Psychological Methods
1998, Vol. 3, No. 1, 46–54

1082-989X/98/$3.00

Using the Normal Quantile Plot to Explore Meta-Analytic Data Sets

Morgan C. Wang
University of Central Florida

Brad J. Bushman
Iowa State University

In a meta-analysis, graphical displays can be used to check statistical assumptions for numerical procedures and they can be used to discover important patterns in the data. The authors propose the normal quantile plot as a preferred alternative to the funnel plot for such purposes. The normal quantile plot, like the funnel plot, can be used to investigate whether all studies come from a single population and to search for publication bias. However, the normal quantile plot is easier to interpret than the funnel plot, especially when it includes 95% confidence bands. In addition, the normal quantile plot can be used to check the normality assumption for numerical procedures. The funnel plot cannot be used for this latter purpose.

Most people have heard the phrase, "A picture is worth a thousand words." This phrase seems especially applicable to producers and consumers of meta-analytic reviews. In a meta-analysis, graphical displays (figures, plots) can be used to enhance numerical analyses in at least two ways. First, graphical displays can be used to discover patterns and relations among variables in a meta-analysis. Second, graphical displays can be used to check statistical assumptions on which numerical analyses are based.

One of the most popular graphical displays for exploring meta-analytic data sets is the funnel plot (Light & Pillemer, 1984). In this article, we begin by describing the funnel plot and its uses. We then propose the normal quantile plot as a preferred alternative to the funnel plot.

The Funnel Plot

A *funnel plot* is a two-dimensional graph with sample size on one axis and effect-size estimate on the other axis. Most statisticians would recognize a funnel plot as a special type of scatter plot. The funnel plot capitalizes on the well-known statistical principle that sampling error decreases as sample size increases. In a meta-analysis, the funnel plot can be used to investigate whether all studies come from a single population and to search for publication bias. We discuss each of these uses in turn.

Investigating Whether All Studies Come From a Single Population

If the studies all come from a single population, then the plot should look like a funnel with the diameter of the funnel decreasing as sample size increases. As sample size increases, the effect-size estimates narrow in on the true population effect size. This is an important use of funnel plots because if all the studies come from a single population, then it makes sense to average the sample effect sizes to estimate the true population effect size.

Figure 1 shows a funnel plot using a simulation data set of 120 studies that compared the means from two groups. The data were simulated to have a population mean difference of 0 and a common variance of 1. Note that as sample size increases the diameter of the funnel plot decreases and converges to the true effect size. This funnel plot suggests that the 120 studies do come from a single population with an effect size of zero. Of course, one would also want to formally test whether the 120 effect size estimates are homogeneous (see Hedges & Olkin, 1985, pp. 122–128).

If studies come from two or more populations, the funnel plot will not converge to a single value. For example, Figure 2 shows a funnel plot using a simulation data set of 120 studies. The mean difference was set at 0 for half of the studies and at 0.8 for the

Correspondence concerning this article should be addressed to Morgan C. Wang, Department of Statistics, University of Central Florida, Orlando, Florida 32816-2370, or to Brad J. Bushman, Department of Psychology, Iowa State University, Ames, Iowa 50011. Electronic mail may be sent to Morgan C. Wang at cwang@pegasus.cc.ucf.edu or to Brad J. Bushman at bushman@iastate.edu.

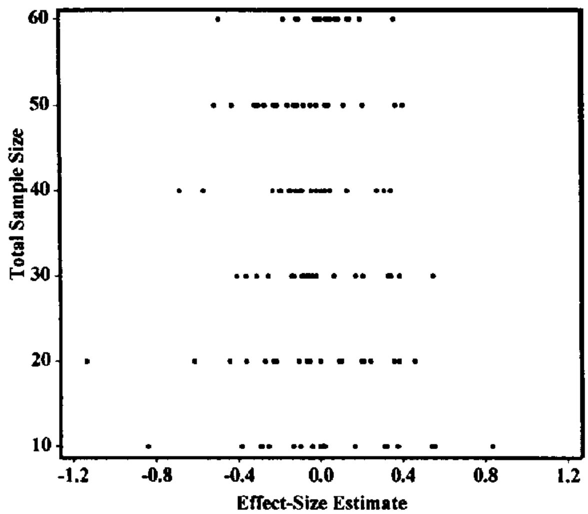

Figure 1. Funnel plot for simulated set of 120 studies with mean difference set at 0 and common variance set at 1.

other half. The common variance was set at 1 for both groups of studies. Note that as the sample size increases, the diameter of the funnel plot decreases only slightly. This funnel plot suggests that the 120 studies do not come from a single population. One should also formally test whether the effect-size estimates are homogenous. It is worth noting that although the homogeneity test is useful in determining whether the studies come from a single population, it is not very useful in determining how many populations the studies come from. The funnel plot can help one decide how many populations the studies come from if the population effect sizes are quite different. However, it is difficult to determine how many populations the studies come from in Figure 2 because the population effect sizes are not different enough.

