



Interpreting effect sizes when controlling for stability effects in longitudinal autoregressive models: Implications for psychological science

Paul Adachi & Teena Willoughby

To cite this article: Paul Adachi & Teena Willoughby (2015) Interpreting effect sizes when controlling for stability effects in longitudinal autoregressive models: Implications for psychological science, *European Journal of Developmental Psychology*, 12:1, 116-128, DOI: [10.1080/17405629.2014.963549](https://doi.org/10.1080/17405629.2014.963549)

To link to this article: <https://doi.org/10.1080/17405629.2014.963549>



Published online: 26 Sep 2014.



Submit your article to this journal [↗](#)



Article views: 917



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 91 View citing articles [↗](#)

Interpreting effect sizes when controlling for stability effects in longitudinal autoregressive models: Implications for psychological science

Paul Adachi and Teena Willoughby

Department of Psychology, Brock University, Ontario, Canada

Effect sizes in longitudinal studies often are *dramatically* smaller than effect sizes in cross-sectional studies. Indeed, autoregressive models (which are often used in longitudinal studies but not in cross-sectional studies) control for past levels on the outcome (i.e., stability effects) in order to predict change in levels of the outcome over time and thus may greatly reduce the magnitude of the effect of a predictor on the outcome. Unfortunately, however, there have been no attempts to differentiate guidelines for interpreting effect sizes for longitudinal studies versus cross-sectional studies. Consequently, longitudinal effect sizes that fall below the universal guidelines for “small” may be incorrectly dismissed as trivial, when they might be meaningful. In the current paper, we first review the present guidelines for interpreting effect sizes. Next, we discuss several examples of how controlling for stability effects can dramatically attenuate effect sizes of other predictors, in order to support our argument that the current guidelines may be misleading for interpreting longitudinal effects. Finally, we conclude by making recommendations for researchers regarding the interpretation of effect sizes in longitudinal autoregressive models.

Keywords: Longitudinal; Effect size; Autoregressive models.

Emphasis on the importance of reporting effect sizes in psychological research has increased over the past decade (Ferguson, 2009), particularly since the APA Task Force on Statistical Inference recommended that effect size estimates and their accompanying confidence intervals should be reported (Wilkinson & Task Force on Statistical Inference, 1999). The critical interpretation of these effect sizes, however, remains a problem (Peng, Chen, Chiang, & Chiang, 2013). Indeed, Rosenthal and Rubin in 1979 argued that while some researchers may discount small effects as trivial and not worth further investigation, small effects,

Correspondence should be addressed to Paul Adachi, Department of Psychology, Brock University, St. Catharines, Ontario L2S 3A1, Canada. E-mail: pa08fg@brocku.ca

in fact, may be meaningful depending on the context of the data. The interpretation of effect sizes when controlling for stability effects in longitudinal autoregressive models is one cogent example of a situation in which small effect sizes often are meaningful.

Specifically, the common practice of simply citing universal guidelines for interpreting the magnitude of effect size coefficients, without considering important factors such as the type of study design, has been criticized as too rigid and potentially misleading (Lipsey et al., 2012; Peng et al., 2013). Unfortunately, there have been no attempts to differentiate guidelines for interpreting effect sizes for different types of study designs, specifically longitudinal studies versus cross-sectional studies. This is a significant problem given that autoregressive models (which often are used in longitudinal studies but not in cross-sectional studies) control for past levels on the outcome variable (stability effects) in order to assess change in levels of the outcome variable (note that change here refers to inter-individual differences in change on the outcome, not mean level change), and thus may dramatically reduce the magnitude of the association between the predictor and the outcome. Hence, longitudinal effect sizes that fall below the universal guidelines for “small” may be incorrectly dismissed as trivial, when they might be meaningful. In fact, we have had several reviewers raise concerns about small effect sizes in our longitudinal autoregressive models. For example, one reviewer stated that “the effect sizes are quite small . . . which raises questions about the significance of the findings . . .” Similarly, another reviewer stated that “the path coefficients are incredibly small (most around .05–.07). Accordingly, the language in terms of effect sizes should be tempered throughout”. The goal of the current paper, therefore, is to highlight this gap in the literature and make a case for why very small effect sizes in longitudinal autoregressive models still may be important. First, we review the current guidelines for interpreting effect sizes, and then discuss why these guidelines may be misleading for interpreting effect sizes in longitudinal research when controlling for stability effects. We cite several examples from longitudinal studies to support our arguments. Finally, we conclude by making recommendations for researchers regarding the interpretation of effect sizes in longitudinal autoregressive models.

