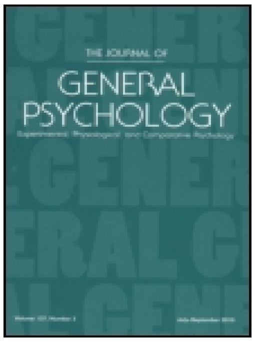

The Journal of General Psychology

ISSN: 0022-1309 (Print) 1940-0888 (Online) Journal homepage: https://www.tandfonline.com/loi/vgen20

Averaging Correlations: Expected Values and Bias in Combined Pearson *r*s and Fisher's *z* Transformations

David M. Corey , William P. Dunlap & Michael J. Burke

The Journal of General Psychology, 1998, *125*(3), 245–261

Averaging Correlations: Expected Values and Bias in Combined Pearson *r*s and Fisher's *z* Transformations

DAVID M. COREY
WILLIAM P. DUNLAP
Department of Psychology
Tulane University

MICHAEL J. BURKE
Department of Psychology and Freeman School of Business
Tulane University

ABSTRACT. R. A. Fisher's z (z'; 1958) essentially normalizes the sampling distribution of Pearson r and can thus be used to obtain an average correlation that is less affected by sampling distribution skew, suggesting a less biased statistic. Analytical formulae, however, indicate less expected bias in average r than in average z' back-converted to average $r_{z'}$. In large part because of this fact, J. E. Hunter and F. L. Schmidt (1990) have argued that average r is preferable to average $r_{z'}$. In the present study, bias in average r and average $r_{z'}$ was empirically examined. When correlations from a matrix were averaged, the use of z' decreased bias. For independent correlations, contrary to analytical expectations, average $r_{z'}$ was also generally the less biased statistic. It is concluded that (a) average $r_{z'}$ is a less biased estimate of the population correlation than average r and (b) expected values formulae do not adequately predict bias in average $r_{z'}$ when a small number of correlations are averaged.

THE CORRELATION COEFFICIENT can be a highly variable statistic. Very large sample sizes can reduce its variability; however, such samples are not always available in empirical settings. One means of addressing the problem of variability is repeatedly to compute correlation coefficients and take the average as a more accurate estimate of the population correlation (ρ).

The study reported in this article was supported in part by a Louisiana Education Quality Support Fund Fellowship awarded to David M. Corey. The authors thank Ron Landis and Billy Robinson for commenting on preliminary drafts of this article.

Address correspondence to David M. Corey, 2007 Stern Hall, Department of Psychology, Tulane University, New Orleans, LA 70118; e-mail: <dave@mailhost.tcs.tulane.edu>.

One instance in which this method may be used is in evaluating performance stability. For example, in learning paradigms, intertrial correlations tend to be low for early trials, but as the participant learns the task, performance stabilizes across later trials. To evaluate performance once it has stabilized, one may wish to identify the stable portion of the intercorrelation matrix and then obtain an average intercorrelation for that portion of the matrix (Carter, Kennedy, & Bittner, 1981). Meta-analysis is another instance in which an average correlation is used. Reporting findings across many correlational studies requires computation of the average correlation across studies. Such use of average correlation is mathematically distinct from the first example in that the correlations being averaged are not obtained from a common matrix but are independent.

For such situations, two methods of obtaining average correlations have been identified. First, an estimate of ρ can be obtained directly from the *r*s as a sample-weighted average of the observed *r*s:

$$\text{average } r = \frac{\sum(N_i r_i)}{\sum N_i}, \tag{1}$$

where N = sample size and r = observed correlation for each study.

However, the distribution of r is skewed when the absolute value of ρ is greater than zero, introducing a potential source of bias in the correlation coefficient. As a result, r is biased toward zero when the absolute value of ρ is greater than zero (i.e., the average r will underestimate the magnitude of ρ). In an alternate method of averaging, each r can first be converted into a Fisher's z (z'; Fisher, 1958):

$$z' = 0.5 \ln \frac{1+r}{1-r}. \tag{2}$$

Then, z' can be averaged and the result back-converted to r via Equation 3:

$$r_{z'} = \frac{e^{2z'} - 1}{e^{2z'} + 1}. \tag{3}$$

The advantage of this technique is that the z' transformation greatly reduces the skew in the r distribution, potentially yielding a more accurate estimate of ρ.

Some researchers (Hunter & Schmidt, 1990; Schmidt, Hunter, & Raju, 1988) have argued that the average untransformed r (average r) is a more accurate estimate of ρ than the average obtained using z' (average $r_{z'}$) is. In support of this claim, Hunter and Schmidt (1990) computed averages in both ways for a contrived set of correlation coefficients presented by James, Demaree, and Mulaik (1986). For those data, average $r = .50$, whereas average $r_{z'} = .51$. Hunter and Schmidt concluded that through the use of average $r_{z'}$, the "mean validity has been inflated by .01—in this case—a 2% increase" (p. 217). Though true, their statement is misleading; such wording implies that average r is an accurate estimate of ρ and that the average $r_{z'}$ is inflated relative to some desired outcome (i.e., ρ). It could just as accurately be stated that the average r is deflated relative to average $r_{z'}$. In other words, the fact that average $r_{z'}$ is larger than average r does not support the hypothesis that averaging untransformed rs provides a more accurate estimate of ρ than averaging z's and back-converting the outcome to r.