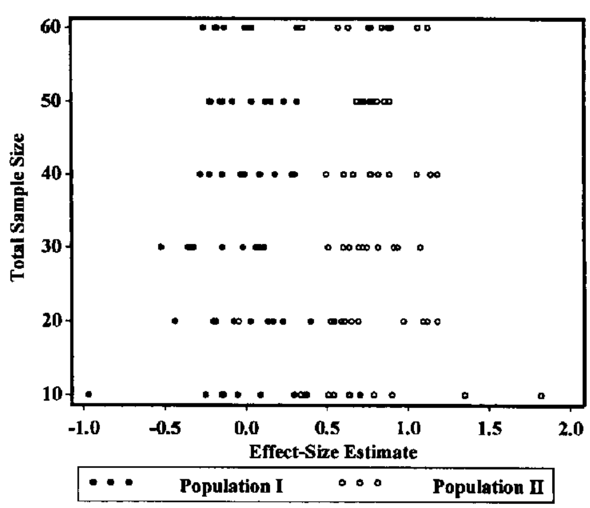

Figure 2. Funnel plot for simulated set of 120 studies from two populations, 60 studies with mean difference set at 0 and the other 60 with mean difference set at 0.8. The common variance was set at 1 for both groups of studies.

Searching for Publication Bias

Funnel plots also can be used to identify publication bias. It is well documented that studies reporting statistically significant results are more likely to be published than are studies reporting nonsignificant results (e.g., Greenwald, 1975). In meta-analysis, this conditional publication of studies with significant results has been labeled the "file drawer problem" (Rosenthal, 1979). The most extreme version of this problem would result if only 1 out of 20 studies conducted was published and the remaining 19 studies were located in researchers' file drawers (or garbage cans). If publication bias is a problem, then the studies included in a meta-analysis may represent a biased subset of the total number of studies conducted.

The shape of a funnel plot can suggest whether publication bias exists. If the true effect size is zero, a few studies will still have significant results—those with very large effect-size estimates (positive or negative) and those with very large sample sizes. Studies with small effect-size estimates and small sample sizes will have nonsignificant results. If publication bias exists, the middle of the funnel plot will be "hollow" (i.e., there are no effect-size estimates in the center of the funnel plot). The data plotted in Figure 3 were obtained from a simulation study in which the mean difference was set at 0 and the common variance was set at 1, and studies with nonsignificant results at the .05 level were deleted. Note that Figure 3 is "hollow."

On the other hand, if the true effect size differs from zero, then publication bias shows up in a different form on the funnel plot. Studies with small effect size estimates and small sample sizes will still have nonsignificant results and will not appear in the plot. If publication bias exists, there will be a "bite" out of the funnel where sample sizes and effect-size estimates are small. The data plotted in Figure 4 were obtained from a simulation study in which the mean difference was set at 0.5 and the common variance was set at 1, and studies with nonsignificant results at the .05 level were deleted. Note that there is a "bite" out of the lower left-hand corner of the plot, where sample sizes and effect-size estimates are small.

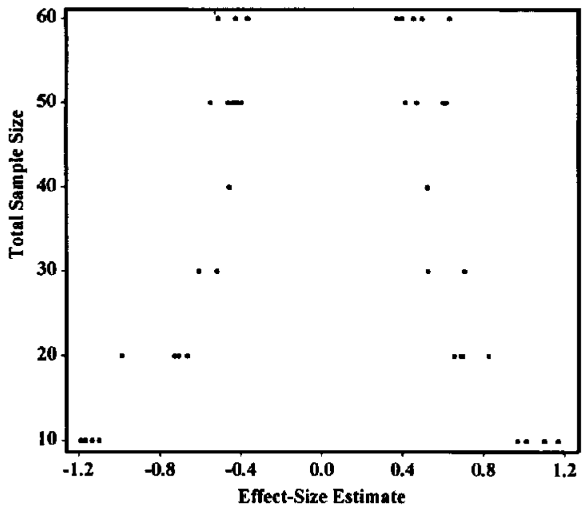

Figure 3. Funnel plot for simulated set of 120 studies with mean difference set at 0 and common variance set at 1. Studies with nonsignificant results at the .05 level were deleted from the data set.

