Journal of Modern Applied Statistical Methods

Volume 4 | Issue 1 Article 11

5-1-2005

Bias Affiliated With Two Variants Of Cohen's d When Determining U1 As A Measure Of The Percent Of Non-Overlap

David A. Walker
Northern Illinois University, dawalker@niu.edu

Journal of Modern Applied Statistical Methods
May, 2005, Vol. 4, No.1, 100-105

1538 – 9472/05/$95.00

Bias Affiliated With Two Variants Of Cohen's *d* When Determining U_1 As A Measure Of The Percent Of Non-Overlap

David A. Walker
Educational Research and Assessment Department
Northern Illinois University

Variants of Cohen's d, in this instance d_t and d_{adj}, has the largest influence on U_1 measures used with smaller sample sizes, specifically when n_1 and $n_2 = 10$. This study indicated that bias for variants of *d*, which influence U_1 measures, tends to subside and become more manageable, in terms of precision of estimation, around 1% to 2% when n_1 and $n_2 = 20$. Thus, depending on the direction of the influence, both d_t and d_{adj} are likely to manage bias in the U_1 measure quite well for smaller to moderate sample sizes.

Key words: Non-overlap, effect size, Cohen's *d*

Introduction

In his seminal work on power analysis, Jacob Cohen (1969; 1988) derived an effect size measure, Cohen's *d*, as the difference between two sample means. Using n, M, and SD from two sample groups, *d* provided "score distances in units of variability" (p. 21), by translating the means into a common metric of standard deviation units pertaining to the degree of departure from the null hypothesis.

The common formula for Cohen's *d* (1988) is

$$d = \frac{\bar{X}_1 - \bar{X}_2}{\hat{\sigma}_{pooled}} \qquad (1)$$

where $\bar{X}_1$ and $\bar{X}_2$ are sample means and

$$\hat{\sigma}_{pooled} = \frac{(n_1-1)(sd_1)^2 + (n_2-1)(sd_2)^2}{(n_1+n_2-2)}$$

David Walker is an Assistant Professor at Northern Illinois University. His research interests include structural equation modeling, effect sizes, factor analyses, predictive discriminant analysis, predictive validity, weighting, and bootstrapping. Email: dawalker@niu.edu.

Cohen's *d* can be calculated if no n, M, or SD for two groups is reported via *t* values and degrees of freedom, termed d_t here, where it is assumed that n_1 and n_2 are equal (Rosenthal, 1991):

$$d_1 = \frac{2t}{\sqrt{df}} \qquad (2)$$

where $t = t$ value, and df = $n_1 + n_2 - 2$

Kraemer (1983) noted that the distribution of Cohen's *d* was skewed and heavy tailed, and Hedges (1981) found that *d* was a positively biased effect size estimate. Hedges proposed an approximate, modified estimator of *d*, which will be termed d_{adj} here, where:

$$c(m) \approx 1 - \frac{3}{4m-1} \qquad (3)$$

where $m = n_1 + n_2 - 2$.

Cohen (1969; 1988) revisited the idea of group overlap, which was studied by Tilton (1937), and the degree of overlap (O) between two distributions; and also in close proximity to the time of Cohen's initial work (i.e., 1969) by Elster and Dunnete (1971). This resulted in the U_1 measure, which was derived from *d* as a percent of non-overlap. As Cohen (1988) explained, "If we maintain the assumption that the populations being compared are normal and with equal variability, and conceive them further as equally numerous, it is possible to define measures of non-overlap (U_1) associated with *d*"

(p. 21).

Algebraically, U_1 is related to the cumulative normal distribution and is expressed as (Cohen, 1988):

$$U_1 = \frac{2P_{d/2} - 1}{P_{d/2}} = \frac{2U_2 - 1}{U_2} \quad (4)$$

where d = Cohen's *d* value, P = percentage of the area falling below a given normal deviate, and $U_2 = P_{d/2.}$

In SPSS (Statistical Package for the Social Sciences) syntax, U_1 is calculated using the following expressions:

```
Compute U = CDF.NORMAL((ABS(d)/2),0,1).
Compute U1 = (2*U-1)/U*100.
Execute.
```

where d = Cohen's d value, ABS = absolute value, CDF. NORMAL = cumulative probability that a value from a normal distribution where M = 0 and SD = 1 is < the absolute value of d/2.