Another argument for using untransformed rs to obtain validity estimates centers on analytical evidence from Hunter, Schmidt, and Coggin (1988). Their analytical formulae (obtained from Hotelling, 1953) suggest that ρ may be better estimated by averaging rs directly than by using z' (Hunter & Schmidt, 1990; Schmidt et al., 1988). The claim by Hunter and Schmidt and Schmidt et al. that the expected bias using z' is always larger in magnitude than the bias in r is false; for N = 10 to 50 and ρ = .8 or .9, the expected value of average $r_{z'}$ is less biased than average r, though the difference in bias magnitude is small (e.g., .0007, when N = 10 and ρ = .9).

However, a more general and significant problem with this approach is that Hotelling's formulae (Equations 4 and 5; note that $n = N - 1$) are asymptotic; they reflect expected bias only when an infinite number of correlations are averaged.

$$E(r - \rho) = -\frac{\rho(1-\rho^2)}{2n} + \frac{\rho - 10\rho^3 + 9\rho^5}{8n^2} + \frac{\rho + 41\rho^3 - 117\rho^5 + 75\rho^7}{16n^3} + \dots \quad (4)$$

$$E(z' - \zeta) = \frac{\rho}{2n} + \frac{5\rho + \rho^3}{8n^2} + \frac{11\rho + 2\rho^3 + 3\rho^5}{16n^3} + \dots \quad (5)$$

Under these conditions, there is usually a small advantage associated with averaging rs directly rather than using z'; however, such conditions never exist in empirical settings. For instance, when ρ is estimated within meta-analytic studies, the number of studies is never infinite and, other than in the area of selection test validation, is often small (i.e., less than 20 effects per meta-analytic distribution). In Cornwell's (1987) survey of 625 meta-analyses in a variety of journals, the modal number of studies used in a meta-analysis was 6 and the median was 12. After Cornwell's review, the use of small numbers of studies for computing mean effect sizes in meta-analysis has continued (Driskell, Willis, & Cooper, 1992; Finkelstein, Burke, & Raju, 1995; Fried, 1991; Mitra, Jenkins, & Gupta, 1992; Narby, Cutler, & Moran, 1993).

Empirical evidence from Silver and Dunlap (1987) indicates that when a finite number of correlations from a matrix of intercorrelations are averaged, bias in average $r_{z'}$ is smaller than bias in average r in magnitude. In Silver and Dunlap's study, the number of correlations averaged was not varied and the correlations came from an intercorrelation matrix; it has not yet been determined empirically how the two procedures compare when different numbers of correlations are averaged or when independent correlations are averaged, as in the case of meta-analysis.

A final argument presented by proponents of using average r rather than average $r_{z'}$ involves rounding of reported results. Often, meta-analyses are performed on correlation coefficients that have been rounded to two decimal places for reporting. If it is assumed that correlations are rounded to two decimal places and that $N \geq 40$, then expected bias in average r is always less than rounding error—that is, bias is less than .005, as calculated via Hotelling's (1953) formulae. In other words, when $N \geq 40$, expected bias in r has no effect on validities that appear in the literature, assuming that those validities have been rounded to two decimal places.

In comparison, Hunter and Schmidt (1990) noted that use of z' can lead to inflation of averages of ".06 or more under extreme circumstances" (p. 217). It is worth mentioning, however, that under the same extreme conditions, the expected bias in average r is also large: .05 or more. These figures also highlight the fact that for small sample sizes (i.e., small N, the extreme circumstances to which Hunter and Schmidt refer), bias in either validity estimate is expected to exceed .005 by a large margin and bias is no longer expected to be less than rounding error. Although large sample sizes are available to researchers in some areas of applied psychology and the social sciences, small samples are often all that is available. For instance, the growing popularity of meta-analysis in subfields such as social, health, and experimental psychology (cf. Cooper & Lemke, 1991) and in disciplines such as education, social policy analysis, and the medical sciences (cf. Cooper & Hedges, 1994)—areas in which researchers may not have access to large samples—indicates a need for identification of statistics that are minimally biased for small samples as well as large samples.

It should be noted that although Hunter and Schmidt's (1950) decision to switch from z'- to r-based meta-analysis was predicated in large part on unpublished findings (i.e., Hunter et al., 1988), these findings have been presented in a widely cited text (Hunter & Schmidt). According to the *Social Sciences Citation Index* (Institute for Scientific Information, 1990–1997), Hunter and Schmidt's text was cited over 360 times from its publication through mid-1997—an average of over 45 citations per year. The conclusions drawn in that text regarding the use of z' may thus influence a large proportion of the scientific community that uses meta-analysis. Given the significant impact of Hunter and Schmidt's text on meta-analytic procedures, and considering the lack of peer-reviewed evidence supporting their claim that using Fisher's z' increases bias, one must ask whether Hunter and Schmidt's suggestion to abandon Fisher's z' in favor of untransformed r was justified.

In the present study, we used Monte Carlo techniques to empirically evaluate the bias associated with the two methods for estimating population correlations. In Experiment 1, an extension of Silver and Dunlap's (1987) investigation, we averaged correlations from an intercorrelation matrix. Data from populations with correlations from 0.0 to 0.9 were generated for several conditions comprising sample sizes (Ns), with various numbers of correlations averaged. Bias was then evaluated for both average r and average $r_{z'}$. In Experiment 2, we compared the two types of correlation estimates when independent correlations were averaged. In both experiments, varying the number of correlations averaged allowed evaluation of bias under more realistic conditions than those assumed under the expected values formulae (i.e., using an infinite number of studies).

