Correlation Coefficients: Mean Bias and Confidence Interval Distortions

Richard L. Gorsuch   **Curtis S. Lehmann**

Fuller Theological Seminary

Non-zero correlation coefficients have non-normal distributions, affecting both means and standard deviations. Previous research suggests that *z* transformation may effectively correct mean bias for *N*'s less than 30. In this study, simulations with small (20 and 30) and large (50 and 100) *N*'s found that mean bias adjustments for larger *N*'s are seldom needed. However, *z* transformations improved confidence intervals even for $N = 100$. The improvement was not in the estimated standard errors so much as in the asymmetrical CI's estimates based upon the *z* transformation. The resulting observed probabilities were generally accurate to within 1 point in the first non-zero digit. These issues are an order of magnitude less important for accuracy than design issues influencing the accuracy of the results, such as reliability, restriction of range, and *N*.

The observed correlation coefficient, *r,* is known to slightly under estimate the population correlation, ρ, because the distribution of *r* is asymmetrical, being skewed toward zero (Fisher, 1921). Thus, correlations that are averaged together will underestimate — in the sense of being closer to zero whether positive or negative — the population correlation. The skewed distribution of *r* also affects its standard error, *SE r,* as the standard error is distorted by the skew, as deviations above and below the correlation will be unequal. Both of these conditions affect the confidence intervals, which should be asymmetrical and based on a non-distorted estimate of the standard error. Fisher (1915, 1921, 1924) first demonstrated these two problems, noted they are greatest in small samples ($N < 30$), and suggested solutions. He also noted one other oddity: the greatest underestimate of the ρ is not, as one might expect, with high absolute values but actually is worse for mid-range correlations.

Figure 1 and 2 show two distributions, one for a population correlation (ρ) of .5 with an *N* of 20 where the problem is expected to be clear (Figure 1) and another for a ρ of .2 and $N = 100$, where the effects are much less (Figure 2). Note that the modal and median correlation encountered by investigators is considerably above the population ρ for the $N = 20$. So while the average correlation will be underestimated when averaging across studies, the average investigator will observe a correlation which overestimates the population value, a point that has not been widely discussed.
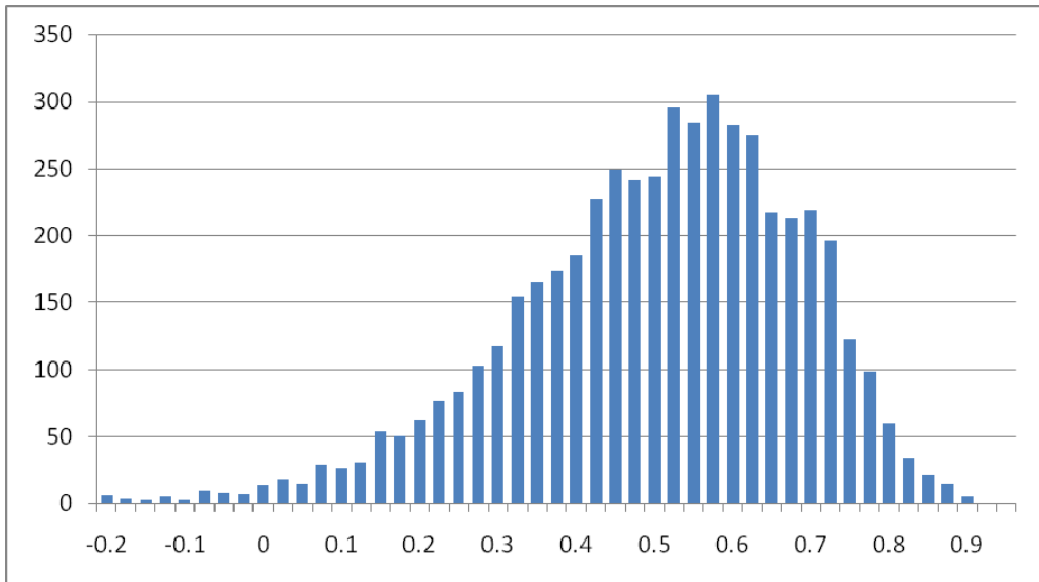
Figure 1: Histogram of Observed r with a Sample Size of 20 and $\rho = .5$
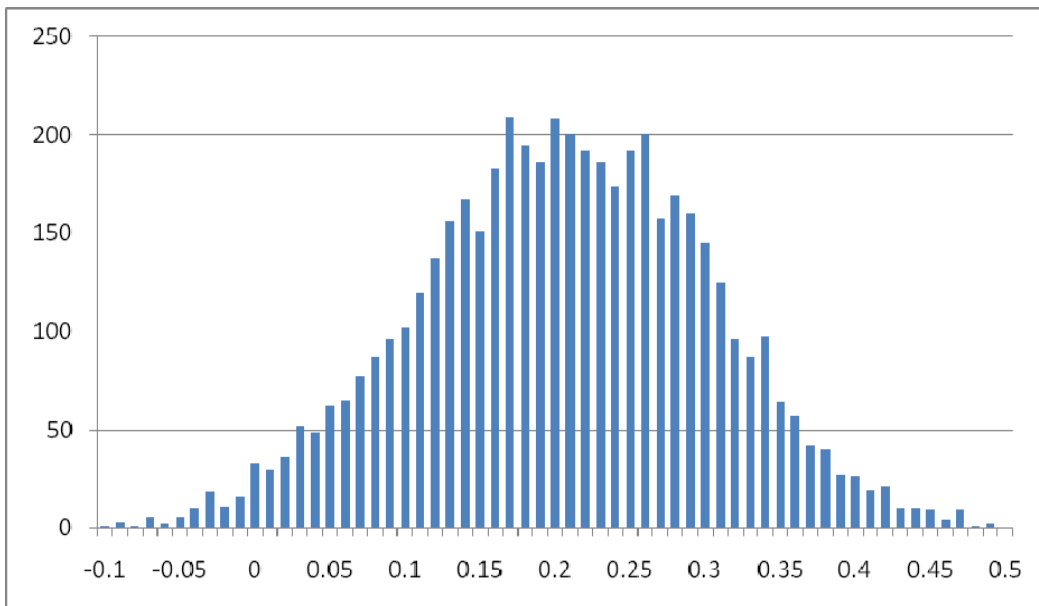
Figure 2: Histogram of Observed r with a Sample Size of 100 and $\rho = .2$