CURRENT GUIDELINES FOR INTERPRETING EFFECT SIZES AND THEIR APPLICATION TO LONGITUDINAL DESIGNS

Although there are several commonly used standardized effect size coefficients (e.g., bivariate r , β , d , f), we will restrict our discussion to the bivariate correlation coefficient (bivariate r) and the standardized regression coefficient (β), which are most relevant to cross-sectional and longitudinal correlational models (see Baguley, 2009, for a discussion of the potential advantages of interpreting unstandardized effects when the original units of measurement are

meaningful). Specifically, r and β represent product-moment correlation coefficients and regression coefficients, which are produced in cross-sectional and autoregressive regression/path models. In contrast, the often used effect size coefficient Cohen's d , for example, is most relevant for testing the difference between two independent means, such as in an experiment examining group differences on a dependent variable (Cohen, 1992).

Bivariate r is restricted to the association between two variables. Bivariate r may not accurately represent the true association between two variables, however, because there may be "third" variables that are driving the association that should be controlled. In contrast to bivariate r , β most often reflects the association between two variables after controlling for third variables in the analysis model. For example, when examining the association between video game play and aggression, it is important to control for gender, as males tend to play more video games and behave more aggressively than females (Ferguson, 2009; Willoughby, Adachi, & Good, 2011). Removing variance in aggression that is predicted by both gender and video game play thus reduces the magnitude of the predictive effect of video game play on aggression. An example from our data shows that a concurrent bivariate $r = .35$ between competitive video game play and aggression is attenuated to $\beta = .24$ after controlling for gender. It is important to note that when there are only two variables in the model (e.g., competitive video game play and aggression), the bivariate r is equal to the value of β . Guidelines for interpreting effect sizes generally do not distinguish between bivariate r and β that controls for other variables. For example, Cohen's (1992) guidelines for interpreting bivariate r are $.10 =$ small, $.30 =$ medium, and $.50 =$ large. Other researchers have suggested more stringent guidelines. For example, Ferguson (2009) suggests that for both bivariate r and β , $.20 =$ small, $.50 =$ medium, and $.80 =$ large. Importantly, however, neither Cohen nor Ferguson distinguish between effect sizes that control for stability effects (i.e., β in longitudinal designs) and those that do not (e.g., bivariate r or β in cross-sectional designs).

Longitudinal research designs in which autoregressive models are used have increased greatly in popularity over the past two decades (Ployhart & Ward, 2011) and are essential to the study of lifespan *development*. Autoregressive models not only allow researchers to control for covariates and other predictors, but also for stability effects (previous scores on the outcome). Indeed, controlling for stability effects is the gold standard of longitudinal designs as it enables researchers to examine whether the variable(s) of interest predicts the outcome over time, controlling for previous levels (the inter-individual differences in change) of the outcome (e.g., Taris, 2000). Specifically, to control for stability in the outcome variable, researchers must control for the product of (a) the stability path from the outcome at time 1 (T1) to the outcome at time 2 (T2) and (b) the initial correlation between the predictor at T1 and the outcome at T1 (i.e., the indirect path from the predictor at T1 to the outcome at T2, through the outcome

at T1). Thus, the focus of longitudinal autoregressive models is to predict inter-individual differences in *change* in levels of the outcome over time, as opposed to cross-sectional research where the focus is not on change. The disadvantage of autoregressive models, however, is that by controlling for previous scores on the outcome, a large portion of the variance in the outcome is removed. In fact, most behavioural (e.g., aggression: Olweus, 1979; drinking alcohol: Ouellette & Wood, 1998) and psychological (e.g., self-esteem: Marsh, 1993; intelligence: Thorndike, 1940) outcomes show strong stability over time (i.e., change is often gradual). Hence, the strongest predictor of behavioural or psychological outcomes often is their previous level of these outcomes (e.g., levels of aggression and self-esteem measures last year). Furthermore, the fact that controlling for stability effects removes variance in the outcome that is shared with the predictor (i.e., accounting for the correlation between the predictor at T1 and the outcome at T1) suggests that the effect size of the predictor on change in levels of the outcome at T2 likely will be small when there is at least moderate overlap between the predictor at T1 and the outcome at T1.