EXPERIMENT 1: AVERAGING CORRELATION COEFFICIENTS FROM A MATRIX

Method

We used CHFAC and RNMVN subroutines from the International Mathematics Subroutine Library for FORTRAN (IMSL, 1987) to generate correlated data. Population parameters were entered in the form of uniform intercorrelation matrices for each value of ρ; values ranged from 0.0 to 0.9 in increments of 0.1. Data were generated for 3×3 through 10×10 matrices. We used Equation 6 (in which k = the number of intercorrelated measures) to compute the number of unique off-diagonal correlations (j) in an intercorrelation matrix.

$$j = \frac{k(k-1)}{2} \tag{6}$$

Thus, 3, 6, 10, 15, 21, 28, 36, and 45 correlations were averaged for 3×3 through 10×10 matrices, respectively. We also varied the number of score pairs (N) making up a single correlation. Specifically, we used values of 10, 20, 30, 40, and 50.

We computed intercorrelation matrices for the data generated by the subroutines. The correlations were then averaged both untransformed and after being converted to z's (Equation 2). Each average z' was then back-converted to r (Equation 3). This process was repeated 10,000 times for each of the 400 conditions (10 ρs $\times$ 8 ks $\times$ 5 Ns). We summed average rs and average $r_{z'}$s across iterations and divided the grand sum by the number of iterations.

Results

Given the large number of conditions in this study (10 ρs $\times$ 8 ks $\times$ 5 Ns = 400 conditions), one can view only a portion of the results in any one figure. We have tried to use figures that best illustrate our most important findings.

Under most conditions, averaging using z's produced less biased estimates of ρ than did averaging untransformed rs. The data for estimates obtained from small samples ($N = 10$) are shown in Figure 1. The most important feature of Figure 1 is that the bias in average $r_{z'}$ (dotted curves labeled "z") tended to be closer to zero (solid horizontal line) than did the bias in average r (solid curves labeled "r"), supporting the conclusion that average $r_{z'}$ is a less biased statistic than average r. Although the magnitudes of the bias for both methods was smaller when N was large, the pattern of results was consistent across the larger values of N (not shown).

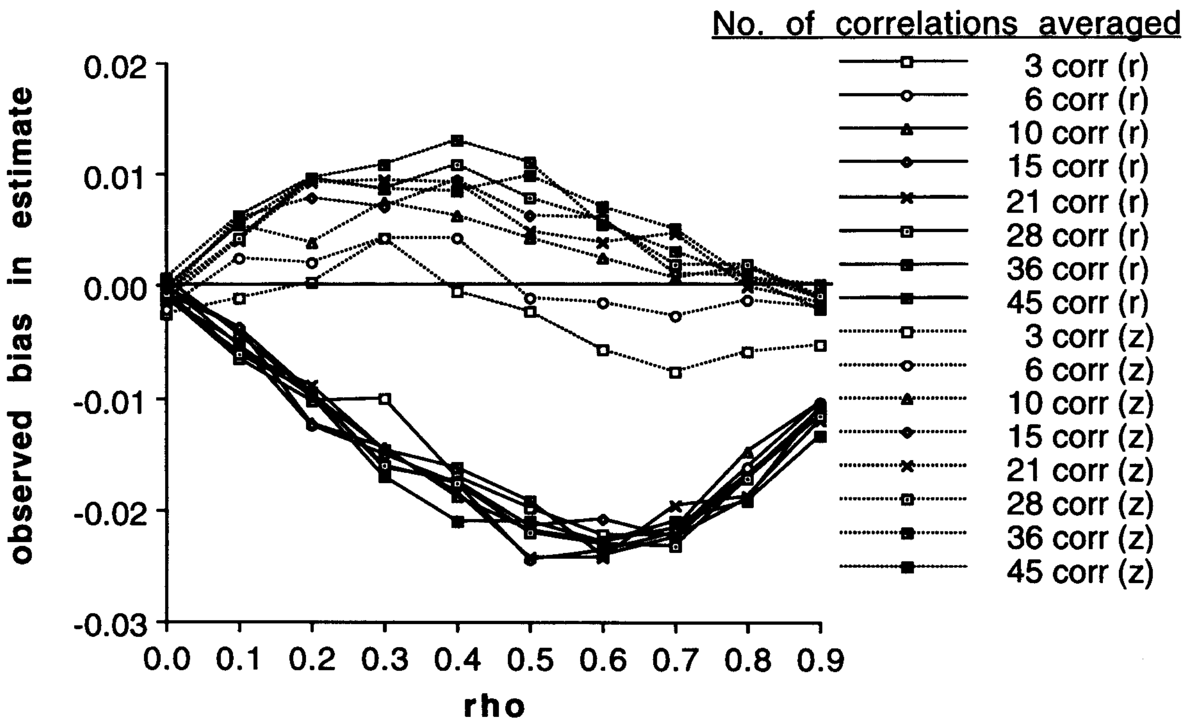

FIGURE 1. Observed bias in average r and average $r_{z'}$ when various numbers of correlations from a matrix were averaged. Positive numbers indicate positively biased estimates of ρ and vice versa. For these data, $N = 10$.

Another interesting feature of the data is that, although bias in average $r_{z'}$ depended on the number of correlations averaged (k), bias in average r was independent of k (see Figure 1). That is, there was a negative shift in the curves showing bias in average $r_{z'}$ as k increased. Yet we found no corresponding shift in average r's bias across levels of k. Surprisingly, average $r_{z'}$ actually became slightly negatively biased when very few correlations were averaged and when ρ was .5 or greater. However, even under those conditions, the magnitude of the bias was smaller for average $r_{z'}$ than for average r.

Across the sample sizes examined, the observed bias in average $r_{z'}$ was smaller than the bias in average r in magnitude. In Figure 2, data are shown for conditions in which a relatively large number of correlations were averaged ($k = 45$). For each sample size, the dotted curves representing bias in average $r_{z'}$ are closer to the zero-bias line than are the solid curves showing bias in average r. Thus, even when as many as 45 correlations were averaged, bias in average $r_{z'}$ was smaller in magnitude than was bias in average r.