Mean bias (underestimation)

The Figures illustrate the mean bias that occurs due to the distribution of observed correlations. Fisher (1915, 1924) considered this problem and introduced two formulas for correcting r's so that means will most closely approximate ρ. The first was so complex it has not been used. Fisher felt

the second one was more useful, and labeled it a "*z* transformation".

Fisher developed the *z* transformation to normalize the distribution of the correlation coefficients, allowing average correlations and standard deviations to be calculated more accurately. This transformation is

$$z = .5 * (\log(1 + r) - \log(1 - r)) \tag{1}$$

where *r* is the observed correlation coefficient. The difference between *r* and *z* is trivial below .25 – within rounding error to two decimal places — but increases with the absolute value of the correlation. For example, a .5 correlation gives a *z* of .55 and a .8 correlation gives a *z* of 1.10 Comparing these to the figures illustrates how *z* stretches out the upper range to provide a more normal distribution.

To average correlations, each observed correlation is transformed to its *z*, the *z*'s are averaged, and the average *z* is transformed back to a correlation by the following formula:

$$r = (\exp(2 * z) - 1)/\exp(2 * z) + 1) \tag{2}$$

The evaluation of Fisher's *z* has always dealt with small samples. Fisher used an *N* of 18 for an example. Others have tested *z* for small samples also. For example, Meng, Rosenthal, and Rubin (1992) used $N = 15$, Strube (1988) 20, and Silver and Dunlap (1987) report that the correction is less than rounding error with *N*'s of 30 to 40. These studies did find that *z* overestimated the average correlation, with the over-estimate being greatest with a population correlation in the mid range but that the bias of *z* was less than the negative bias of *r*. As the impact decreases with *N*, it has been assumed that the impact is trivial for larger *N*'s. However, there are few tests of this proposed effect of *N*.

Olkin and Pratt (1958) have proposed another equation to correct for mean bias to improve on *z*:

$$G(r) = r (1 + ((1 - r^2)/2 * (N - 3)) \tag{3}$$

where G(*r*) is the corrected estimate of the correlation. They show that an observed correlation of .5 becomes a corrected correlation of .521 for $N = 20$ and .513 for $N = 30$, but do not give the magnitude of the correction for larger sample sizes. In their analyses, Zimmenman, Zumbo, and Williams (2003) reported that this correction was only better than Fisher's with *N* less than 20.

The discussion above and analyses below are applicable to correlations which are random samples from the same population. Another variable influencing the results of averaging correlations is the number of correlations in the set being averaged. These issues have been addressed

(e.g., Alexander, Scozzaro, & Borodkin, 1989; Callender & Osburn, 1988; Corey, Dunlap, & Burke, 1998; Hafdahl, 2010; Hafdahl & Williams, 2009; Hunter & Schmidt, 2004; Silver & Dunlap, 1987) because of the importance for meta-analysis. But those issues are not addressed here. The concern of this study is for correlations drawn from the same population.

