



# A practical guide to calculating Cohen's $f^2$ , a measure of local effect size, from PROC MIXED

Arielle S. Selya<sup>1\*</sup>, Jennifer S. Rose<sup>1</sup>, Lisa C. Dierker<sup>1</sup>, Donald Hedeker<sup>2</sup> and Robin J. Mermelstein<sup>2</sup>

<sup>1</sup> Psychology Department, Wesleyan University, Middletown, CT, USA

<sup>2</sup> Institute for Health Research and Policy, University of Illinois at Chicago, Chicago, IL, USA

## Edited by:

Jeremy Miles, Research and Development Corporation, USA

## Reviewed by:

Anne C. Black, Yale University School of Medicine, USA

Daniel Saverio John Costa, University of Sydney, Australia

## \*Correspondence:

Arielle S. Selya, Psychology Department, Wesleyan University, 207 High Street, Middletown, CT 06459, USA.

e-mail: aselya@wesleyan.edu

Reporting effect sizes in scientific articles is increasingly widespread and encouraged by journals; however, choosing an effect size for analyses such as mixed-effects regression modeling and hierarchical linear modeling can be difficult. One relatively uncommon, but very informative, standardized measure of effect size is Cohen's  $f^2$ , which allows an evaluation of local effect size, i.e., one variable's effect size within the context of a multivariate regression model. Unfortunately, this measure is often not readily accessible from commonly used software for repeated-measures or hierarchical data analysis. In this guide, we illustrate how to extract Cohen's  $f^2$  for two variables within a mixed-effects regression model using PROC MIXED in SAS<sup>®</sup> software. Two examples of calculating Cohen's  $f^2$  for different research questions are shown, using data from a longitudinal cohort study of smoking development in adolescents. This tutorial is designed to facilitate the calculation and reporting of effect sizes for single variables within mixed-effects multiple regression models, and is relevant for analyses of repeated-measures or hierarchical/multilevel data that are common in experimental psychology, observational research, and clinical or intervention studies.

**Keywords:** Cohen's  $f^2$ , effect size, hierarchical linear modeling, mixed-effects regression, PROC MIXED,  $R^2$ , SAS

## INTRODUCTION

Effect sizes are an important complement to null hypothesis significance testing (e.g.,  $p$ -values), in that they offer a measure of practical significance in terms of the magnitude of the effect, and are independent of sample size. As an additional benefit, dimensionless, or standardized measures of effect size allow direct comparison of two or more quantities, for example variables measured on different scales or independent studies in a meta-analysis. The statistical community has long encouraged researchers to report effect sizes (Wilkinson and Task Force on Statistical Inference, 1999; Kline, 2004; Nakagawa and Cuthill, 2007), and scientific journals are increasingly requesting or requiring authors to report them along with  $p$ -values (e.g., Snyder, 2000; Huberty, 2002; Fidler et al., 2005).

One practical difficulty for researchers who wish to include effect size in their results is the large number of potential measures available. Kirk (1996) reported 40 different effect size measures, of which several may be appropriate for any given data structure. For example, four effect size measures exist for dichotomous outcomes (Pace, 2011), three for continuous outcomes across groups (Huberty, 2002), and still others for multilevel data, for which there is often no consensus on which is most appropriate (Peugh, 2010). Three commonly used types of effect size are suitable for the majority of relatively simple analyses, and fall into the  $r$  family (measures of correlation between continuous variables), the  $d$  family (standardized mean differences in a continuous dependent variable across levels of a categorical independent variable), or ratio statistics (measures of comparative risk for

dichotomous outcomes; Rosenthal, 1994; Nakagawa and Cuthill, 2007).

However, these frequently used measures of effect size are inappropriate for more complex data structures such as hierarchical data, which may consist of observations nested within subjects (e.g., repeated observations of the same participants) or observations nested within groups (e.g., observational studies of students within different classrooms). Additionally, many effect sizes are unable to address research questions which may involve competing effect sizes of different variables within the same multivariate model, rather than a variable's effect size for the overall model; this distinction has been described as local vs. global effect sizes (Peugh, 2010). A researcher developing an intervention program with limited resources, for example, may wish to know which of many simultaneous risk factors has the largest effect on the outcome to determine which intervention strategy may be most effective.