In summary, when we averaged multiple rs (or z's) from an intercorrelation matrix, the bias in average $r_{z'}$ was smaller in magnitude than was the bias in average r. With increasing N, the $r_{z'}$ advantage became smaller, however; even when both k and N were relatively large (i.e., the curves in Figure 2 where $N = 50$), there was a small advantage in using average $r_{z'}$. For no combination of N and k was average r consistently less biased than average $r_{z'}$.

Discussion

Our findings supported Silver and Dunlap's (1987) conclusion that using z' to average correlations from a matrix produces less bias than averaging untransformed rs. Bias in average r was unaffected by the size of the matrix, regardless of the sample size, but bias in average $r_{z'}$ became increasingly positive as the number of correlations averaged increased. Nevertheless, when data came from a matrix, bias in average $r_{z'}$ was smaller than bias in average r across conditions. Thus, to obtain a less biased estimate of the population value underlying correlations in a matrix, Fisher's z transformation, rather than untransformed rs, should be used to average correlations, regardless of sample size and the number of correlations in the matrix.

EXPERIMENT 2: AVERAGING INDEPENDENT CORRELATION COEFFICIENTS

Method

We used the random normal number generator RNNOF(0) from the International Mathematics Subroutine Library for FORTRAN (IMSL, 1987) to generate

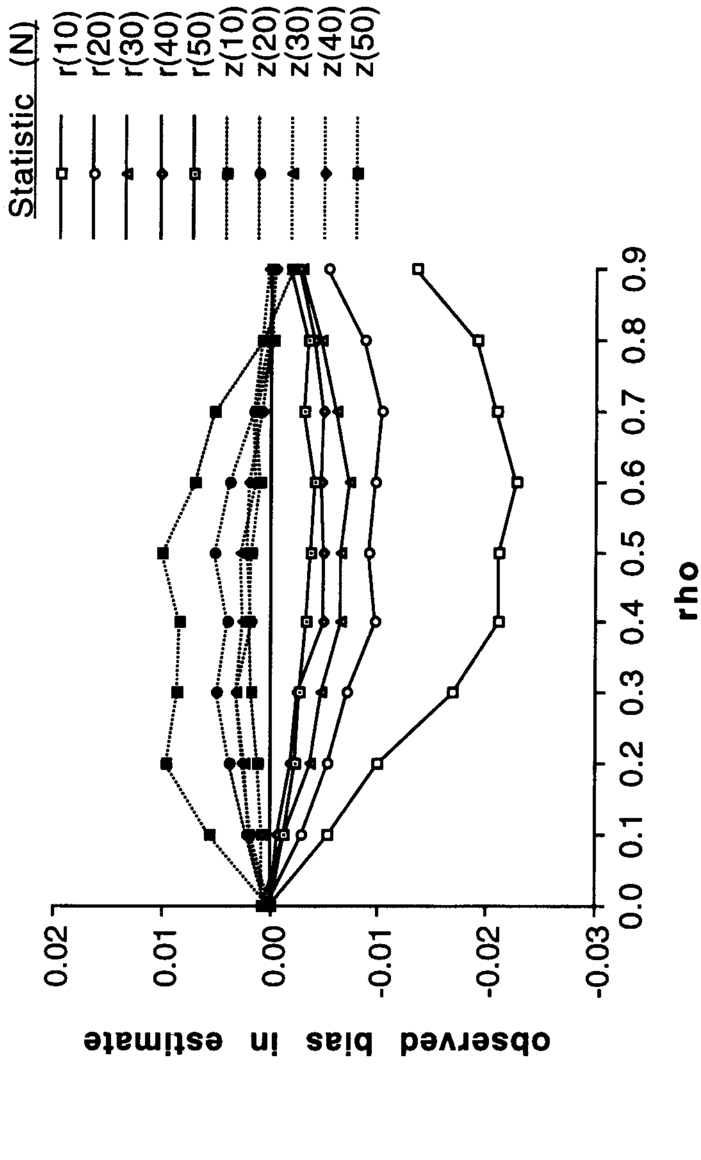

FIGURE 2. Observed bias in average *r* and average $r_{z'}$ for various sample sizes when correlations from a matrix were averaged. Positive numbers indicate positively biased estimates of ρ and vice versa. For these data, 45 correlations from a matrix were averaged.

data. We used the following FORTRAN statements to generate X and Y datasets with a population correlation of ρ.

```
COMM = RNNOF(0)
ERR1 = RNNOF(0)
ERR2 = RNNOF(0)
x = COMM * SQRT(ρ) + ERR1 * SQRT(1 – ρ)
y = COMM * SQRT(ρ) + ERR2 * SQRT(1 – ρ).
```

Values of ρ ranged from 0.0 to 0.9 in increments of 0.1. As in Experiment 1, we averaged 3, 6, 10, 15, 21, 28, 36, and 45 correlations. The number of scores (N) making up a single correlation also paralleled the values in Experiment 1 (N = 10, 20, 30, 40, or 50).

The correlations were averaged both untransformed and after being converted to z' (Equation 2). Each average z' was then back-converted to r (Equation 3). We repeated this process 10,000 times for each of the 400 conditions used in Experiment 1 (10 ρs × 8 number of correlations × 5 Ns). We summed average rs and average $r_{z'}$s across iterations and divided the grand sum by the number of iterations.