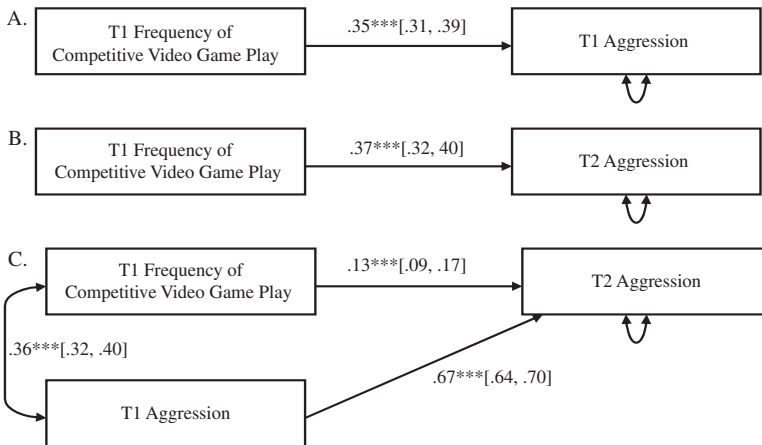
Just as one example, aggression has been shown to be quite stable over time (e.g., Cairns & Cairns, 1994; Dodge, Coie, & Lynam, 2006; Vaillancourt, Brendgen, Boivin, & Tremblay, 2003) and meta-analyses have demonstrated that the average longitudinal predictive effect of video game play on aggression is $r = .08$ (Ferguson, *in press*). If the correlation between the predictor at T1 and the outcome at T1 is small, however, then controlling for even high stability in the outcome should not greatly attenuate the predictive effect of the predictor at T1 on the outcome at T2. Yet, if the initial correlation between the predictor and T1 and the outcome at T1 is small, then the path between the predictor at T1 and the outcome at T2 often will be small as well. Nevertheless, small effects may still be meaningful when predicting change, as they can suggest, for example, that the predictor is associated with change in levels of the outcome over time during a particular period of development (e.g., adolescence), which cannot be assessed in a cross-sectional design. Furthermore, predictive effects on change in levels of the outcome may reflect an ongoing process of cumulative effects and thus may have a substantial impact on the outcome over time (e.g., Abelson, 1985; Cui, Donnellan, & Conger, 2007).

In addition, the goal of controlling for stability effects is conceptually different than the goal of controlling for a covariate that is a confound, such as gender. The purpose of controlling for gender is to remove extraneous variance in the outcome that is shared between the predictor and gender, in order to examine the unique effect of the predictor on the outcome. In contrast, the purpose of controlling for previous levels of the outcome variable is to predict change in levels of the outcome, not to remove extraneous variance in the outcome that is shared between the predictor and a confound. Therefore, it is misleading to use the same guidelines to interpret both longitudinal effect size coefficients from autoregressive models that control for stability effects and cross-sectional effect

size coefficients that control for confounds, but do not control for stability effects (see also Baguley, 2004, 2009, and Lenth, 2001, for additional arguments for avoiding the use of “canned” effect sizes when interpreting effects). Indeed, longitudinal effects often may fall below the current guidelines for a small effect, yet still may be meaningful. To illustrate this point, several examples from psychological research will be considered in which medium predictive effects are attenuated to very small effects by controlling for stability in the outcome.

EXAMPLES OF GREATLY ATTENUATED EFFECT SIZES WHEN CONTROLLING FOR STABILITY EFFECTS

We begin with an example from our own data in which we examine in path analyses whether frequency of competitive video game play predicts aggression among a large sample of adolescents. First, we examine the concurrent association between T1 frequency of competitive video game play and T1 aggression (see Figure 1A). The bivariate r is .35 for the predictive effect of T1 frequency of competitive video game play on T1 aggression. Next, we examine the longitudinal association between T1 frequency of competitive video game play and T2 aggression without controlling for stability effects (see Figure 1B). Similar to the concurrent association, the bivariate r is .37 for the predictive effect of T1 frequency of competitive video game play on T2 aggression (again, when there are only two variables in the model, the bivariate r is equal to the



Notes. T1 = Time 1; T2 = Time 2. Standardized coefficients (95% confidence intervals are in brackets) are reported. * $<.05$, ** $<.01$, *** $<.001$.