Once a suitable effect size measure has been chosen, researchers unfamiliar with effect sizes often face practical challenges in obtaining or calculating their values. Effect sizes are often not directly provided by statistical software output of multivariate regression analyses. In some cases, even statistics which could be used to calculate effect sizes (e.g.,  $R^2$ , the proportion of variance accounted for, in a multiple regression) are not readily accessible from the output. Additional caution is needed when calculating effect sizes using hierarchical or repeated-measures data, as researchers must account for variance accounted for by different sources, for example fixed effects vs. random effects in a mixed-effects regression model.

This article first explains the choice of Cohen's  $f^2$  as a useful measure of local effect size appropriate for hierarchical and repeated-measures data. Next, a step-by-step guide is provided for calculating Cohen's  $f^2$  from PROC MIXED in SAS software, a popular and powerful tool for mixed-effects regression models and hierarchical linear models. Finally, using data from a longitudinal cohort study of adolescent smoking to compare local effect sizes of separate variables, two examples are given in which Cohen's  $f^2$  is calculated to answer variants of this research question.

## METHODS

### DATA AND RESEARCH QUESTION

Data were drawn from the Social-Emotional Contexts of Adolescent Smoking Patterns (SECASP) Study which has been described elsewhere (e.g., Dierker and Mermelstein, 2010). Briefly, sampling was conducted within high schools in the greater Chicago metropolitan area, and was designed to obtain a racially and ethnically diverse sample of adolescents who were at the earliest stages of smoking exposure. The final sample included 1,263 students across 16 high schools, and the current focus is on experimental and light smokers ( $N = 726$  at baseline who had smoked in the past 90 days, but not more than 5 cigarettes/day and on a daily basis). Follow-up assessments were conducted at 6, 15, and 24 months following baseline. Though the full study also contains other smoking status groups and later assessment waves, for simplicity, the demonstrations presented below will be restricted to these observations.

Measures of interest include smoking frequency (number of days out of the past 30 on which the adolescent smoked), smoking quantity (number of cigarettes smoked in the past 7 days), nicotine dependence syndrome scale (NDSS; mean of 10 nicotine dependence items assessed on a four-point Likert-type scale) modified for use in adolescents (Sterling et al., 2009), and gender. Smoking frequency, smoking quantity, and NDSS score were assessed at each follow-up assessment. Additionally, a categorical variable indicating the assessment (baseline, 6, 15, or 24 month follow-up) is included in the repeated-measures analyses.

The research questions involve the independent effects of smoking quantity and nicotine dependence on the dependent variable of smoking frequency, while controlling for gender. The example analyses below will address whether smoking quantity or nicotine dependence (as measured by the NDSS), relative to each other, has a greater association with smoking frequency (1) within a single assessment and (2) in general across each of the four assessments.

### CHOOSING A MEASURE OF EFFECT SIZE

As the dependent variable is continuous, an effect size in the  $d$  family of standardized mean differences might be considered first. This effect size measure has been used elsewhere in the context of multivariate mixed-effects regression models using repeated measures of subjects, for example mean differences in an outcome across groups (Friedmann et al., 2008); however, it is inadequate for the current research question for the following reasons. First, it is a comparison of groups and thus requires the independent variable of interest to be categorical. While effect sizes of different assessment waves may be of tangential interest to the research topic,

the primary question relates to the continuous variables of smoking quantity and nicotine dependence. Second, standardized mean differences cannot determine local effect sizes, that is, individual effect sizes of particular variables within a multivariate model that includes other categorical and continuous independent variables.

Cohen's  $f^2$  (Cohen, 1988) is appropriate for calculating the effect size within a multiple regression model in which the independent variable of interest and the dependent variable are both continuous. Cohen's  $f^2$  is commonly presented in a form appropriate for global effect size:

$$f^2 = \frac{R^2}{1 - R^2}. \quad (1)$$

However, the variation of Cohen's  $f^2$  measuring local effect size is much more relevant to the research question:

$$f^2 = \frac{R_{AB}^2 - R_A^2}{1 - R_{AB}^2} \quad (2)$$

where  $B$  is the variable of interest (i.e., either smoking quantity or nicotine dependence score),  $A$  is the set of all other variables (i.e., gender and depending on what  $B$  is at the moment, nicotine dependence score or smoking quantity),  $R_{AB}^2$  is the proportion of variance accounted for by  $A$  and  $B$  together (relative to a model with no regressors), and  $R_A^2$  is the proportion of variance accounted for by  $A$  (relative to a model with no regressors). Thus, the numerator of (2) reflects the proportion of variance uniquely accounted for by  $B$ , over and above that of all other variables (Cohen, 1988).