Results

Once again, because of the large number of conditions, we cannot show all of our data; the results shown in the figures represent our most important findings. Figure 3 (Experiment 2) corresponds to Figure 1 (Experiment 1). As we found for correlations from a matrix, when independent correlations were averaged, bias in average $r_{z'}$ (thin dotted curves labeled "z" in the figure) tended to be smaller (closer to the zero-bias line) than bias in average r (solid curves labeled "r"). This finding supports the conclusion that, when independent correlations are averaged, average $r_{z'}$ is a less biased statistic than average r.

As in Experiment 1, we found a negative shift in bias for average $r_{z'}$ as k decreased (see Figure 3; cf. Figure 1). And, once again, no corresponding shift was seen for the average r curves. Thus, one can conclude that average r for independent correlations is unaffected by the number of correlations averaged, whereas average $r_{z'}$ is less biased when few, rather than many, correlations are averaged.

Another interesting and important finding is related to the expected values for bias obtained via Hotelling's analytical formulae (Equations 4 and 5; see Figure 3). In Figure 3, the amount of bias in each statistic predicted by Hotelling's formulae is indicated by the large open circles and squares and heavy dashed curves labeled "r(analyt.)" and "z(analyt.)," respectively. For each level of k, bias in average r was predicted quite well analytically via Equation 4, as seen in the clustering of the average r curves around the analytical estimate of bias. However, that was not the case for average $r_{z'}$. Equation 5 provided an accurate prediction of bias in average $r_{z'}$ only when a large number of correlations were aver-

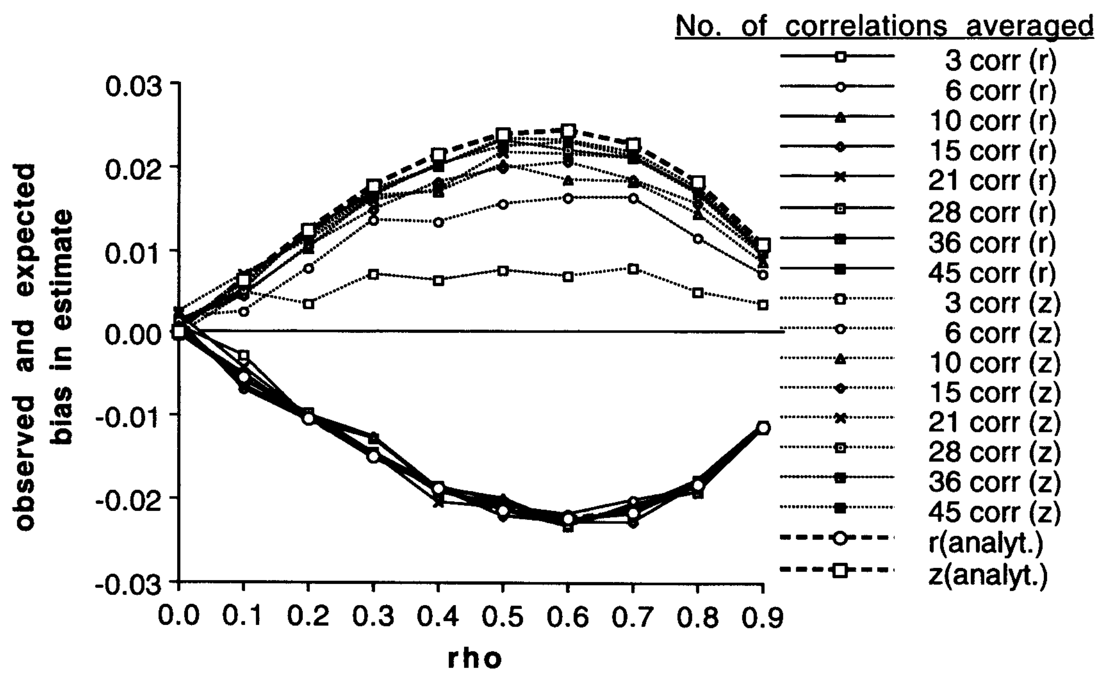

FIGURE 3. Observed bias in average r and average $r_{z'}$ when various numbers of independent correlations were averaged. The large open circles and squares labeled "r(analyt.)" and "z(analyt.)" represent expected bias based on Hotelling's (1953) asymptotic formulae. Positive numbers indicate positively biased estimates of ρ and vice versa. For these data, $N = 10$.

aged. When a small number of correlations were averaged, Hotelling's (1953) formula for expected bias in average $r_{z'}$ overestimated bias by a large margin. See Figure 3, in which there are large differences between the "z(analyt.)" curve and the curves showing observed bias in average $r_{z'}$ when $k = 3$ or 6 but not when $k = 36$ or 45. Thus, one can conclude that, although bias in average r conforms quite well to analytically derived expected values, bias in average $r_{z'}$ approaches the expected value asymptotically as the number of correlations averaged rises. When a small number of correlations are averaged, Hotelling's formula is inaccurate, with average $r_{z'}$ being considerably less biased than predicted by the analytical formula.

Data for the modal meta-analytic case in which six studies were averaged are shown in Figure 4. The difference in magnitude of observed bias for the two averaging techniques is plotted for each value of ρ. We obtained the differences by subtracting the magnitude (i.e., absolute value) of the bias in average $r_{z'}$ from the magnitude of the bias in average r. Values greater than zero therefore indicate less bias in average $r_{z'}$. In the large proportion of positive values, average $r_{z'}$ was usually less biased than average r (see Figure 4). Again, that was true when relatively few correlations were averaged, resulting in observed bias that was smaller than expected bias. As one can see in Figure 3, when 36 or more correlations were averaged, bias in average $r_{z'}$ was predicted fairly well by the analytical formula. As Hotelling's (1953) formulae predict, when k was large, bias in the two estimates was opposite in direction and approximately equal in magnitude (see Figure 5; note the high degree of symmetry about the zero-bias line, which shows data for the conditions in which 45 correlations were averaged).