Figure 1. Results for models assessing the concurrent and longitudinal associations between frequency of competitive video game play and aggression.

value of β). Finally, we conduct an autoregressive path analysis to examine the longitudinal association between T1 frequency of competitive video game play and T2 aggression, controlling for stability effects (see Figure 1C). The stability path from T1 aggression to T2 aggression is $\beta = .67$, and the concurrent association between T1 frequency of competitive video game play and T1 aggression is bivariate $r = .36$. Thus, the indirect path from T1 frequency of competitive video game play on T2 aggression through T1 aggression is $.67 \times .36 = .24$. Controlling for stability effects, therefore, removes variance (i.e., .24) in T2 aggression that is shared with T1 frequency of competitive video game play and reduces the predictive effect of T1 frequency of competitive video game play on T2 aggression by almost 2/3, from bivariate $r = .37$ to $\beta = .13$. Therefore, controlling for stability in aggression dramatically reduces the magnitude of the effect size of another predictor of T2 aggression.

Next, we discuss two examples of longitudinal analyses by other researchers that also clearly illustrate our point. In these examples, we compare the bivariate r between the predictors at an earlier time point and the outcomes at a later time point to that of the β which controls for stability effects. First, Lee (2011) conducted a four-wave autoregressive cross-lagged path analysis to examine the longitudinal association between stress and internalizing problems among 3188 youth. Table 1 shows that the bivariate r s (not controlling for stability effects) between earlier predictors and later outcomes range from .26 to .33 (see path bivariate r). The stability effects for the outcomes ranged from $\beta = .38$ to .53 (see stability effect in Table 1), and the concurrent associations between the predictors and outcomes ranged from bivariate $r = .44$ to .49 (see concurrent bivariate r in Table 1), which suggests that controlling for stability effects will remove variance in the later outcome that is shared with the earlier predictor. A comparison of the bivariate r s to the β s between the earlier predictors and the later outcomes confirms that controlling for stability effects dramatically reduces the effect sizes (by as much as seven times), from bivariate r s ranging from .26 to .33, to β s ranging from .04 to .15 (see path β in Table 1).

Second, Martin and Liem (2010) conducted several two-wave autoregressive cross-lagged path analyses to examine the longitudinal associations between achieving academic personal bests and several measures of academic engagement and achievement, of which we chose four to illustrate our point (i.e., educational aspirations, homework completion, class participation, and literacy effort; see Table 1), among 1866 adolescents. Table 1 shows that the bivariate r s between the T1 predictors and the T2 outcomes for the first three paths (i.e., all paths except T1 academic personal bests \rightarrow T2 literacy effort) ranged from .31 to .33. The stability effects for these outcomes ranged from $\beta = .46$ to .53, and the concurrent associations between the predictors and outcomes ranged from bivariate $r = .47$ to .52, suggesting that controlling for stability effects will remove variance in the T2 outcome that is shared with the T1 predictor. Consistent with the prior two examples, Table 1 confirms that

TABLE 1
 Comparing effect size coefficients that control for stability effects to effect size coefficients that do not control for stability effects

Study	Path	Concurrent bivariate <i>r</i>	Stability effect	Path bivariate <i>r</i>	Path β
Lee (2011)	T1 internalizing problems → T2 stress	.44***	.45***	.26**	.07***
	T2 internalizing problems → T3 stress	.47***	.53***	.29**	.04*
	T3 internalizing problems → T4 stress	.49***	.48***	.31**	.08***
	T1 stress → T2 internalizing problems	.44***	.38***	.32**	.15***
	T2 stress → T3 internalizing problems	.47***	.41***	.33**	.14***
	T3 stress → T4 internalizing problems	.49***	.43***	.33**	.12***
Martin and Liem (2010)	T1 academic personal bests → T2 educational aspirations	.52***	.46***	.31***	.07**
	T1 academic personal bests → T2 homework completion	.48***	.53***	.33***	.08**
	T1 academic personal bests → T2 class participation	.47***	.53***	.31***	.06**
	T1 academic personal bests → T2 literacy effort	.06*	.16***	.05*	.04**

