observed for the continuous predictor.

One potential explanation for the patterns observed in Study 2 is that there is a higher prevalence of contemporaneous correlations among the models that result in suppression effects relative to models that do not lead to suppression. A contemporaneous correlation occurs when variables that are measured at the same time are more correlated than variables measured at different times. Specifically, there may be contemporaneous correlations between pretest and predictors given that they were measured together prior to the intervention targeting emotional regulation. For instance, the mood of the parent reporting on

Figure 4 ■ Scatterplots depicting the association between posttest emotion regulation and pretest ADIS severity. On the left is the bivariate association between ADIS severity and posttest emotion regulation. On the right is the association between posttest ERSSQ and pretest ADIS after introducing pretest ERSSQ. Red circles are participants low on pretest ERSSQ, green triangles are participants with moderate levels of pretest ERSSQ, and blue squares are those high on pretest ERSSQ.
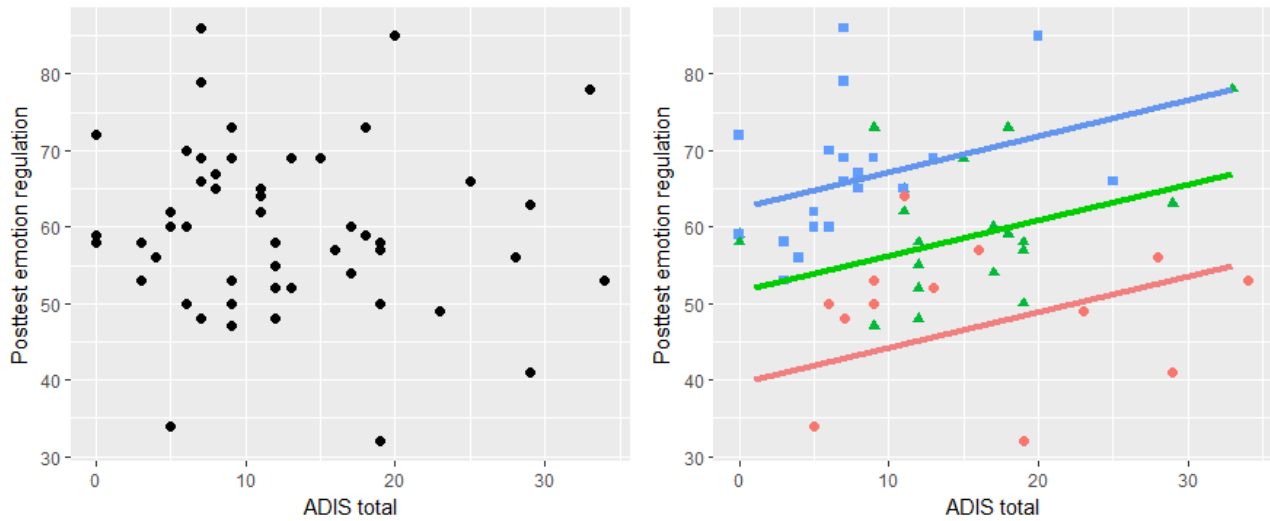

the child's symptoms may influence how they respond to both the pretest measure and predictor measure and create shared variability among these measures that is contemporaneous. By controlling for pretest, we effectively remove contemporaneous variation between pretest and predictor that contributes noise to the predictor-posttest relationship, leading to an enhancement of the association between predictor and change. Furthermore, we observe that a weaker pretest-posttest association is correlated with a larger magnitude of suppression ($r = -.43$), suggesting that the more variability in the pretest that is contemporaneous, controlling for pretest results in stronger associations between predictor and change.

Results from both studies highlight that it might be advantageous to include pretest measures when assessing predictors of change. Although previous work (e.g., Erikson & Häggström, 2014; Farmus, Arpin-Cribbie, & Cribbie, 2019) on Lord's Paradox warns against controlling for the pretest whenever it relates to a predictor to evade spurious associations arising between the predictor and outcome, our findings suggest the potential importance of pretest measures to help clarify a predictor's relation to change. Our findings suggest that what many methodologists would label as a spurious association may in fact be statistical suppression at play, an entirely legitimate regression phenomenon, and leading researchers to miss meaningful relationships and inaccurately depict associa-

tions between predictors and outcomes.

Our findings also highlight that predictors should not be disregarded based on weak bivariate associations with change outcomes. If theory links a predictor to change, researchers should compare the regression coefficients and semipartial correlations that control for the pretest measure to the bivariate correlations to determine if the pretest is accounting for irrelevant variance in the predictor, thereby allowing for a better estimate of the association with the posttest that is reduced of noise.