Thus, the link between *d* and U_1 was seen by Cohen (1988) in that, "d is taken as a deviate in the unit normal curve and P [from expression 4] as the percentage of the area (population of cases) falling below a given normal deviate" (p. 23).

For Cohen (1998), non-overlap was the extent to which an experiment or intervention had had an effect of separating the two populations of interest. A high percentage of non-overlap indicated that the two populations were separated greatly. When *d* = 0, there was 0% overlap and $U_1 = 0$ also, or as Cohen (1988) noted "either population distribution is perfectly superimposed on the other" (p. 21). Therefore, the two populations were identical.

The assumptions for the percentage of population non-overlap are: 1) the comparison populations have normality and 2) equal variability. Further, Cohen (1988) added that the U_1 measure would also hold for samples from two groups if "the samples approach the conditions of normal distribution, equal variability, and equal sample size" (p. 68).

Cohen (1988, p. 22) went on to produce Table 2.2.1, which consisted of non-overlap percentages for values of *d*. Assuming a normal distribution, this table showed that, for example, a value of *d* = .20 would have a corresponding U_1 = 14.7%, or a percentage of non-overlap of just over 14%. That is, the distribution of scores for the treatment group overlapped only a small amount with the distribution of scores for the non-treatment group, which was manifested in the small effect size of .20. As the value of *d* increased, so would the percentage of non-overlap between the two distributions of scores, which indicated that the two groups differed considerably.

Methodology

After an extensive review of the literature, it was found that very few studies included effect size indices with tests for statistical significance and none produced a U_1 measure when any of the variants of *d* were reported. Further, beyond studies, for example, by Hedges (1981) or Kraemer (1983) related to the upward bias and skewness associated with *d* in small samples, it appears in the scholarly literature that *d* as a percent of non-overlap has not been studied to evaluate any bias affiliated with variants of *d*, d_t and d_{adj}, substituted for it in the calculation of U_1, except for what has been provided by Cohen (1988).

Thus, the intent of this research was to examine U_1 under varying sizes of *d* and n (i.e., $n_1 = n_2$). That is, this research looked at *d* values of .2, .5, .8, 1.00, and 1.50, which represent in educational research typically small to extremely large effect sizes. The sizes of n were 10, 20, 40, 50, 80, and 120, which represent in educational research small to large sample sizes. It should be noted, though, as was first discussed by Glass, McGaw, and Smith (1981), and reiterated by Cohen (1988), about the previously-mentioned *d* effect size target values and their importance:

> these proposed conventions were set forth throughout with much diffidence, qualifications, and invitations not to employ them if possible. The values chosen had no more reliable a basis than my own intuition. They were offered as conventions because they were needed in

a research climate characterized by a neglect of attention to issues of magnitude (p. 532).

Using the work of Aaron, Kromrey, and Ferron (1998), this study's tables will display the bias and proportional bias found in each U_1 measure found via both d_t and d_{adj}. As noted in the Aaron et al. research, the current study defines bias as the difference between the tabled value of U_1, derived from the standard *d* formula and presented by Cohen (1988) as Table 2.2.1, and the presented U_1 value resultant from d_t and d_{adj}., respectively. Proportional bias, or the "size of [the] bias as a proportion of the actual effect size estimate" (Aaron et al., p. 9), will be defined as the bias found above divided by the presented estimate for U_1 derived from both d_t and d_{adj}, respectively (see Tables 1 and 2).

Results

Using syntax written in SPSS v. 12.0 to obtain the results of the study, Tables 1 and 2 indicated, as would be expected, that regardless of the variant of *d* used, as the value of *d* increased, the bias in U_1 decreased. For example, Table 1 shows that at a small value of d = .2, and also at a moderate value of d = .5, the bias for small to moderate sample sizes ranged from about 1% to over 4%. As the value of *d* increased into the large effect size range of d = .8 to 1.50, the bias for the same sample sizes ranged from about 3% to under 1%.