Notes: Cross-sectional bivariate *r* refers to the concurrent association between the predictor and the outcome. Stability effect refers to the β coefficient between the outcome at an earlier time point and the outcome at a later time point. Path bivariate *r* refers to the association between the predictor at an earlier time point and the outcome at a later time point without controlling for stability effects. Path β refers to the association between the predictor at an earlier time point and the outcome at a later time point, controlling for stability effects. *** $< .001$, ** $< .01$, * $< .05$.

controlling for stability effects dramatically reduces the effect sizes (by as much as five times) for the paths between the T1 predictors and the T2 outcomes, from the bivariate $r_s = .31$ to $.33$ to $\beta_s = .06$ to $.08$. In contrast, the stability coefficient for T2 literacy effort (see the last path in Table 1) of $\beta = .16$ is much smaller than the other stability coefficients in Table 1. In addition, the concurrent association between T1 academic personal bests and T1 literacy effort of bivariate $r = .06$ is much smaller than the other concurrent associations. Controlling for stability in literacy effort, therefore, removed very little variability ($.16 \times .06 = .01$) in T2 literacy effort that is associated with T1 academic personal bests and only reduced the predictive effect of T1 academic personal bests on T2 literacy effort from bivariate $r = .05$ to $\beta = .04$. Thus, in less common situations when the stability coefficient and the concurrent association between the T1 predictor and the T1 outcome are small, controlling for stability effects should have little impact on reducing the effect size of the path between the T1 predictor and the T2 outcome.

CONCLUSION AND IMPLICATIONS FOR RESEARCHERS

In the current paper, our goal was to highlight a gap in the literature regarding the interpretation of longitudinal effect sizes that control for stability effects and make a case for why very small effect sizes in longitudinal autoregressive models are common, yet still may be important. Citing examples from our own and others' research, we demonstrated that controlling for stability effects greatly reduces the magnitude of effect size coefficients between earlier predictors and later outcomes in longitudinal designs. In fact, we demonstrated that the bivariate r_s between an earlier predictor and a later outcome can be anywhere from *three* times, to *four to five* times (Martin & Liem, 2010), to *seven* times (Lee, 2011) larger than the β_s that control for stability effects. Therefore, we conclude that it may be misleading to apply Cohen's (1992) or Ferguson's (2009) guidelines for interpreting effect sizes when examining longitudinal effect sizes that control for stability effects.

Specifically, given that most behaviours show strong stability over time, often the amount of change in levels of the outcome will be small. In addition, controlling for stability effects often removes a large portion of variance in the outcome that is shared with the predictors. For example, given a bivariate longitudinal association of $r = .40$ between a predictor at T1 and an outcome at T2, if the stability effect of the outcome at T1 on the outcome at T2 is $r = .70$, and the overlap between the predictor at T1 and the outcome at T1 is $r = .40$, then controlling for stability will reduce the path between the predictor at T1 on the outcome at T2 from $r = .40$ to $\beta = .12$. Thus, 70% of the variability shared by the outcome at T2 and the predictor at T1 will be removed by controlling for stability. In contrast, in cross-sectional models where stability is not controlled, the effect size of $r = .40$ would be well above Cohen's (1992) cut-off for a

medium effect, and Ferguson's (2009) cut-off for a small effect. Therefore, even very small predictor effects (i.e., $\beta < .10$) in autoregressive models may be meaningful when there is strong stability in the outcome and at least moderate overlap between the predictor at T1 and the outcome at T1. For instance, competitive video game play may have a very small predictive effect on changes in levels of aggression over the span of one year. Over several years, however, these small effects may become additive, in that competitive video game play may have a significant cumulative impact on aggression. In addition, it also is important to interpret a longitudinal effect size coefficient in the context of prior studies (e.g., Lipsey et al., 2012; Peng et al., 2013). For example, if the effect sizes in cross-sectional studies on a particular topic tend to be small (e.g., $\beta = .20$), then we would expect even smaller effect sizes (e.g., $\beta < .10$) in longitudinal autoregressive models that control for stability effects.