### EXTRACTING COHEN'S $f^2$ FROM PROC MIXED

The PROC MIXED procedure in SAS software is used in this guide, as it is a powerful and widely used tool for running mixed-effects regression models and hierarchical linear models. It is important to note that PROC MIXED assumes the dependent variable to be continuous and approximately normally distributed; thus, the calculations described in the examples below are for continuous dependent variables. As the calculation for Cohen's  $f^2$  is based on  $R^2$  values of different versions of the full regression model, it is necessary to extract  $R^2$  from the output of PROC MIXED in SAS. Though  $R^2$  is not output directly, it is the proportion of variance accounted for by the regressors in the model and can thus be calculated using the residual variance of the full model ( $V_{\text{full}}$ ) and the residual variance of the model with no regressors ( $V_{\text{null}}$ ):

$$R^2 = \frac{V_{\text{null}} - V_{\text{full}}}{V_{\text{null}}} \quad (3)$$

Calculating Cohen's  $f^2$  in Eq. 2 requires both  $R_{AB}^2$  and  $R_A^2$ , thus several variants of the regression must be run in order to obtain  $V_{AB}$  (residual variance of the model containing  $A$  and  $B$  together),  $V_A$  (residual variance of the model without  $B$ ) for each  $B$ , and  $V_{\text{null}}$  (residual variance of the model without regressors). Once these residual variances are obtained,  $R_{AB}^2$  and  $R_A^2$  can be calculated as in Eq. 3, and these in turn can be used to calculate Cohen's  $f^2$  as in Eq. 2.

A further consideration when using a mixed-effects regression model or hierarchical linear model is that the reduction of variance between  $V_{\text{full}}$  and  $V_{\text{null}}$  must be assured to be from only the fixed effects, and not from the random effects. It is possible that, after removing regressors, the random effects account for more than their previous share of the now-increased total residual variance. To accurately assess the reduction in variance due to the fixed effects of the variable of interest, it is therefore necessary to hold constant the variance accounted for by random effects when running the  $A$ -only or null models. For example, in a mixed-effects regression model that involves fixed effects of variables  $A$  and  $B$ , and random intercepts across subjects,  $A$ -only and null models must have the same random subject effects as the full  $AB$  model.

## DEMONSTRATIONS

### EXAMPLE 1: EFFECT SIZES OF SMOKING QUANTITY AND NICOTINE DEPENDENCE WITHIN ASSESSMENT WAVES

The main research question is whether smoking quantity and nicotine dependence have time-varying effects on smoking frequency. The motivation for obtaining effect sizes is to supplement results from a previously run mixed-effects regression model investigating these variables' interactions with time (among other terms). Random effects included slope and intercept which were allowed to vary by subject. The analysis showed significant time-varying effects; that is, the strengths of associations that smoking quantity and nicotine dependence score have with smoking frequency change over time. For this reason, obtaining local effect sizes for (1) smoking quantity and (2) the nicotine dependence score *within each assessment wave* is helpful to better characterize exactly how their relationships with smoking frequency change over time as well as how they compare to each other. Since evaluating effect sizes within each assessment wave no longer involves repeated measures, there are no random effects to fix and thus the reduction in variance is purely due to the fixed effects.

First,  $R_{AB}^2$  is calculated for smoking quantity at the baseline wave. This involves finding  $V_{AB}$  and  $V_{\text{null}}$ . To find  $V_{AB}$ , a multiple regression model is run at the baseline wave:

```
if time = 0;
PROC MIXED data=new method=ml;
class id time gender;
model SmkFreq = SmkQuant NDSS gender /solution cl;
run;
```

$V_{AB}$  is printed in the output and is shown in **Table 1**; the value is 18.3555.