If publication bias exists, one would expect unpublished studies (e.g., master's theses, doctoral dissertations) to report smaller effect-size estimates than published studies. A real-world illustration of the use of funnel plots to detect publication bias is shown in Figure 5. The data were taken from a meta-analytic review of 54 studies of the effects of psychoeducational interventions on postsurgical hospital stay

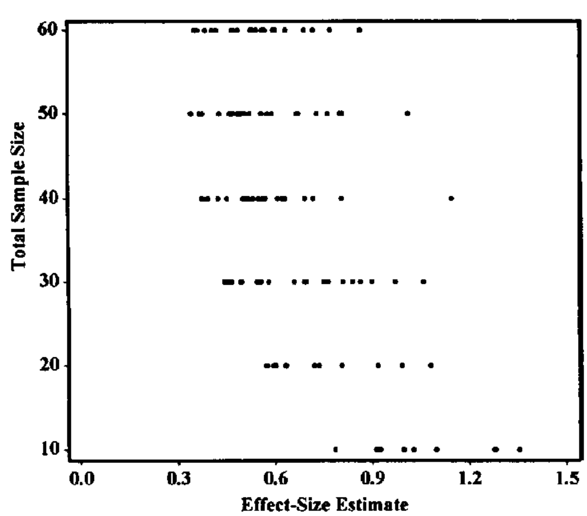

Figure 4. Funnel plot for simulated set of 120 studies with mean difference set at 0.5 and common variance set at 1. Studies with nonsignificant results at the .05 level were deleted from the data set.

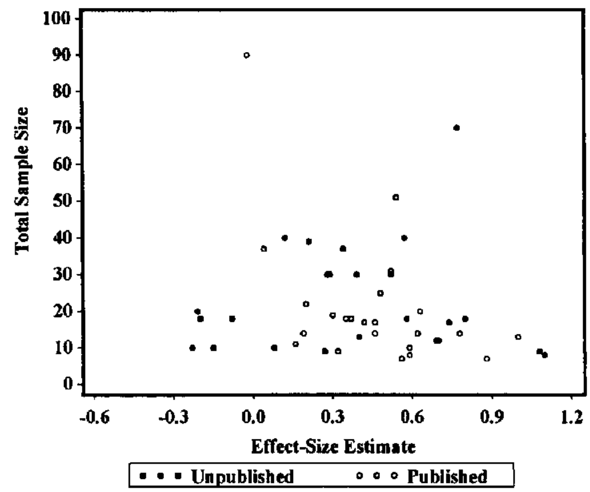

Figure 5. Funnel plot for Devine and Cook's (1983) meta-analysis of the effects of psychoeducational interventions on postsurgical hospital stay.

(Devine & Cook, 1983). Overall, psychoeducational interventions reduced hospital stay by about 3 days. However, as can be seen in left corner of Figure 5, unpublished studies had smaller effects than did published studies, suggesting the presence of publication bias.

Problems With Funnel Plots

Unfortunately, there are three problems with funnel plots. First, it is very difficult to determine (using one's eye) whether the data are shaped like a funnel, especially when the number of studies included in a meta-analytical review is small. For example, Figure 6 shows a funnel plot using a simulation data set of 15 studies that compared the means from two groups. It is very difficult to determine whether the data in Figure 6 are shaped like a funnel. In contrast, it is much easier to determine whether the data fall on a straight line in a normal quantile plot (see description below). If the normal quantile plot also includes 95% confidence bands, then even more reliable judgments about the data can be made.[1] For example, Figure 7 uses the

[1] Several formulae can be used to approximate the standard error estimate used in constructing the 95% confidence interval bands (Chambers, Cleveland, Kleiner, & Tukey, 1983; Kendall & Stuart, 1977). We used the formula by Chambers et al. to create the normal quantile plots in this article. A SAS MACRO (SAS Institute, 1990) for constructing the quantile normal plot with confidence interval bands is given in the Appendix.

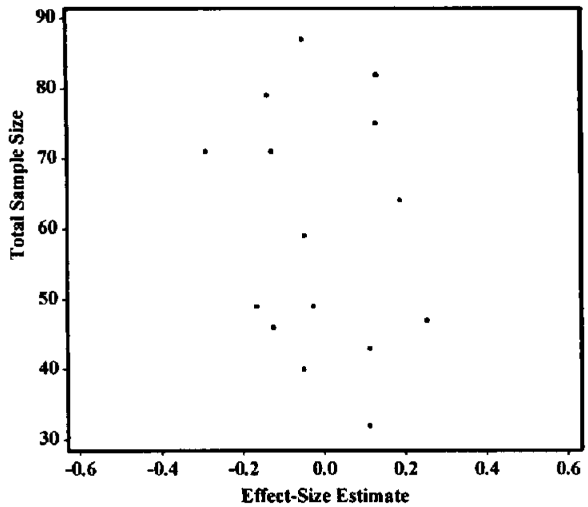

Figure 6. Funnel plot for simulated set of 15 studies with mean difference set at 0 and common variance set at 1.

same simulation data set as Figure 6. It is easy to see that the data in Figure 7 fall on a straight line and within the 95% confidence bands.

A second problem with funnel plots is that they do not use the important fact that the effect-size estimate in each study in a meta-analysis has an approximately normal distribution if the study has a large enough sample size. Most meta-analytical procedures are based on the assumption of normality (or asymptotic normality). It is therefore important to check this assumption before applying meta-analytic procedures that assume normality.