Yet, Cohen's (1992) and Ferguson's (2009) guidelines may be appropriately applied to longitudinal effect size coefficients in certain situations. For example, Martin and Liem (2010) demonstrated that when the stability effects and the concurrent association between the T1 predictor and the T1 outcome were small, controlling for stability effects had very little impact in terms of reducing the effect size for the path between the T1 predictor and the T2 outcome. Thus, small stability coefficients suggests that there is a large amount of change in levels of the outcome over time, and small concurrent associations suggest that controlling for stability effects removes little variance in the outcome that is shared with the predictor. In the case of small stability coefficients and/or small concurrent associations, therefore, it may be more appropriate to consider a very small effect (i.e., $\beta < .10$) between a T1 predictor and a T2 outcome to be trivial, consistent with Cohen (1992) and Ferguson (2009). Thus, it is important to consider the implications of the magnitude of stability effects on the effect sizes of the predictors in autoregressive models in order to further clarify whether small effect sizes may be meaningful.

Because researchers often are primarily focused on predicting change in levels of the outcome, however, they rarely discuss the stability of the outcome and the impact of controlling for this stability on attenuating longitudinal predictive effects. For example, we searched for studies that were published in one of the top developmental journals, *Developmental Psychology*, over the last 10 years in which longitudinal autoregressive predictive models were employed, using the search terms "Developmental Psychology", "longitudinal OR prospective", and "time 1" in the PsycINFO database. Of the 38 studies that were retrieved, 34 of the studies did not discuss the impact of the stability of the outcome on the size of the predictive effects (see Figure 2). Only three studies discussed the stability effects in relation to the presence of absence of statistically significant predictive effects such as "... even after accounting for initial levels of aggression at age 3 and the strong stability in child aggression between age 3 and age 5, increases in spanking between ages 1 and 3 predicted increases in child aggression between

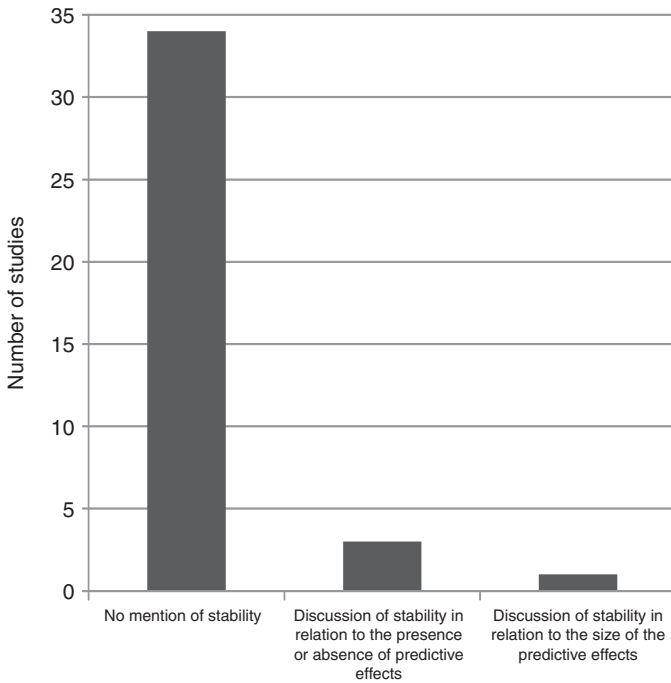


Figure 2. Studies published in *Developmental Psychology* in the past 10 years in which longitudinal autoregressive designs were employed and the researchers either did or did not discuss the stability of the outcome.

age 3 and age 5” (Lee, Altschul, & Gershoff, 2013, p. 9). Furthermore, only one study discussed the stability effect in relation to the *size* of the predictive effects such as “our results have to be weighted by an appropriate consideration of the stability of pathways among variables . . . Whereas it is not at all surprising that small effects have to be forcefully observed in stringent mediational models such as the present one, where variable stability dominates” (Caprara et al., 2013, p. 12).