Discussion

The present results do not support the conclusion of Schmidt et al. (1988) and Hunter and Schmidt (1990) that average r is a less biased estimate of ρ than is average $r_{z'}$. Although expected bias in average r is smaller than expected bias in average $r_{z'}$, the analytical formulae upon which these values are based assume that the number of correlations averaged is infinite. Expected bias formulae thus do not reflect the observed effect of number of correlations averaged on bias in the $r_{z'}$ statistic. Expected values accurately predict the behavior of average $r_{z'}$ only when a large number of correlations are averaged. When a small number of correlations are averaged, validities estimated using average $r_{z'}$ are considerably less biased than expected values predict. Thus, the present results indicate that under conditions common in meta-analysis, average $r_{z'}$ provides a less biased validity estimate than does average r. It must therefore be concluded that for independent correlations, average $r_{z'}$ is generally a better estimate of ρ than is average r.

It must be acknowledged that the superiority of average $r_{z'}$ in estimating population correlations does not necessarily constitute evidence that average $r_{z'}$ should be used in meta-analyses. A detailed discussion of the relative merits of

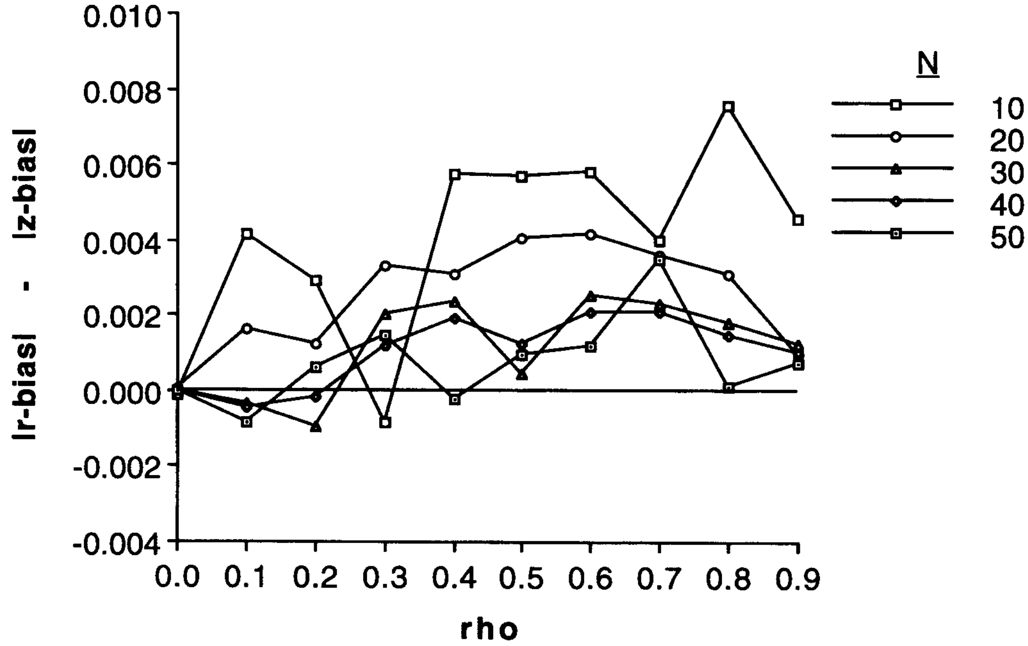

FIGURE 4. Relative accuracy of average r and average $r_{z'}$ when independent correlations were averaged. Positive numbers indicate an average $r_{z'}$ advantage; negative numbers indicate an average r advantage. For these data, six correlations were averaged.

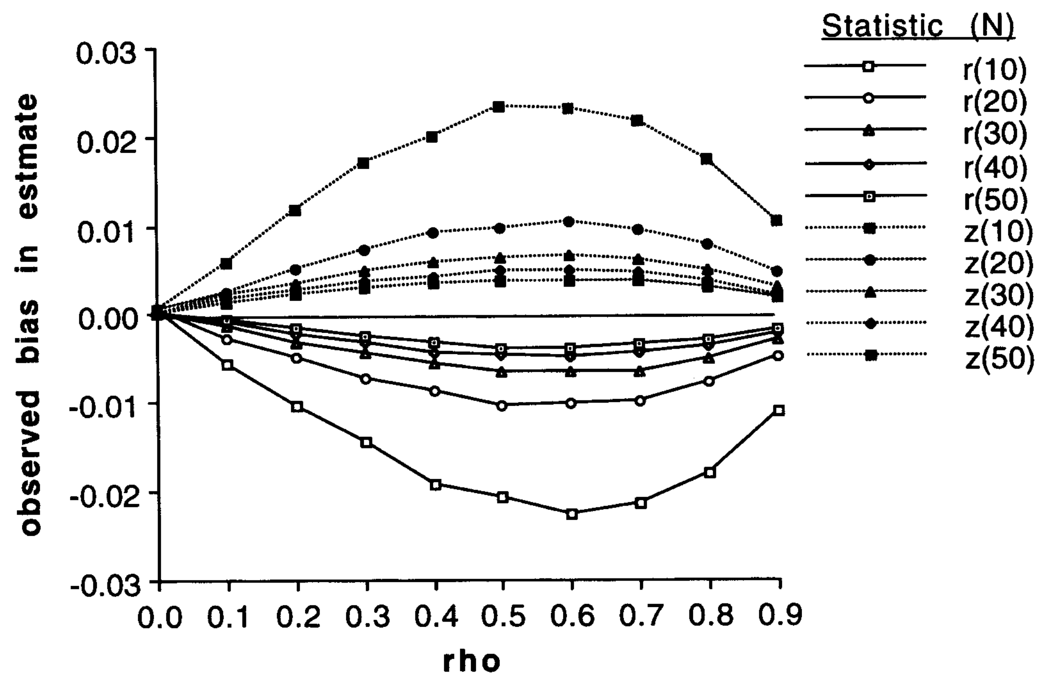

FIGURE 5. Observed bias in average r and average $r_{z'}$ for various sample sizes when independent correlations were averaged. Positive numbers indicate positively biased estimates of ρ and vice versa. For these data, 45 correlations were averaged.