By modifying the *model* statement, values can be obtained for  $V_{\text{null}}$  as well as the variance of the model excluding smoking quantity (denoted as  $V_{-SQ}$  rather than  $V_A$ , which in this example can refer to excluding smoking quantity or excluding the nicotine dependence score).

**Table 1 |  $V_{AB}$  at baseline in example 1.**

Covariance parameter estimates	
Cov parm	Estimate
Residual	18.3555

Calculating  $R^2$  and  $f^2$  values from the residual variance estimates can be automated by using the SAS output delivery system (ODS) to store the desired parameters in separate data sets, and subsequently using a merged dataset of variances to calculate  $R^2$  and  $f^2$  using Eqs 3 and 2, respectively. The following syntax calculating  $R^2$  and  $f^2$  for smoking quantity illustrates this:

```
if time = 0;
* Obtain  $V_{AB}$ 
ods output CovParms = VABex1SQ;
PROC MIXED data=new method=ml;
class id time gender;
model smkFreq = smkQuant NDSS gender /solution cl;
run;
quit;
ods output close;
```

```
* Obtain  $V_{-SQ}$ ;
ods output CovParms = VAex1SQ;
PROC MIXED data=new method=ml;
class id time gender;
model smkFreq = NDSS gender /solution cl;
run;
quit;
ods output close;
```

```
* Obtain  $V_{\text{null}}$ ;
ods output CovParms = VnullEx1SQ;
PROC MIXED data=new method=ml;
class id time gender;
model smkFreq = /solution cl;
run;
quit;
ods output close;
```

```
* Merge datasets for  $V_{AB}$ ,  $V_{-SQ}$ , and  $V_{\text{null}}$  into one dataset, while
renaming the variable name "Estimate" with unique names;
DATA VallEx1SQ;
merge VABex1SQ(rename=(Estimate=VAB))
VAex1SQ(rename=(Estimate=VA)) VnullEx1SQ
(rename=(Estimate=Vnull)); by CovParm;
DROP CovParm;
run;
```

```
* Calculate  $R_{AB}^2$ ,  $R_{-SQ}^2$ , and  $f_{SQ}^2$  from the merged dataset and
save to a new dataset;
DATA resultsEx1SQ; set VallEx1SQ;
DROP VAB VA Vnull;
R2AB = (Vnull - VAB)/Vnull;
R2A = (Vnull - VA)/Vnull;
f2 = (R2AB - R2A)/(1 - R2AB);
run;
```

```
* Print the results to the Results window
PROC PRINT data=resultsEx1SQ;
run;
```

This procedure reveals that at the baseline wave, Cohen's local  $f^2$  for smoking quantity is 0.69 (**Table 2**).

Calculating  $R^2_{-ND}$  and Cohen's  $f^2$  for nicotine dependence simply requires appropriate modifications to dataset names and to the *model* statement in the syntax above. Since  $V_{AB}$  and  $V_{null}$  are the same regardless of the variable of interest, only the *model* statement for  $V_A$  must be modified to hold out nicotine dependence from the model rather than smoking quantity:

```
* Obtain  $V_{-ND}$ ;
ods output CovParms = VAex1ND;
PROC MIXED data=new method=ml;
class id time gender;
model smkFreq = smkQuant gender /solution cl;
run;
quit;
ods output close;
Cohen's local  $f^2$  for nicotine dependence is 0.32 (Table 3).
```

Repeating this process for the remaining assessment waves yields the local effect sizes for smoking quantity and nicotine dependence over time (Table 4). This allows a direct comparison of the two regressors, revealing that at the baseline wave, smoking quantity has a stronger relationship with smoking frequency than does nicotine dependence, while two years later, nicotine dependence has a stronger relationship with smoking frequency than does smoking quantity. These values signify approximately moderate ( $f^2 \geq 0.15$ ) to large ( $f^2 \geq 0.35$ ) effect sizes according to Cohen's (1988) guidelines.