Finally, it is important to consider how large a longitudinal effect size must be in order to be considered meaningful. We propose that this depends on the size of the stability coefficients and the concurrent associations between the predictor and the outcome (i.e., stronger stability effects and concurrent associations = smaller effect sizes for predictors), as well as the effect sizes in cross-sectional studies on the particular topic. For example, in a four-wave cross-lagged path analysis examining the bidirectional associations between competitive video game play and aggression among adolescents, we found significant predictive effects of earlier competitive video game play on later aggression that were as low as $\beta = .03$ (Adachi & Willoughby, 2013). However, because strong stability

in aggression implies that change in levels of aggression over time is small and controlling for this stability removed a large portion of variance in later aggression that was shared with earlier competitive video game play, we believe that these very small effects are not trivial, but in fact are meaningful. Specifically, our results suggest that competitive video game play predicts change in levels of aggression over time during the developmental period of adolescence, which demonstrates that this effect is not limited only to another period of development, such as childhood. Furthermore, these small long-term effects are consistent with meta-analytic studies regarding the link between video game play and aggression (e.g., Ferguson, [in press](#)). Moreover, predictive effects of competitive video game play on change in levels of aggression may reflect an ongoing process of cumulative effects, and thus may have a meaningful impact on aggression over time.

HOW SHOULD RESEARCHERS INTERPRET EFFECT SIZES IN AUTOREGRESSIVE MODELS?

We believe that recommending new guidelines for interpreting small, medium, and large effect sizes in autoregressive models may not be practical, as the degree to which small longitudinal effects are meaningful is largely dependent on the size of the bivariate correlations for these effects and the stability of the outcome. For example, a small longitudinal effect size (controlling for stability in the outcome) of $\beta = .07$ is more likely to be meaningful if the bivariate correlation for this effect is much larger (e.g., $r = .30$) and the stability effect is large (e.g., $\beta = .70$; controlling for this stability effect greatly attenuates the longitudinal effect size from $r = .30$ to $\beta = .07$), than if the bivariate correlation also is very small (e.g., $r = .09$) and the stability effect is small (e.g., $\beta = .20$; controlling for this stability effect has a minimal impact on reducing the longitudinal effect size from $r = .09$ to $\beta = .07$). Furthermore, given that longitudinal models vary greatly from study to study in terms of the number of waves examined (and hence whether lag 2 stability paths, lag 3 stability paths, etc., are included), and the number of predictors and covariates in the model, we believe that a concrete recommendation for what small, medium, and large effects should mean would be impractical. Interpreting effect sizes in autoregressive models, therefore, should involve a more dynamic method than simply citing universal guidelines for small, medium, and large effects. Specifically, we suggest the following for researchers attempting to interpret effect sizes in autoregressive models:

- Always report and discuss the stability effect for the outcome of interest;
- interpret the effect size of the predictor(s) on the outcome by putting it in perspective. What is the bivariate correlation for the predictive effect and how does the impact of controlling for stability attenuate that effect?

We conclude that this dynamic and individualized method of interpreting longitudinal effects in autoregressive models is more precise than simply referring to the current effect size guidelines and dismissing all very small effects (i.e., $\beta < .10$) as trivial, or proposing a more liberal cut-off for small effects in autoregressive models (e.g., small = .05) and then concluding that all very small effects (e.g., $\beta > .05$) are meaningful.

Manuscript received 9 April 2014

Revised manuscript accepted 4 September 2014

First published online 26 September 2014

REFERENCES

- Abelson, R. P. (1985). A variance explanation paradox: When a little is a lot. *Psychological Bulletin*, *97*, 129–133. doi:10.1037/0033-2909.97.1.129
- Adachi, P. J. C., & Willoughby, T. (2013). Demolishing the competition: The longitudinal link between competitive video games, competitive gambling, and aggression. *Journal of Youth and Adolescence*, *42*, 1090–1104. doi:10.1007/s10964-013-9952-2
- Baguley, T. (2004). Understanding statistical power in the context of applied research. *Applied Ergonomics*, *35*, 73–80. doi:10.1016/j.apergo.2004.01.002
- Baguley, T. (2009). Standardized or simple effect size: What should be reported? *British Journal of Psychology*, *100*, 603–617. doi:10.1348/000712608X377117
- Cairns, R. B., & Cairns, B. D. (1994). *Lifelines and risks: Pathways of youth in our time*. Cambridge: Cambridge University Press.
- Caprara, G. V., Tisak, M. S., Alessandri, G., Fontaine, R. G., Fida, R., & Paciello, M. (2013). The contribution of moral disengagement in mediating individual tendencies toward aggression and violence. *Developmental Psychology*. Advance online publication. doi:10.1037/a0034488
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159. doi:10.1037/0033-2909.112.1.155
- Cui, M., Donnellan, M. B., & Conger, R. D. (2007). Reciprocal influence between parents' marital problems and adolescent internalizing and externalizing behavior. *Developmental Psychology*, *43*, 1544–1552. doi:10.1037/0012-1649.43.6.1544
- Dodge, K. A., Coie, J. D., & Lynam, D. (2006). Aggression and antisocial behavior in youth. In N. Eisenberg, W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology, Vol. 3: Social, emotional, and personality development* (6th ed.) (pp. 719–788). Hoboken, NJ: Wiley.
- Ferguson, C. J. (in press). Do angry birds make for angry children? A meta-analysis of video game influences on children's and adolescents' aggression, mental health, prosocial behavior and academic performance. *Perspectives on Psychological Science*.
- Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, *40*, 532–538. doi:10.1037/a0015808
- Lee, J. (2011). Reciprocal influences between stress and internalizing problems in Korean adolescents: A cross-lagged, longitudinal study. *Asian Journal of Social Psychology*, *14*, 189–196. doi:10.1111/j.1467-839X.2011.01345.x
- Lee, S. J., Altschul, I., & Gershoff, E. T. (2013). Does warmth moderate longitudinal associations between maternal spanking and child aggression in early childhood? *Developmental Psychology*. doi:10.1037/a0031630