The bias related to the U_1 measure for both forms of *d* used in this study was similar with both variants of *d*, the bias was constant with small sample sizes having 3% to 4% bias, moderate sample sizes having about 1%, and large sample sizes having very small amounts of bias. More specifically, it did appear, though, that the bias related to d_{adj} decreased more readily after d =.2 than was seen with d_t. That is, when d = .20, the bias for d_t = 4.5% and the bias for d_{adj} = 4.3 %, which were very similar. However, when d = .5, d_t incurred a bias of 4.4%, while the bias for d_{adj} = 3.5%. This trend continued to d = 1.50, with d_{adj} incurring less bias than d_t, or stated another way, d_t had more of a biased effect on U_1. d_t's over-estimation property was also noted by Thompson and Schumacker (1997) in a study that assessed the effectiveness of the binomial effect size display.

Conclusion

Finally, as was found by Aaron et al. (1998), Hedges (1981), and Kraemer (1983), this study added to the literature that the biases found in variants of *d*, in this instance d_t and d_{adj}, had the largest influence on U_1 measures used with smaller sample sizes, specifically when n_1 and n_2 = 10. Although not looking at U_1 measures per se, the Aaron et al., Hedges, and Kraemer studies showed the effect of small sample sizes on *d* and variants of *d* when n_1 and n_2 = 5 or 10.

The current study indicated that bias for variants of *d* tended to subside and become more manageable, in terms of precision of estimation, around 1% to 2% when n_1 and n_2 = 20, or beyond very small sample sizes of n_1 and n_2 = 5 and 10. This is favorable for educational and behavioral sciences research designs that contain sample sizes typically of less than 100 participants (Huberty & Mourad, 1980). Thus, both d_t and d_{adj} tended to manage bias in the U_1 measure quite well for smaller to moderate sample sizes.

Table 1: Bias Affiliated with Estimates of U_1 Derived from d_t

$n_1 = n_2$	**d**	**U_1**	**U_1 via d_t**	**Bias ($U_1 - U_1\ d_t$)**	**Proportional Bias (Bias / $U_1\ d_t$)**
10	.2	14.7	15.4	.7	.045
20	.2	14.7	15.1	.4	.026
40	.2	14.7	14.9	.2	.013
50	.2	14.7	14.8	.1	.007
80	.2	14.7	14.8	.1	.007
120	.2	14.7	14.7	0	0

Table 1 Continued.

$n_1 = n_2$	d	U_1	U_1 via d_t	Bias ($U_1 - U_1\ d_t$)	Proportional Bias (Bias / $U_1\ d_t$)
10	.5	33.0	34.5	1.5	.044
20	.5	33.0	33.7	.7	.021
40	.5	33.0	33.4	.4	.012
50	.5	33.0	33.3	.3	.009
80	.5	33.0	33.2	.2	.006
120	.5	33.0	33.1	.1	.003

$n_1 = n_2$	d	U_1	U_1 via d_t	Bias ($U_1 - U_1\ d_t$)	Proportional Bias (Bias / $U_1\ d_t$)
10	.8	47.4	49.2	1.8	.037
20	.8	47.4	48.3	.9	.019
40	.8	47.4	47.8	.4	.008
50	.8	47.4	47.7	.3	.006
80	.8	47.4	47.6	.2	.004
120	.8	47.4	47.5	.1	.002

$n_1 = n_2$	d	U_1	U_1 via d_t	Bias ($U_1 - U_1\ d_t$)	Proportional Bias (Bias / $U_1\ d_t$)
10	1.00	55.4	57.4	2.0	.035
20	1.00	55.4	56.4	1.0	.018
40	1.00	55.4	55.9	.5	.009
50	1.00	55.4	55.8	.3	.005
80	1.00	55.4	55.6	.2	.004
120	1.00	55.4	55.5	.1	.002

$n_1 = n_2$	d	U_1	U_1 via d_t	Bias ($U_1 - U_1\ d_t$)	Proportional Bias (Bias / $U_1\ d_t$)
10	1.50	70.7	72.7	2.0	.028
20	1.50	70.7	71.7	1.0	.014
40	1.50	70.7	71.2	.5	.007
50	1.50	70.7	71.1	.4	.006
80	1.50	70.7	70.9	.2	.003
120	1.50	70.7	70.8	.1	.001