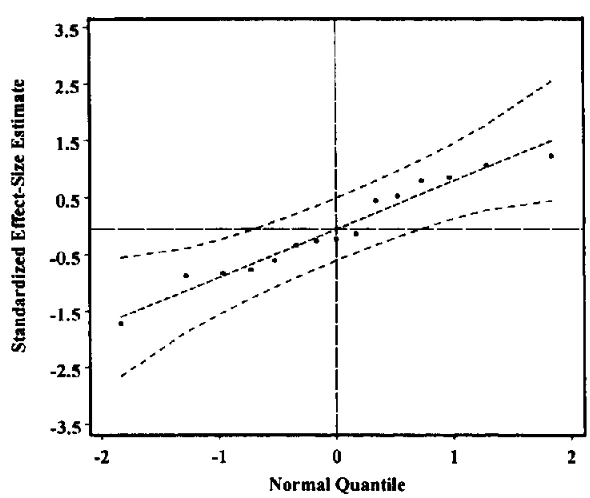

Figure 7. Normal quantile plot for simulated set of 15 studies with mean difference set at 0 and common variance set at 1.

A third problem with funnel plots is that data can look like a funnel even if the studies come from more than one population if the populations have the same mean but different variances. For example, Figure 8 shows a funnel plot using a simulation data set of 120 studies. Half of the studies were simulated to have a mean difference of 0 and a common variance of 1; the other half were simulated to have a mean difference of 0 and a common variance of 4. Note that as the sample size increases, the diameter of the funnel plot decreases and converges to the value zero. The funnel plot suggests that the 120 studies come from a single population with a mean difference of zero. This conclusion is wrong, however, because the studies come from two different populations.

Normal Quantile Plots

In a *quantile–quantile plot,* two distributions are compared by plotting their quantiles (or percentiles) against each other. The quantiles for one distribution are plotted on the X axis and the quantiles for the other distribution are plotted on the Y axis. If the two distributions are similar, then the quantiles will also be similar and the points will fall close to the line $X = Y$. Any deviation from the $X = Y$ reveals how the distributions differ.

In a *normal quantile plot,* the quantiles of an observed distribution are plotted against the quantiles of the standard normal distribution (i.e., the normal dis-

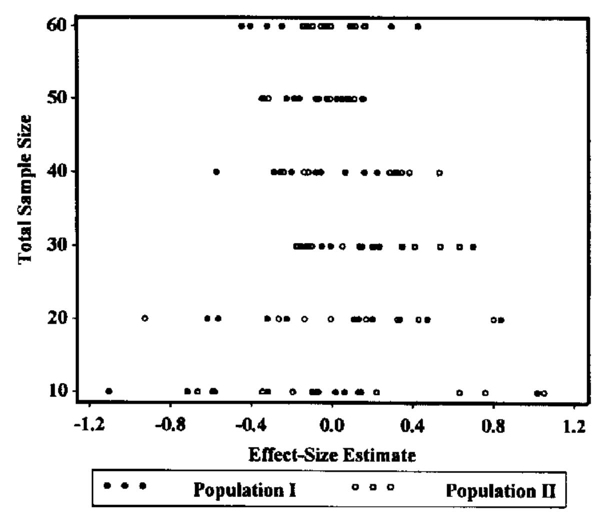

Figure 8. Funnel plot for simulated set of 120 studies from two populations, 60 studies with mean difference set at 0 and common variance set at 1 and the other 60 with mean difference set at 0 and common variance set at 4.

tribution with mean of 0 and standard deviation of 1). If the observed data have a standard normal distribution, the points on the plot will fall close to the line $X = Y$ and the plot should look like Figure 9. The slope of the plotted line is approximately equal to 1, the standard deviation of the observed distribution, and the Y coordinate of the center point is equal to 0, the population mean of the observed distribution. Figure 10 depicts data from a normal distribution with a mean of 0.5 and a standard deviation of 1. Note that the points on the plot fall close to the line $X = Y + 0.5$ (i.e., the slope is 1 and the Y coordinate of the center point is 0.5). Figure 11 depicts that data from a normal distribution with a mean of 0 and a standard deviation of 2. Note that the points on the plot fall close to the line $X = 2Y$. Figure 12 depicts the data from a normal distribution with a mean of 0.5 and a standard deviation of 2. Note that the points on the plot fall close to the line $X = 2Y + 0.5$. Thus, the points on the normal quantile plot should look like an "approximate straight line" and most of the points should be located inside the 95% confidence interval bands if the observed data do come from a single normal population. If the data do not fall on a straight line or if several points are located outside the 95% confidence bands, the observed data probably do not come from a single normal population.

In a meta-analysis, the observed distribution consists of the effect-size estimates from a set of studies. Each standardized effect-size estimate (i.e., effect-size estimate divided by its estimated standard error) should have a standard deviation of 1. According to the central limit theorem, the distribution of effect-size estimates will be approximately normal if the sample size in each study is large enough (say at least 30 in each study). Thus, plotted line in the normal quantile plot for meta-analytical data should be fairly straight and should have a slope close to 1 if the studies come from a single population and if the sample size for each study is large enough.