Confidence interval distortion

Confidence intervals give the range within which a specified percent of observed correlations would occur for the N used with random samples from the same population if the observed correlation is actually the population correlation. They are computed by, first, finding the standard error (SE) of the statistic. SE's are standard deviations and so 66% of the values would, with normal distributions and large Ns, occur between the lower bound – found by subtracting the standard error from the statistic – and the upper bound – found by adding the SE to the statistic. If other confidence intervals are desired, then the SE is multiplied by the appropriate value from a normal curve table to give the desired percent. Applied to a correlation the confidence intervals are:

$$CI = r - a * SE\,r \text{ to } r + a * SE\,r \tag{4}$$

where a is the distance needed in z score units for the desired percent within half of the bounded interval. The confidence intervals are usually based on an estimate of the SE for the statistic. The lower and upper intervals derive their meaning from the assumption that the SE describes a normal distribution so that probabilities based on the normal curve (and its derivatives such as t) can apply.

Figures 1 and 2 show that the skew of the distributions of non-zero correlations also affects the $SE\,r$ (which in this case is the SD of the distribution of r's). Except for a ρ of zero, there are more extreme correlations on the side towards 0.0. This distorts confidence intervals distributions as compared to a normal curve. One such distortion is that the lower and upper CI's are not equally distant from the mean correlation. Instead the lower 2.5% bound is a greater absolute distance from the mean than the upper 2.5% bound.

A second distorting effect arises from using this formula to estimate the standard error:

$$SE\,r = (1 - r^2)/\sqrt{(N - 1)} \tag{5}$$

Using this formula gives the standard deviation of the errors for the observed r, assuming $r's$ are normally distributed. However, the

distribution of r is skewed and so the formula is an under estimate of the standard deviations for a given correlation. But how distorted is the standard error by this formula which assumes a normal distribution?

Fisher (1915, 1921) noted that the z transformation gives a statistic which has an identical *SE* regardless of the value of z or r. The formula is:

$$SE\,z = 1 / \sqrt{(N - 3)} \tag{6}$$

This differs from the *SE* of a correlation whose error decreases as ρ increases. Perhaps more important, it provides lower and upper CI's which, when translated back to correlations, are asymmetrical with the lower CI being farther from the correlation than the upper CI.

Using the z transformation for lower and upper bounds of a confidence interval requires the a for formula (4). Three values for a have been suggested. First, classical literature (e.g., Guilford, 1956) suggested using the value from the z table. Current literature commonly draws the value from a t table to correct for sample size. However, Cohen, Cohen, West, and Aiken (2003), suggest that a be set at a "good enough" approximation in some circumstances, given the possible problems from assuming r is ρ and in estimating *SE r*. For 95% CI's, the z table gives 1.96, the t table value varies, for example, from 2.045 for $N = 20$ and 1.98 for $N = 100$, and Cohen et al. recommend 2 for all N's, which is what the other values all round to if one decimal place accuracy is "good enough."

While the studies of the z transformation have examined the mean distortion, the *SE r* and resulting confidence intervals have been ignored. Both classical and current texts (e.g., Cohen, et al., 2003; Guilford, 1956) recommend Fisher's z transformation be used for CI's, but the need for this recommendation has seldom been investigated empirically. Neither have the several recommendations for a for this situation.[1]

Sample size

Fisher (1921) noted the problems from the non-normal distributions were most important with small N's. His major example (1921) used an N of 18. The simulation studies investigating these affects have investigated a variety of small sizes, typically 10 and 20. The conclusions from past research noted above directly apply to such small samples, and that Fisher's transformation can produce positively biased estimates for them. The math and the studies generally agree that the problems became less with larger N's. Strube (1988) did suggest that as the number of

---

[1] For the reliability coefficient alpha– which can be considered a correlation coefficient – other procedures for confidence intervals have been proposed which might be applied to correlations, but Romano, Kromrey, and Hibbard (2010) found all to give poorer confidence intervals than Fisher's z transformation.

correlations being averaged increased, *z* may become as biased as *r* but in the opposite direction. But there are no examples with *N*'s > 30 to provide data on the seriousness of the problems with larger *N*'s, and whether the seriousness is the same for the mean bias and CI distortion.