This research has focused on associations, but one important consideration is whether the substantive interest pertains to modeling of causal effects. In situations where the pretest is not a mediator in the causal path between a predictor and posttest outcome, it may be an instrumental variable that has a causal effect on the predictor, which then itself has a causal effect on the posttest scores. If the pretest is an instrumental variable, it should not be controlled for as it may bias the causal effects of predictor on posttest (Kim, 2019). Researchers should be cautious about including pretest scores in regression-based models when it is possible that the pretest can act as a mediator or instrumental variable (Pearl, 2014). If however, the pretest is a confounder, than it should be controlled for. An acyclic causal diagram may assist researchers in their decision to include pretest measures when there is an underlying causal structure to the model.
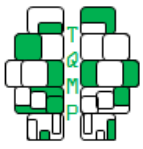
Figure 5 ■ Scatterplot depicting the magnitude of suppression effects predicted from the association between continuous predictors and pretest measures. Blue squares are instances of absolute suppression, red circles are instances of mutual suppression, and green triangles are instances of negative suppression.
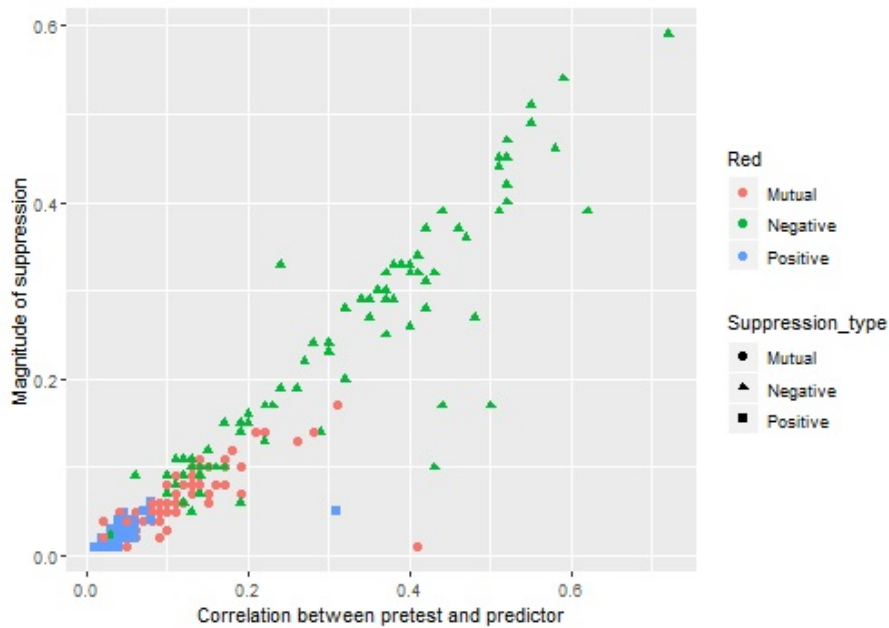
One natural question that may arise from this research is whether it is possible to differentiate between results that arise after the research question has changed due to the statistical model changing. Regressing posttest on a continuous predictor asks a different research question than regressing posttest on a continuous predictor after holding constant pretest scores. Whereas the first question is a simple examination of the association between a predictor and a posttest measure, adding a pretest covariate asks whether the predictor is associated with posttest scores *among subgroups of individuals with the same pretest scores*. Therefore, a reasonable interpretation of our results is that because the models address different research questions and hypotheses, the results are *expected* to change, precluding the potential that the findings can be attributed to statistical suppression effects. However, when there is no relationship between the pretest and the predictor, then the partial regression coefficient for the predictor should equate the raw bivariate relation between the predictor and posttest (Darlington & Hayes, 2017). In this situation, that relationship remains the same whether one adds the pretest or not, despite the changing nature of the research question. Hence, suppression by the pretest can only occur in the presence of some relationship between the predictor and the pretest, and not due to the changing research question. Note that what appears to be suppression by simply adding pretest to the model cannot occur if the pretest and predictor are unrelated. Thus, we can rule out that the suppression effects observed are strictly due to the changing nature of the question the model addresses.

Some key limitations to this research should be noted. Only 80 studies were included in Study One due the strict inclusion criteria required to detect the nature and type of suppression. For example, the nature of our research question necessitated an examination of regression based models that were restricted to a pretest and a single predictor, in order to identify whether the pretest was the suppressor. However, many authors do not report basic correlations or standardized regression coefficients or semipartial correlations. Study Two was based on a single clinical sample, and so the results may not generalize to all clinical samples or samples from other populations. However, the numerous variables were correlated to varying degrees, which allowed us greater range to explore our research hypotheses.