Table 2: Bias Affiliated with Estimates of U_1 Derived from d_{adj}

$n_1 = n_2$	d	U_1	U_1 via d_{adj}	Bias ($U_1 - U_1\ d_{adj}$)	Proportional Bias (Bias / $U_1\ d_{adj}$)
10	.2	14.7	14.1	.6	.043
20	.2	14.7	14.4	.3	.021
40	.2	14.7	14.6	.1	.007
50	.2	14.7	14.6	.1	.007
80	.2	14.7	14.6	.1	.007
120	.2	14.7	14.7	0	0

$n_1 = n_2$	d	U_1	U_1 via d_{adj}	Bias ($U_1 - U_1\ d_{adj}$)	Proportional Bias (Bias / $U_1\ d_{adj}$)
10	.5	33.0	31.9	1.1	.035
20	.5	33.0	32.5	.5	.015
40	.5	33.0	32.8	.2	.006
50	.5	33.0	32.8	.2	.006
80	.5	33.0	32.9	.1	.003
120	.5	33.0	33.0	0	0

$n_1 = n_2$	d	U_1	U_1 via d_{adj}	Bias ($U_1 - U_1\ d_{adj}$)	Proportional Bias (Bias / $U_1\ d_{adj}$)
10	.8	47.4	45.9	1.5	.033
20	.8	47.4	46.7	.7	.015
40	.8	47.4	47.1	.3	.006
50	.8	47.4	47.1	.3	.006
80	.8	47.4	47.2	.2	.004
120	.8	47.4	47.3	.1	.002

$n_1 = n_2$	d	U_1	U_1 via d_{adj}	Bias ($U_1 - U_1\ d_{adj}$)	Proportional Bias (Bias / $U_1\ d_{adj}$)
10	1.00	55.4	53.8	1.6	.030
20	1.00	55.4	54.7	.7	.013
40	1.00	55.4	55.1	.3	.005
50	1.00	55.4	55.1	.3	.005
80	1.00	55.4	55.2	.2	.004
120	1.00	55.4	55.3	.1	.002

$n_1 = n_2$	d	U_1	U_1 via d_{adj}	Bias ($U_1 - U_1\ d_{adj}$)	Proportional Bias (Bias / $U_1\ d_{adj}$)
10	1.50	70.7	69.1	1.6	.023
20	1.50	70.7	69.9	.8	.011
40	1.50	70.7	70.3	.4	.006
50	1.50	70.7	70.4	.3	.004
80	1.50	70.7	70.5	.2	.003
120	1.50	70.7	70.6	.1	.001

Reference

Aaron, B., Kromrey, J. D., & Ferron, J. M. (1998, November). *Equating r-based and d-based effect size indices: Problems with a commonly recommended formula*. Paper presented at the annual meeting of the Florida Educational Research Association, Orlando, FL.

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Elster, R. S., & Dunnette, M. D. (1971). The robustness of Tilton's measure of overlap. *Educational and Psychological Measurement, 31*, 685-697.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.

Hedges, L. V. (1981). Distribution theory for Glass' estimator of effect size and related estimators. *Journal of Educational Statistics, 6*, 107-128.

Huberty, C. J., & Mourad, S. A, (1980). Estimation in multiple correlation/prediction. *Educational and Psychological Measurement, 40*, 101-112.

Kraemer, H. C. (1983). Theory of estimation and testing of effect sizes: Use in meta-analysis. *Journal of Educational Statistics, 8*, 93-101.

Rosenthal, R. (1991). (Series Ed.), *Meta-analytic procedures for social research*. Newbury Park, CA: Sage Publications.

Thompson, K. N., & Schumacker, R. E. (1997). An evaluation of Rosenthal and Rubin's binomial effect size display. *Journal of Educational and Behavioral Statistics, 22*, 109-117.

Tilton, J. W. (1937). The measurement of overlapping. *Journal of Educational Psychology*, *28*, 656-662.