As a cautionary note, stratifying models by assessment wave is appropriate for calculating effect sizes of the original mixed-effects regression model only because the original model (1) used time

as a categorical variable (due to the irregular intervals between assessment waves) and (2) allowed smoking quantity and nicotine dependence to vary over time (i.e., included interactions between these variables and time). Had time been continuous or the regressors had not been allowed to vary over time, the effect sizes in Table 4 would be less consistent with the original model, possibly resulting in inaccurate estimates of effect size.

**EXAMPLE 2: EFFECT SIZES OF SMOKING QUANTITY AND NICOTINE DEPENDENCE ACROSS TIME**

Having obtained local effect sizes *within* each assessment wave, the current question is whether smoking quantity or nicotine dependence has the greater effect *across* assessments. To calculate this, random effects must be held constant between  $R^2_{AB}$  and  $R^2_A$  so that the change in variance after adding either smoking quantity or the nicotine dependence score reflects only a change in the variance due to fixed effects. This example thus extends the procedure described in example 1 to include random effects and repeated measures in the calculation of Cohen's  $f^2$ .

First, the full AB model must be run in order to obtain the values of random effect variance at which subsequent models must be fixed. The covariance parameters in this example include both random effect variances (due to the inclusion of random slopes and intercepts across subjects) and residual variance; these parameters can be saved into a new dataset using ODS (Table 5).

Calculation of  $R^2_{AB}$  and  $R^2_A$  is based on the residual variance ( $V_{AB} = 20.2417$ ); however, the random effect variance must be held at the values shown in Table 5 when running the A-only and null models in order to accurately measure the reduction of residual variance. Using the *parms* statement, the covariance dataset from the full AB model can be read into in the  $V_A$  and  $V_{null}$ , in order to fix the random effect variance. The following syntax runs the mixed models, holding the random effects constant, merges the residual variance from each model, and calculates  $R^2_{AB}$ ,  $R^2_{SQ}$ , and Cohen's  $f^2$  for smoking quantity across all assessment waves.

```
PROC SORT data=new;
by id;

* Obtain  $V_{AB}$ ;
ods output CovParms = VABex2SQ;
PROC MIXED data=new method=ml;
class id time gender;
model smkFreq = smkQuant NDSS gender time smkQuant*time
NDSS*time gender*time /solution cl;
random int wave /SUBJECT=id type=un;
run;
quit;
ods output close;
```

**Table 2 |  $R^2_{AB}$ ,  $R^2_{SQ}$ , and Cohen's  $f^2$  values for smoking quantity at baseline in example 1.**

Obs	$R^2_{AB}$	$R^2_A$	$f^2$
1	0.74650	0.57232	0.68711

**Table 3 |  $R^2_{AB}$ ,  $R^2_{-ND}$ , and Cohen's  $f^2$  values for nicotine dependence at baseline in example 1.**

Obs	$R^2_{AB}$	$R^2_A$	$f^2$
1	0.74650	0.66663	0.31507

**Table 4 | Effect sizes for smoking quantity and nicotine dependence within assessment waves.**

	Cohen's $f^2$ for smoking quantity	Cohen's $f^2$ for nicotine dependence
Baseline wave	0.69	0.32
6-Month wave	0.48	0.46
15-Month wave	0.15	0.80
24-Month wave	0.15	0.70

Cohen's  $f^2$  for local effect sizes of smoking quantity and nicotine dependence within a multiple regression performed within each assessment wave are shown. According to Cohen's (1988) guidelines,  $f^2 \geq 0.02$ ,  $f^2 \geq 0.15$ , and  $f^2 \geq 0.35$  represent small, medium, and large effect sizes, respectively.