- Lenth, R. V. (2001). Some practical guidelines for effective sample size determination. *The American Statistician*, 55, 187–193. doi:10.1198/000313001317098149
- Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., Roberts, M., Anthony, K. S., & Busick, M. D. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms (NCSE 2013–3000)*. Washington, DC: National Center for Special Education Research, Institute of Education Sciences, US Department of Education.
- Marsh, H. W. (1993). Self-esteem stability and responses to the stability of self scale. *Journal of Research in Personality*, 27, 253–269. doi:10.1006/jrpe.1993.1018
- Martin, A. J., & Liem, G. A. D. (2010). Academic personal bests (PBs), engagement, and achievement: A cross-lagged panel analysis. *Learning and Individual Differences*, 20, 265–270. doi:10.1016/j.lindif.2010.01.001
- Olweus, D. (1979). Stability of aggressive reaction patterns in males: A review. *Psychological Bulletin*, 86, 852–875. doi:10.1037/0033-2909.86.4.852
- Ouellette, J. A., & Wood, W. (1998). Habit and intention in everyday life: The multiple processes by which past behavior predicts future behavior. *Psychological Bulletin*, 124, 54–74. doi:10.1037/0033-2909.124.1.54
- Peng, C. -Y., Chen, L. -T., Chiang, H. -M., & Chiang, Y. -C. (2013). The impact of APA and AERA guidelines on effect size reporting. *Educational Psychology Review*, 25, 157–209. doi:10.1007/s10648-013-9218-2
- Ployhart, R. E., & Ward, A. -K. (2011). The “quick start guide” for conducting and publishing longitudinal research. *Journal of Business and Psychology*, 26, 413–422. doi:10.1007/s10869-011-9209-6
- Rosenthal, R., & Rubin, D. B. (1979). A note on percent variance explained as a measure of the importance of effects. *Journal of Applied Social Psychology*, 9, 395–396. doi:10.1111/j.1559-1816.1979.tb02713
- Taris, T. W. (Ed.). (2000). 4 issues in discrete-time panel analysis. *A primer in longitudinal data analysis* (pp. 54–75). London: Sage. doi:10.4135/9781849208512.n4
- Thorndike, R. L. (1940). “Constancy” of the IQ. *Psychological Bulletin*, 37, 167–186. doi:10.1037/h0061268
- Vaillancourt, T., Brendgen, M., Boivin, M., & Tremblay, R. E. (2003). A longitudinal confirmatory factor analysis of indirect and physical aggression: Evidence of two factors over time? *Child Development*, 74, 1628–1638. doi:10.1046/j.1467-8624.2003.00628.x
- Wilkinson, L. & Task Force on Statistical Inference (1999). Statistical methods in psychological journals: Guidelines and explanations. *American Psychologist*, 54, 594–604. doi:10.1037/0003-066X.54.8.594
- Willoughby, T., Adachi, P. J. C., & Good, M. (2011). A longitudinal study of the association between violent video game play and aggression among adolescents. *Developmental Psychology*, 48, 1044–1057. doi:10.1037/a0026046