In a meta-analysis, the normal quantile plot can be used to (a) check the normality assumption, (b) investigate whether all studies come from a single popula-

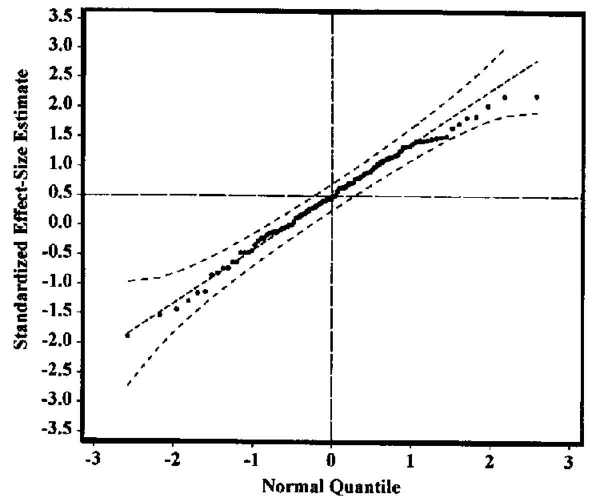

Figure 10. Normal quantile plot for simulated data set with a mean of 0.5 and standard deviation of 1 (N = 100).

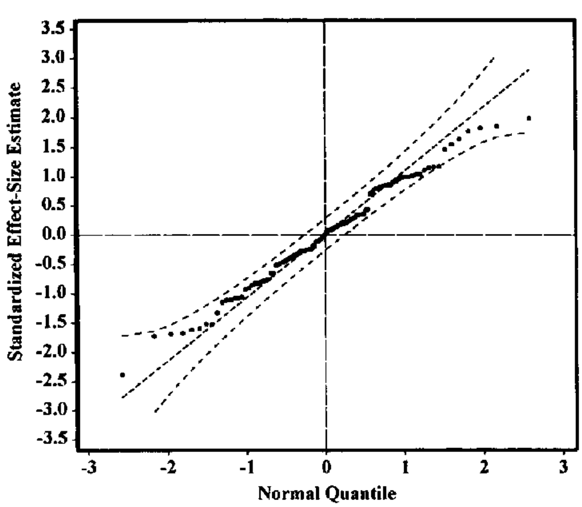

Figure 9. Normal quantile plot for simulated set with a mean of 0 and standard deviation of 1 (N = 100).

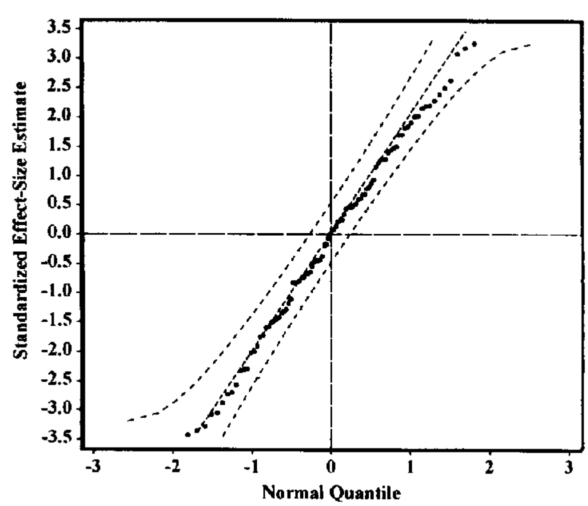

Figure 11. Normal quantile plot for simulated data set with a mean of 0 and standard deviation of 2 (N = 100).

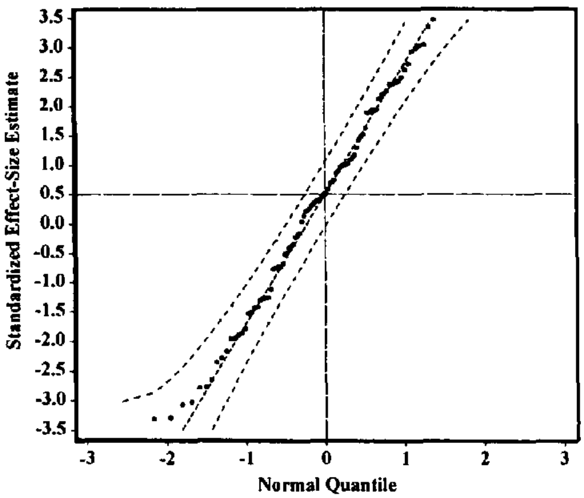

Figure 12. Normal quantile plot for simulated data set with a mean of 0.5 and standard deviation of 2 ($N = 100$).

tion, and (c) search for publication bias. We discuss each of these uses in turn.

Checking the Normality Assumption

Figure 13 depicts the results of a meta-analysis of 43 studies on sex differences in physical aggression (Bettencourt & Miller, 1996). Note that the data fall on a straight line and within the 95% confidence interval bands. Figure 13 suggests that these data are normally distributed.