average r and average $r_{z'}$ is beyond the scope of this article. However, James et al. (1986) suggested that average $r_{z'}$ should replace average r as a matter of course in meta-analysis. Hunter and Schmidt (1990) and Schmidt et al. (1988) responded with arguments in favor of average r.

GENERAL DISCUSSION

The Importance of Small Improvements in Accuracy

All else being equal, a less biased statistic is preferable. Nevertheless, there are conditions under which bias may be accepted. For example, one would be justified in ignoring a difference in bias when that difference is less than rounding error; if bias never influences results reported in the literature, no practical benefit results from correcting the bias. As the present data show, the same does not apply to correlation coefficients. Another possibility is that some desired procedure may be available for a more biased statistic but not for a less biased one. Fortunately, averaging through the use of z' does not preclude the use of procedures that are available only in the r metric (e.g., correcting unreliability in measures). The actual averaging of the coefficients is all that need be done to reduce bias in the z' metric. Furthermore, averaging in the z' metric does not require prohibitively complex mathematical procedures. Thus, there are no obvious reasons not to take advantage of the reduction of bias that results from averaging in the z' metric.

The practical value of small improvements in accuracy can also be seen for procedures for translating correlation coefficients into utility estimates (i.e., estimates of the practical gain or improvement related to a behavioral intervention). It has long been recognized that the gain in economic utility or improvement associated with a behavioral intervention (e.g., training program, selection procedure) is a direct function of the correlation or validity coefficient (Brogden, 1946; Cronbach & Gleser, 1965; Raju, Burke, & Normand, 1990). Furthermore, it is well known that the percentage improvement in performance or productivity resulting from the use of a valid selection procedure is directly proportional to the validity coefficient itself (Jarrett, 1948). Together, these decision-theoretic procedures for utility analysis indicate that small correlation coefficients (or small differences between correlation coefficients' magnitudes) can have significant practical value—a point demonstrated empirically by Burke and Doran (1989) within the context of estimating the economic utility of small differences in validity (correlation) coefficients for alternative personnel selection procedures.

Competing Estimates of ρ

Combined estimates of ρ other than average r and average $r_{z'}$ were not evaluated here. Some of these are obtained simply by adding values to r (Hunter &

Schmidt, 1990; Olkin & Pratt, 1958) or subtracting values from z' (Hotelling, 1953) to compensate for expected bias in the statistics. Others are obtained via maximum likelihood estimation using untransformed *r*s (Pearson, 1933; Viana, 1982).

In the present study, we were concerned only with r and $r_{z'}$ for several reasons. First, the bulk of the popular meta-analytic literature has been dedicated to procedures involving r and z' (e.g., Hunter & Schmidt, 1990; Raju, Burke, Normand, & Langlois, 1991). Second, although several researchers have used empirical studies to compare combined correlation statistics (Donner & Rosner, 1980; Silver & Dunlap, 1987; Viana, 1982), none have compared observed bias in average r with that of average $r_{z'}$. Finally, we are aware of no studies in which expected values of r and z' were compared with observed values. In the present study, we sought to examine the accuracy of analytical estimates of the statistics and found that the expected value of z' did not adequately predict the behavior of the statistic.

Variability in the Statistics

The reader has likely noticed that we have not mentioned variability in the statistics. This omission is not an oversight. The primary benefit that would appear to be associated with a less variable statistic is narrower confidence intervals. However, such confidence intervals would be constructed in different metrics, depending on which estimate was used. Confidence intervals around estimates of ρ would be constructed in the r metric, and confidence intervals around estimates of ζ (zeta, the parameter estimated by z') would be constructed in the z' metric. Because different standard error estimates are used to construct confidence intervals in the two metrics, an observed difference between the variability of r and $r_{z'}$ in the r metric would not necessarily indicate differences in confidence interval width. Furthermore, there is no proven method by which variance in the z' metric can be converted to the r metric for comparison.

The only adequate means by which to compare the effective difference in variance in the two average correlation statistics is therefore to compare alpha and power in a paradigm that tests null hypothesis significance. A comparison of this type would require an evaluation of the field of potential variance estimation procedures for average r and average z' (e.g., Donner & Rosner, 1980; Finkelstein, et al., 1995; Fisher, 1958; Hunter & Schmidt, 1990). Such an evaluation lies beyond the scope of the present study. However, in one such study, no systematic difference between the two statistics was found (Donner & Rosner, 1980), showing that, for the purpose of constructing confidence intervals, there is no meaningful difference in variance between the two estimates of ρ that we have considered. In addition, the first author, in an independent simulation study, found that the average observed standard errors of the aggregate r and $r_{z'}$ statistics were approximately equal (.055 and .058, respectively; *SE*s were averaged across ks = 5, 10, 30, 90; Ns = 5, 10, 50, 100; ρs = 0, .1, .3, .5, .7, .9).