To date it has not been recommended that researchers use regression-based models when the pretest and the predictor are correlated. However, if it is possible that the pretest variable may act as a suppressor then it is ac-

tually recommended that researchers adopt regression-based models. Particularly, our results demonstrate that pretest inclusion can dramatically change the interpretation and magnitude of relations between predictors and posttest outcomes, which should be interpreted within the context/framework of the research. This clarification can help researchers gain greater understanding of substantive phenomenon. More rigorous research methods may be used to investigate situations where it is, or is not, recommended that a regression-based model is utilized is warranted and encouraged.

References

Arah, O. (2008). The role of causal reasoning in understanding simpson's paradox, lord's paradox, and the suppression effect: Covariate selection in the analysis of observational studies. *Emerging Themes in Epidemiology*, *5*, 1–5. doi:10.1186/1742-7622-5-5

Blonigen, D. M., Patrick, C. J., Douglas, K. S., Poythress, N. G., Skeem, J. L., Lilienfeld, S. O., ... Krueger, R. F. (2010). Multimethod assessment of psychopathy in relatiin to factors of internalizing and externalizing from the personality assessment inventory: The impact of method variance and suppressor effects. *Psychological Assessment*, *22*(1), 96–107. doi:10 . 1037 / a0017240

Bock, R. D. (1975). *Multivariate statistical methods in behavioural research*. New York: McGraw-Hill.

Brown, N. J. L., & Coyne, J. C. (2017). Emodiversity: Robust predictor of outcomes or statistical artifact? *Journal of Experimental Psychology: General*, *146*(9), 1372–1378. doi:10.1037/xge0000330

Campbell, D. T., & Kenny, D. A. (1999). *A primer on regression artifacts*. New York, NY: Guilford Press.

Castro-Schilo, L., & Grimm, K. J. (2018). Using residualized change versus difference scores for longitudinal research. *Journal of Social and Personal Relationships*, *35*(1), 32–58. doi:10.1177/0265407517718387

Conger, A. J. (1974). A revised definition for suppressor variables: A guide to their identification and interpretation. *Educational and Psychological Measurement*, *34*(1), 35–46. doi:10.1177/001316447403400105

Constantino, J. N., & Gruber, C. P. (2012). *Social responsiveness scale-second edition (srs-2)*. Torrance, CA: Western Psychological Services.

Courville, T., & Thompson, B. (2001). Use of structure coefficients in published multiple regression articles: ß is not enough. *Educational and Psychological Measurement*, *61*(2), 229–248. doi:10.1177/0013164401612006

Cox, D. R., & McCullagh, P. (1982). Some aspects of analysis of covariance. *Journal of the Biometric Society*, *38*(3), 541–561.

Darlington, R. B., & Hayes, A. (2017). *Regression analysis and linear models: Concepts, applications, and implementation*. New York: The Guilford Press.

Edwards, J. R. (2001). Ten difference score myths. *Organizational Research Methods*, *4*(3), 265–287. doi:10.1177/109442810143005

Eriksson, K., & Häggström, O. (2014). Lord's paradox in a continuous setting and a regression artifact in numerical cognition research. *PLoS One*, *9*(4), 095949. doi:10.1371/journal.pone.0

Farmus, L., Arpin-Cribbie, C. A., & Cribbie, R. A. (2019). Continuous predictors of pretest-posttest change: Highlighting the impact of the regression artifact. *Frontiers in Applied Mathematics and Statistics*, *4*(64), 1–11. doi:10.3389/fams.2018.00064

Gaylord-Harden, N. K., Cunningham, J. A., Holmbeck, G. N., & Grant, K. E. (2010). Suppressor effects in coping research with african American adolescents from low-income communities. *Journal of Consulting and Clinical Psychology*, *78*(6), 843–855. doi:10.1037/a0020063

Gollwitzer, M., Christ, O., & Lemmer, G. (2014). Individual differences make a difference: On the use and the psychometric properties of difference scores in social psychology. *European Journal of Social Psychology*, *44*, 673–682. doi:10.1002/ejsp.2042

Gutierrez-Martinez, N., & Cribbie, R. A. (2019). Incidence and interpretation of statistical suppression in the educational and behavioural sciences. Retrieved, from https://doi.org/10.31234/osf.io/rke62

Hicks, B. M., & Patrick, C. J. (2006). Psychopathy and negative emotionality: Analyses of suppressor effects reveal distinct relations with emotional distress, fearfulness, and anger-hostility. *Journal of Abnormal Psychology*, *115*(2), 276–287. doi:10.1037/0021-843X.115.2.276

Horst, P., Wallin, P., Guttman, L., Wallin, F. B., Clausen, J. A., Reed, R., & Rosenthal, E. (1941). *The prediction of personal adjustment*. Washington DC: Social Science Research Council.