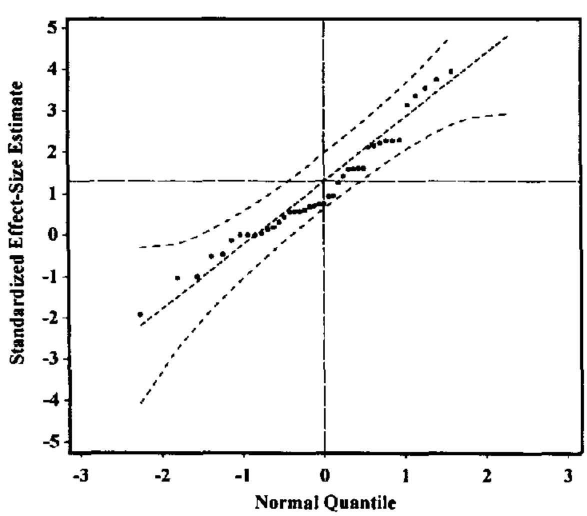

Figure 13. Normal quantile plot for Bettencourt and Miller's (1996) meta-analysis of sex differences in physical aggression.

Investigating Whether All Studies Come From a Single Population

Figure 14 shows a normal quantile plot using the same simulation data set as for Figure 2. The mean difference was set at 0 for half of the studies and at 0.8 for the other half. The common variance was set at 1 for both groups of studies. The curve in Figure 14 is S-shaped and has one "bump" below and one "bump" above the center dashed line, reflecting the fact that the 120 studies come from two different populations. Figure 14 is also easier to interpret than Figure 2. From Figure 14, it is easy to determine that the studies come from two populations. From Figure 2, it is difficult to determine how many populations the studies come from.

Searching for Publication Bias

Figure 15 shows a normal quantile plot using the same simulation data set as for Figure 3. The studies were simulated to have a mean difference of 0 and a common variance of 1, and studies with nonsignificant results at the .05 level were deleted. The fitted curve on Figure 15 has a suspicious gap around zero where there are no data. This gap is due to the fact that studies with nonsignificant effects were deleted. Because the gap is located around $Y = 0$, the true population effect size is very close to zero. Figure 15 sug-

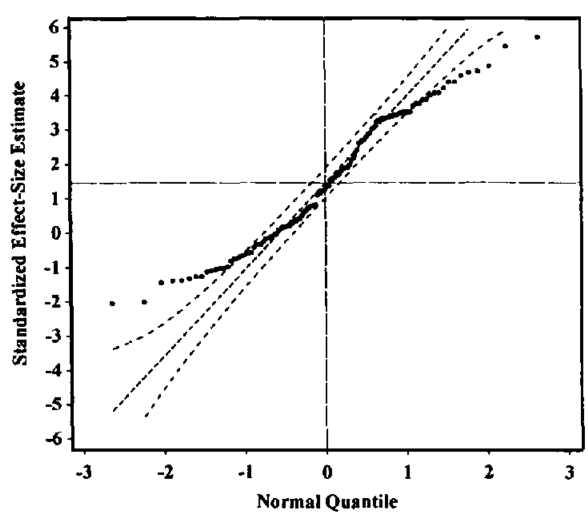

Figure 14. Normal quantile plot for simulated set of 120 studies from two populations, 60 studies with mean difference set at 0 and the other 60 with mean difference set at 0.8. The common variance was set at 1 for both groups of studies.

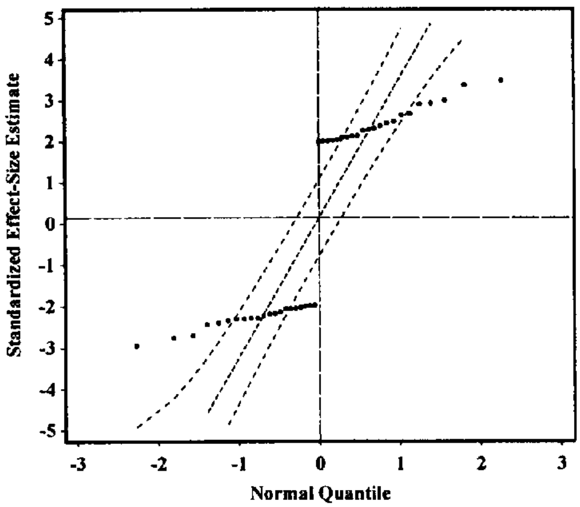

Figure 15. Normal quantile plot for simulated set of 120 studies with mean difference set at 0 and common variance set at 1. Studies with nonsignificant results at the .05 level were deleted from the data set.

gests the presence of the type of publication bias that exists when the population effect size equals zero.