CONCLUSIONS

When a very large number of correlations are averaged, it is analytically reasonable to expect a minute advantage in bias associated with using untransformed r rather than Fisher's (1958) z' transformation (difference in magnitude of expected bias is less than .0028 for $N \geq 10$). However, the present results indicate that the behavior of average-correlation statistics often does not conform to such expectations. When correlations come from a matrix, there is a consistent advantage associated with using z'. Across sample sizes and numbers of correlations averaged, bias in average $r_{z'}$ is smaller than bias in average r.

When independent correlations are averaged, bias in $r_{z'}$ varies considerably, depending on the number of correlations averaged. If a large number of correlations are averaged, the expectation of a very slight advantage associated with using average r is realized. However, when a few correlations are averaged, as is common in meta-analysis, there is a clear advantage in using average $r_{z'}$. Furthermore, the r advantage when many correlations are averaged is virtually invisible compared with the bias discrepancy in favor of average $r_{z'}$ when few correlations are averaged. Thus, considering the frequent use of small numbers of studies in meta-analyses and the relatively large potential advantage of using average $r_{z'}$ we recommend the use of average $r_{z'}$ to estimate population correlations.

REFERENCES

Brogden, H. E. (1946). On the interpretation of the correlation coefficient as a measure of predictive efficiency. *Journal of Educational Psychology, 37,* 64–76.

Burke, M. J., & Doran, L. I. (1989). A note on the economic utility of generalized validity coefficients in personnel selection. *Journal of Applied Psychology, 74,* 171–175.

Carter, R. C., Kennedy, R. S., & Bittner, A. C., Jr. (1981). Grammatical reasoning: A stable performance yardstick. *Human Factors, 23,* 587–591.

Cooper, H. M., & Hedges, L. V. (Eds.). (1994). *The handbook of research synthesis.* New York: Russell Sage Foundation.

Cooper, H. M., & Lemke, K. M. (1991). On the role of meta-analysis in personality and social psychology. *Personality and Social Psychology Bulletin, 17,* 245–251.

Cornwell, J. M. (1987). *Content analysis of meta-analytic studies from I/O psychology.* Unpublished dissertation, University of Tennessee, Knoxville.

Cronbach, L. J., & Gleser, G. C. (1965). *Psychological tests and personnel decisions.* Champaign: University of Illinois Press.

Donner, A., & Rosner, B. (1980). On inferences concerning a common correlation coefficient. *Applied Statistics, 29,* 69–76.

Driskell, J., Willis, R., & Copper, C. (1992). Effect of overlearning on retention. *Journal of Applied Psychology, 77,* 615–622.

Finkelstein, L. M., Burke, M. J., & Raju, N. S. (1995). Age discrimination in simulated employment contexts: An integrative analysis. *Journal of Applied Psychology, 80,* 652–663.

Fisher, R. A. (1958). *Statistical methods for research workers* (13th ed.). Edinburgh: Oliver & Boyd.

Fried, Y. (1991). Meta-analytic comparison of the job diagnostic survey and job characteristics inventory as correlates of work satisfaction and performance. *Journal of*

Applied Psychology, 76, 690–697.
Hotelling, H. (1953). New light on the correlation coefficient and its transforms. *Journal of the Royal Statistical Society, B15,* 193–225.
Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings.* Newbury Park, CA: Sage.
Hunter, J. E., Schmidt, F. L., & Coggin, T. D. (1988). *Meta-analysis of correlation: The issue of bias and misconceptions about the Fisher* z *transformation.* Unpublished manuscript, Department of Psychology, Michigan State University.
IMSL, Inc. (1987). *STAT/LIBRARY user's manual, version 1.0.* Houston: Author.
Institute for Scientific Information. (1990–1997). *Social Sciences Citation Index.* Philadelphia: Author.
James, L. R., Demaree, R. G., & Mulaik, S. A. (1986). A note on validity generalization procedures. *Journal of Applied Psychology, 71,* 440–450.
Jarrett, F. F. (1948). Per cent increase in output of selected personnel as an index of efficiency. *Journal of Applied Psychology, 32,* 135–145.
Mitra, A., Jenkins, G., & Gupta, N. (1992). A meta-analytic review of the relationship between absence and turnover. *Journal of Applied Psychology, 77,* 879–889.
Narby, D., Cutler, B., & Moran, G. (1993). A meta-analysis of the association between authoritarianism and jurors' perceptions of defendant culpability. *Journal of Applied Psychology, 78,* 34–42.
Olkin, I. O., & Pratt, J. W. (1958). Unbiased estimation of certain coefficients. *Annals of Mathematical Statistics, 29,* 201–211.
Pearson, K. (1933). On a method of determining whether a sample of size *n* supposed to have been drawn from a parent population having a known probability integral has probably been drawn at random. *Biometrika, 25,* 379–410.
Raju, N. S., Burke, M. J., & Normand, J. (1990). A new approach for utility analysis. *Journal of Applied Psychology, 75,* 3–12.
Raju, N. S., Burke, M. J., Normand, J., & Langlois, G. M. (1991). A new meta-analytic approach. *Journal of Applied Psychology, 76,* 432–446.
Schmidt, F. L., Hunter, J. E., & Raju, N. S. (1988). Validity generalization and situational specificity: A second look at the 75% rule and Fisher's *z* transformation. *Journal of Applied Psychology, 73,* 665–672.
Silver, N. C., & Dunlap, W. P. (1987). Averaging correlation coefficients: Should Fisher's *z* transformation be used? *Journal of Applied Psychology, 72,* 146–148.
Viana, M. A. (1982). Combined estimators for the correlation coefficient. *Communications in Statistics—Theory and Methods, 11,* 1483–1504.

Manuscript received March 5, 1998
Revision accepted for publication July 9, 1998