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the depression anxiety stress scales (dass) with the beck depression and anxiety inventories. *Behaviour research and therapy*, *33*(3), 335–343.

Lubin, A. (1957). Some formulae for use with suppressor variables. *Educational and Psychological Measurement*, *17*(2), 286–296. doi:10 . 1177 / 001316445701700209

MacKinnon, D., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prevention Science*, *1*(4), 173–181. Re-

trieved from https://www.ncbi.nlm.nih.gov/pubmed/11523746

Oakes, M. J., & Feldman, H. A. (2001). Statistical power for nonequivalent pretest-posttest designs: The impact of change-score versus ancova models. *Evaluation Review*, *25*(1), 3–28. doi:10.1177/0193841X0102500101

Pandey, S., & Elliott, W. (2010). Suppressor variables in social work research: Ways to identify in multiple regression models. *Journal of the Society for Social Work and Research*, *1*(1), 28–40. doi:10.5243/jsswr.2010.2

Pearl, J. (2014). *Lord's paradox revisited – (oh lord! kumbaya!)* Retrieved from https://apps.dtic.mil/dtic/tr/fulltext/u2/a615058.pdf

Reynolds, C. R., & Kamphaus, R. W. (2004). *Behavior assessment system for children—second edition*. Circle Pines, MN: AGS.

Shields, A., & Cicchetti, D. (1997). Emotion regulation among school-age children: The development and validation of a new criterion q-sort scale. *Developmental psychology*, *33*(6), 906.

Tu, Y. K., Gunnell, D., & Gilthorpe, M. S. (2008). Simpson's paradox, lord's paradox, and suppression effects are the same phenomenon–the reversal paradox. *Emerging Themes in Epidemiology*, *5*(2), 1–9. doi:10.1186/1742-7622-5-2

Tzelgov, J., & Henik, A. (1991). Suppression situations in psychological research: Definitions, implications, and applications. *Psychological Bulletin*, *109*, 524–536.

Tzelgov, J., & Stern, I. (1978). Relationships between variables in three variable linear regression and the concept of suppressor. *Educational and Psychological Measurement*, *38*(2), 325–335. doi: 10.1177/001316447803800213

Van Breukelen, G. J. (2013). Ancova versus change from baseline in nonrandomized studies: The difference. *Multivariate Behavioral Research*, *48*(6), 895–922. doi:10.1080/00273171.2013.831743

Velicer, W. F. (1978). Suppressor variables and the semi-partial correlation coefficient. *Educational and Psy-chological Measurement*, *38*(4), 953–958. doi:10.1177/001316447803800415

Wainer, H., & Brown, L. M. (2004). Two statistical paradoxes in the interpretation of group differences: Illustrated with medical school admission and licensing data. *American Statistician*, *58*(2), 117–123.

Watson, D., Clark, L. A., Chmielewski, M., & Kotov, R. (2013). The value of suppressor effects in explicating the construct validity of symptom measures. *Psychological Assessment*, *25*(3), 929–941. doi:10.1037/a0032781

Wechsler, D. (2011). *Wechsler abbreviated scale of intelligence-second edition (wasi-ii)*. San Antonio, TX: NCS Pearson.

Weiss, J. A., Thomson, K., Burnham Riosa, P., Albaum, C., Chan, V., Maughan, A., . . . Black, K. (2018). A randomized waitlist-controlled trial of cognitive behavior therapy to improve emotion regulation in children with autism. *Journal of Child Psychology Psychiatry*, *59*(11), 1180–1192. doi:10.1111/jcpp.12915

Werts, C. E., & Linn, R. L. (1969). Lord's paradox: A generic problem. *Psychological Bulletin*, *72*(6), 423–425. doi:10.1037/h0028331

Werts, C. E., & Linn, R. L. (1971). Problems with inferring treatment effects from repeated measures. *Educational and Psychological Measurement*, *31*(4), 857–866. doi:10.1177/001316447103100407

Wijayatunga, P. (2017). Resolving the lord's paradox. In *Presented at the 32nd international workshop on statistical modelling* (Vol. 32), Rijksuniversiteit Groningen, Netherlands: Johann Bernoulli Institute.

Wright, D. B. (2006). Comparing groups in a before-after design: When t test and ancova produce different results. *British Journal of Educational Psychology*, *76*(3), 663–675. doi:10.1348/000709905X52210

Zeman, J., Cassano, M., Suveg, C., & Shipman, K. (2010). Initial validation of the children's worry management scale. *Journal of Child and Family Studies*, *19*(4), 381–392. doi:10.1007/s10826-009-9308-4

Open practices

The *Open Material* badge was earned because supplementary material(s) are available on https://osf.io/sc9tr/.

Citation