Figure 16 shows a normal quantile plot using the same simulation data set as for Figure 4. The studies were simulated to have a mean difference of 0.5 and a common variance of 1, and studies with nonsignificant results at the .05 level were deleted. The curve is U-shaped, reflecting the fact that the data have an asymmetric distribution that is skewed to the right. The long right tail is due to the fact that studies with nonsignificant results were deleted. Figure 16 suggests the presence of the type of publication bias that exists when the population effect size differs from zero.

Figure 17 shows a normal quantile plot using Devine and Cook's (1983) meta-analysis of the effects of psychoeducational interventions on postsurgical hospital stay, with only published studies. Although the points not exceed the 95% confidence bands, almost all points on the upper end of the curve are above the straight line. The departure from linearity at the upper end of the curve reflects the fact that the data have an asymmetric distribution that is slightly skewed to the right. The long right tail is due to the fact the unpublished studies were deleted.

Figure 18 shows a normal quantile plot from Devine and Cook's (1983) meta-analysis that includes both published and unpublished studies. Note that the points fall on a straight line and do not exceed the 95% confidence bands. Figure 18 suggests that the standardized effect-size estimates have a normal distribution and that no publication bias exists.

Discussion

In this article, we have tried to illustrate how the normal quantile plot can be used to explore the results of a meta-analysis. We believe (and we hope our read-

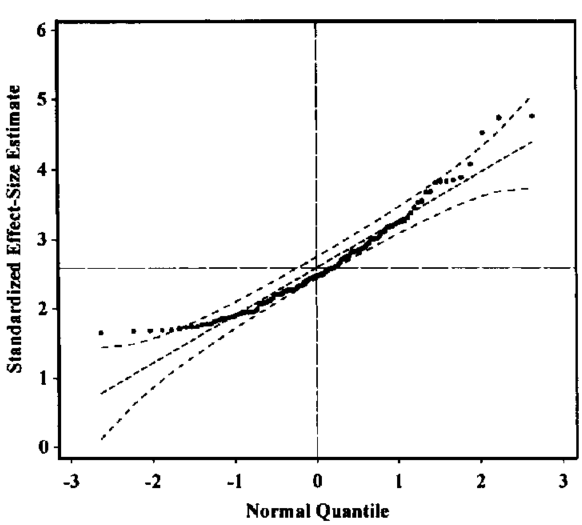

Figure 16. Normal quantile plot for simulated set of 120 studies with mean difference set at 0.5 and common variance set at 1. Studies with nonsignificant results at the .05 level were deleted from the data set.

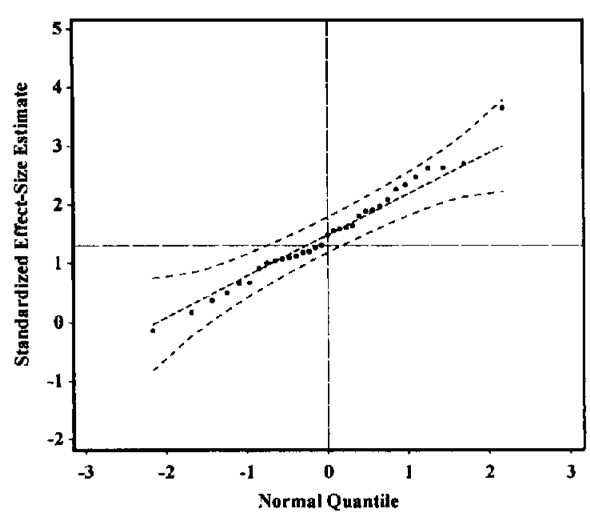

Figure 17. Normal quantile plot for Devine and Cook's (1983) meta-analysis of the effects of psychoeducational interventions on postsurgical hospital stay (published studies only).

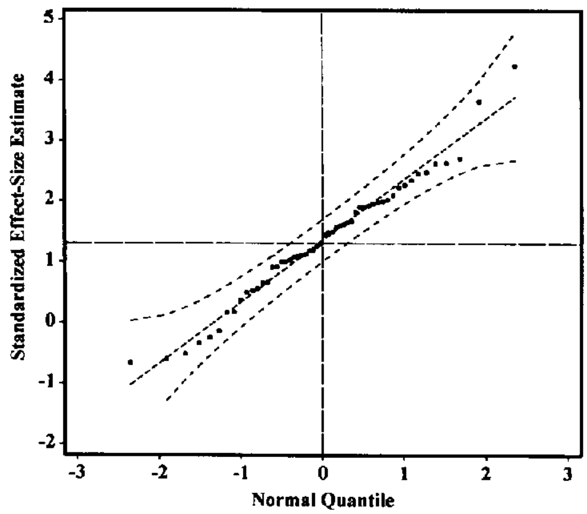

Figure 18. Normal quantile plot for Devine and Cook's (1983) meta-analysis of the effects of psychoeducational interventions on postsurgical hospital stay (published and unpublished studies).

ers agree) that normal quantile plots are less ambiguous to interpret than funnel plots. It is easier to determine whether the effect-size estimates from a meta-analysis fall on a straight line than to determine whether they are shaped like a funnel. The inclusion of 95% confidence bands on the normal quantile plot reduces ambiguity even further. Like all other statistical plots, normal quantile plots have weaknesses. For example, if the studies come from more than one population with means that are very close, one might not be able to detect population mean differences using a normal quantile plot. In general, graphical procedures are not intended to replace the formal statistical tests designed for the same purpose. However, the use of graphical displays can enhance the use of statistical tests and can be used to check the assumptions on which they are based.