**Table 5 | Covariance dataset for the full AB model in example 2.**

Cov parm	Subject	Estimate
UN(1, 1)	id	2.3419
UN(2, 1)	id	1.8809
UN(2, 2)	id	0.8706
Residual		20.2417

```

* Obtain  $V_{SQ}$ ;
ods output CovParms = VAex2SQ;
PROC MIXED data=new method=ml;
class id time gender;
model smkFreq = NDSS gender time NDSS*time gender*time
  /solution cl;
random int wave /SUBJECT=id type=un;
parms /parmsdata=VABex2SQ hold=1,2,3;
run;
quit;
ods output close;

* Obtain  $V_{null}$ ;
ods output CovParms = VnullEx2SQ;
PROC MIXED data=new method=ml;
class id time gender;
model smkFreq = /solution cl;
random int wave /SUBJECT=id type=un;
parms /parmsdata=VABex2SQ hold=1,2,3;
run;
quit;
ods output close;

* Select the residual variance and discard the random-effects
  variances;
DATA VABex2SQ;
set VABex2SQ;
if CovParm = "Residual";
run;
DATA VAex2SQ;
set VAex2SQ;
if CovParm = 'Residual';
run;
DATA VnullEx2SQ;
set VnullEx2SQ;
if CovParm = 'Residual';
run;

* Merge datasets for  $V_{AB}$ ,  $V_{SQ}$ , and  $V_{null}$  into one dataset, while
  renaming the variable name "Estimate" with unique names;
DATA VallEx2SQ;
merge VABex2SQ(rename=(Estimate=VAB)) VAex2SQ(rename
  =(Estimate=VA)) VnullEx2SQ(rename=(Estimate=Vnull));
  by CovParm;
DROP CovParm Subject;
run;

* Calculate  $R_{AB}^2$ ,  $R_{SQ}^2$  and  $f_{SQ}^2$  from the merged dataset and
  save to a new dataset;
DATA resultsExSQ; set VallEx2SQ;
DROP VAB VA Vnull;
R2AB = (Vnull - VAB)/Vnull;
R2A = (Vnull - VA)/Vnull;
f2 = (R2AB - R2A)/(1 - R2AB);
run;

* Print the results to the Results window

```

```
PROC PRINT data=resultsExSQ;
```

```
run;
```

Finally,  $R_{AB}^2$ ,  $R_{ND}^2$ , and Cohen's local  $f^2$  for nicotine dependence can be obtained by modifying the A-only model above to withhold nicotine dependence rather than smoking quantity from the model as shown below; aside from file names, no other modifications are necessary as the full AB model and the null models are the same for smoking quantity and nicotine dependence.

```
* Obtain  $V_{ND}$ ;
```

```
ods output CovParms = VAex2ND;
```

```
PROC MIXED data=new method=ml;
```

```
class id time gender;
```

```
model smkFreq = smkQuant gender time smkQuant*time
  gender*time /solution cl;
```

```
random int wave /SUBJECT=id type=un;
```

```
parms /parmsdata=VABex2ND hold=1,2,3;
```

```
run;
```

```
quit;
```

```
ods output close;
```

These results show that, across assessment waves, nicotine dependence score ( $f^2 = 0.44$ ) has a slightly greater impact on smoking frequency than does smoking quantity ( $f^2 = 0.37$ ), though both have large effect sizes.

## CONCLUSION

In this guide, Cohen's  $f^2$  was chosen as an appropriate measure of local effect size for variables within a multivariate, mixed-effects regression model. However, it is important to note caveats to this approach. First, some statisticians have objected to standardized effect size, as it can be reduced by variables with low reliability/high variance, and is unstable across studies with different designs, such that repeated-measures studies typically yield stronger effect sizes compared to an equivalent independent design (Baguley, 2009). Additionally, standardized effect size can be distorted by the sampling procedure (e.g., sampling from a truncated distribution, sampling from the tails of the distribution, or restricting to a subset of the population; Baguley, 2009), and thus generalizing standardized effect sizes should be done with caution.

A second caveat is that effect size measures other than Cohen's  $f^2$  may be appropriate for similar analyses. Previous literature has suggested as a local effect size the proportional reduction in variance (PRV) after the variable of interest has been added to a mixed-effects model or hierarchical linear model, compared to an A-only model (Peugh, 2010). PRV is related to Cohen's  $f^2$  through  $R^2$ -type measures, but has disadvantages for these particular current research questions. Specifically, PRV does not incorporate the total variance of the model prior to adding regressors, thus a given value of PRV may indicate different effect sizes relative to the overall model, depending on the variance of the intercept-only model. As a result, Cohen's  $f^2$  is standardized with respect to the total variance, allowing more accurate comparison of effect sizes between variables within the model as a whole.