We believe, with Tukey (1977), that a data set should be explored. Exploratory data analysis may be even more important when conducting a meta-analysis than when conducting a primary research study. Readers might be more likely to dismiss the results from a single study than the results from an integrative review of several studies. The reader must have confidence that the results from a meta-analysis can be generalized to the population of interest. If the studies come from more than one population, it is not appropriate to combine the results from all studies to estimate a single population effect size. Likewise, if publication bias exists, the combined estimate based on the studies included in the meta-analysis might be different from the population effect size.

It seems paradoxical that many authors use state-of-the-art procedures in conducting a meta-analysis, yet they still use display methods that were popular before the age of computers, graphics, and exploratory data analysis (Light, Singer, & Willett, 1994). The normal quantile plot can be a useful tool to explore meta-analytic data sets.

References

Bettencourt, B. A., & Miller, N. (1996). Sex differences in aggression as a function of provocation: A meta-analysis. *Psychological Bulletin, 119,* 422–447.

Chambers, J. M., Cleveland, W. S., Kleiner, B., & Tukey, P. A. (1983). Graphical methods for data analysis. Boston, MA: Duxbury.

Devine, E. C., & Cook, T. D. (1983). A meta-analytic analysis of effects of psycho-educational interventions on length of hospital stay. *Nursing Research, 32,* 267–274.

Greenwald, A. G. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin, 82,* 1–20.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* New York: Academic Press.

Kendall, M. G., & Stuart, A. (1977). *The advanced theory of statistics* (Vol. 1). Hafner, NY: Wiley.

Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research.* Cambridge, MA: Harvard University Press.

Light, R. J., Singer, J. D., & Willett, J. B. (1994). The visual presentation and interpretation of meta-analyses. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 439–453). New York: Sage.

Rosenthal, R. (1979). The "file-drawer problem" and tolerance for null results. *Psychological Bulletin, 86,* 638–641.

SAS Institute (1990). *SAS/GRAPH software: Reference* (Version 6, Vols. 1 and 2). Cary, NC: Author.

Tukey, J. W. (1977). *Exploratory data analysis.* Reading, MA: Addison-Wesley.

(*Appendix follows*)

Appendix

SAS Macro for the Normal-Quantile Plot

```
%MACRO CIQQPLOT(AA,BB,CC,DD);
PROC SORT DATA=&AA; BY &DD;
PROC UNIVARIATE NOPRINT DATA=&AA;
  VAR &DD;
  OUTPUT OUT=STATS N=NOBS MEAN=MEAN STD=STD;
DATA QUANTILE;
  SET&AA;
  IF_N_=1 THEN SET STATS;
  P=(_N_-0.5)/NOBS;
  Z=PROBIT(P);
  NORMAL = MEAN + Z * STD;
  SE = (STD/( (1/SQRT(8*ATAN(1) ) )*EXP(-(0.5*Z*Z) ) ) )*SQRT(P*(1-P)/NOBS);
  LOWER = NORMAL - 1.96 * SE;
  UPPER = NORMAL + 1.96 * SE;
  KEEP LOWER UPPER NORMAL &DD Z;
PROC GPLOT DATA = QUANTILE GOUT = &BB;
  PLOT &DD * Z = 1
        NORMAL * Z = 2
        LOWER * Z = 3
        UPPER * Z = 3
        / OVERLAY
        VAXIS = AXIS1 VMINOR = 0
          HAXIS = AXIS2 HMINOR = 0
          VREF = 0/* The value of VREF should set to the mean of the standardized effect-size estimate */
          LVREF = 5
          HREF = 0
          LHREF = 5
          NAME = '&CC'
          FRAME;
  SYMBOL1 V=DOT HEIGHT = 1.5 I = NONE COLOR = BLACK;
  SYMBOL2 V = NONE I = JOINT L = 3 COLOR = BLACK;
  SYMBOL3 V = NONE I = JOINT L = 20 COLOR = BLACK;
  AXIS1 OFFSET = (2) LABEL = (H = 3 A = 90 R = 0 "Standardized Effect-Size Estimate")
        VALUE = (H = 3) WIDTH = 2;
  AXIS2 LENGTH = 80 LABEL = (H = 3 "Normal Quantile")
      OFFSET = (2) VALUE = (H = 3) WIDTH = 2;
RUN; QUIT;
%MEND CIQQPLOT;
```

Received September 6, 1996
Revision received March 5, 1997
Accepted March 19, 1997 ■