Despite the widespread use of repeated-measures data and hierarchical or multilevel data in experimental, observational, and clinical research, reporting effect sizes presents challenges to researchers unfamiliar with this topic. These challenges can

start with the initial process of choosing which measure out of many is most appropriate for the data, the analysis, and the research question. Though effect size measures appropriate for repeated-measures or multilevel data have already been proposed (Cohen, 1988; Nakagawa and Cuthill, 2007; Peugh, 2010) and can be directly or indirectly obtained through common statistical software packages, existing sources rarely provide comprehensive information about these processes that are accessible to naïve researchers. This article presents a practical guide to calculating Cohen's  $f^2$ , an effect size measure for a single variable within a multivariate, mixed-effects regression model, from the output of PROC MIXED in SAS software. This guide is intended for

researchers in a wide variety of scientific fields who are unfamiliar with effect sizes, and aims to facilitate effect size reporting in analyses that are commonly performed, yet too complex for standard measures of effect size.

## REFERENCES

- Baguley, T. (2009). Standardized or simple effect size: what should be reported? *Br. J. Psychol.* 100, 603–617.
- Cohen, J. E. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Dierker, L., and Mermelstein, R. (2010). Early emerging nicotine-dependence symptoms: a signal of propensity for chronic smoking behavior in adolescents. *J. Pediatr.* 156, 818–822.
- Fidler, F., Cumming, G., Thomason, N., Pannuzzo, D., Smith, J., Fyffe, P., Edmonds, H., Harrington, C., and Schmitt, R. (2005). Toward improved statistical reporting in the journal of consulting and clinical psychology. *J. Consult. Clin. Psychol.* 73, 136–143.
- Friedmann, P. D., Rose, J. S., Swift, R., Stout, R. L., Millman, R. P., and Stein, M. D. (2008). Trazodone for sleep disturbance after alcohol detoxification: a double-blind, placebo-controlled trial. *Alcohol. Clin. Exp. Res.* 32, 1652–1660.
- Huberty, C. J. (2002). A history of effect size indices. *Educ. Psychol. Meas.* 62, 227–240.
- Kirk, R. E. (1996). Practical significance: a concept whose time has come. *Educ. Psychol. Meas.* 56, 746–759.
- Kline, R. B. (2004). *Beyond Significance Testing*. Washington: American Psychological Association.
- Nakagawa, S., and Cuthill, I. C. (2007). Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biol. Rev. Camb. Philos. Soc.* 82, 591–605.
- Pace, N. L. (2011). Research methods for meta-analyses. *Best Pract. Res. Clin. Anaesthesiol.* 25, 523–533.
- Peugh, J. L. (2010). A practical guide to multilevel modeling. *J. Sch. Psychol.* 48, 85–112.
- Rosenthal, R. (1994). "Parametric measures of effect size," in *The Handbook of Research Synthesis*, eds. H. Cooper and L. V. Hedges (New York: Russel Sage Foundation), 231–245.
- Snyder, P. (2000). Guidelines for reporting results of group quantitative investigations. *J. Early Interv.* 23, 145–150.
- Sterling, K. L., Mermelstein, R., Turner, L., Diviak, K., Flay, B., and Shiffman, S. (2009). Examining the psychometric properties and predictive validity of a youth-specific version of the nicotine dependence syndrome scale (NDSS) among teens with varying levels of smoking. *Addict. Behav.* 34, 616–619.
- Wilkinson, L., and Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: guidelines and explanations. *Am. Psychol.* 54, 594–604.
- could be construed as a potential conflict of interest.
- Received: 22 February 2012; paper pending published: 14 March 2012; accepted: 27 March 2012; published online: 17 April 2012.
- Citation: Selya AS, Rose JS, Dierker LC, Hedeker D and Mermelstein RJ (2012) A practical guide to calculating Cohen's  $f^2$ , a measure of local effect size, from PROC MIXED. *Front. Psychology* 3:111. doi: 10.3389/fpsyg.2012.00111
- This article was submitted to *Frontiers in Quantitative Psychology and Measurement*, a specialty of *Frontiers in Psychology*.
- Copyright © 2012 Selya, Rose, Dierker, Hedeker and Mermelstein. This is an open-access article distributed under the terms of the Creative Commons Attribution Non Commercial License, which permits non-commercial use, distribution, and reproduction in other forums, provided the original authors and source are credited.