The current study

The purpose of the current study is to provide more information on the seriousness of the problems of mean bias and confidence interval distortion. We expect to replicate the mean bias magnitude, and replicate that adjustments by *z* transformation and by the Olkin and Pratt (1958) adjustment reduce mean bias with Olkin and Pratt (1958) being the better. *SE r* estimates can be calculated in several ways, including by formula from the observed correlation and by *z* transformation, and these are examined for their impact on confidence intervals with the expectation that *z* transformations will be more accurate. The analyses begin with typical *N*'s used previously (20 and 30) to replicate and extend previous findings. *N*'s of 50 and 100 are examined to evaluate how quickly the impact on mean *r* and confidence intervals drops. To provide perspective, the impacts of these problems are briefly compared to the impact of attenuation due to reliability or restriction of range, as well as the discrepancies in percentage tails for confidence intervals.

Method

To create a sufficient data set for this study, 5000 samples were constructed for each combination of ρ and *N*. For each case, X (raw score) and E (error) were created using the SPSS random number generator (Mersenne Twister) for *z* scores (normally distributed with Mean = 0 and *SD* = 1). These two variables were utilized to compute Y using the formula:

$$Y = \rho * X + \sqrt{(1 - \rho^2)} * E \tag{7}$$

where ρ is the population correlation coefficient. X and Y were then correlated.

The *N*'s selected were 20 and 30 to overlap with prior studies and 50 and 100 to provide data on larger samples. The selected population correlations were .2 for a level less effected by skew, .5 as the middle range of correlations which are most impacted (Zimmerman, Zumbo, & Williams, 2003), and .8 as a higher correlation with a greater skew.

The sample descriptive statistics were computed as the simple mean, median, standard deviation, skew, and kurtosis of each sample of correlations. The formulas given above were used to compute the other needed statistics.

Analyses and Results

Descriptive statistics of the correlation coefficients for each ρ and N combination are presented in Table 1. These follow the pattern as displayed in Figures 1 and 2. The displacements of the mean from the median are the expected result from the skew of each. The skew increases with ρ. but decreases with N.

Table 1
Descriptive Statistics of the Distribution of Observed r

	Mean	Median	*SD/SE*	Skew	Kurtosis
			$N = 20$		
$\rho = .2$	0.196	0.205	0.222	-0.271	-0.113
$\rho = .5$	0.490	0.512	0.180	-0.648	0.471
$\rho = .8$	0.791	0.806	0.091	-1.044	1.658
			$N = 30$		
$\rho = .2$	0.196	0.203	0.178	-0.237	-0.051
$\rho = .5$	0.492	0.506	0.143	-0.526	0.322
$\rho = .8$	0.795	0.806	0.070	-0.884	1.154
			$N = 50$		
$\rho = .2$	0.195	0.200	0.136	-0.205	-0.017
$\rho = .5$	0.494	0.502	0.110	-0.499	0.438
$\rho = .8$	0.798	0.804	0.053	-0.752	1.072
			$N = 100$		
$\rho = .2$	0.200	0.201	0.097	-0.107	-0.052
$\rho = .5$	0.497	0.501	0.077	-0.282	0.021
$\rho = .8$	0.799	0.802	0.036	-0.532	0.649

Kurtosis also varied, being negative with low correlations and positive with high correlations. This effect also decreased with increasing N. The past literature has not noted this shift which can also lead to poorer estimates of variability.

The difference between the mean and median suggests that the average investigator sees more values greater than the population value than lower than that value. If the mean and median were the same, 50% of the investigators would report a correlation above the ρ. But with a skewed distribution, more see a higher correlation. Table 2 gives the percent of investigators who would report a correlation above ρ. The worst case is with $N = 20$, when 57% would report a correlation above the population.5, and which drops to 53% for a large N.

Table 2
Discrepancies from 50% of the Percentage of Scores above Mean r and ρ

	N = 20	N = 30	N = 50	N = 100
		Mean r		
ρ = .2	1.88%	2.06%	1.16%	0.54%
ρ = .5	4.92%	3.72%	2.84%	2.26%
ρ = .8	6.74%	6.18%	4.68%	3.86%
		ρ		
ρ = .2	0.94%	0.88%	0.04%	0.46%
ρ = .5	2.82%	1.68%	0.62%	0.62%
ρ = .8	2.66%	3.20%	2.98%	2.80%

To examine mean bias of correlations as a function of ρ and N and the effect both of Fisher and of Olkin and Pratt transformations, each estimate of the population correlation – averaged r, Fisher z transformation averaged r, and Olkin and Pratt adjustment – was computed. Each estimate was then subtracted from the ρ. The discrepancies are in Table 3. At ρ = .2 the discrepancies for all three estimates are almost non-existent. The underestimate for averaged r is greatest for ρ = .5. As the N increases, the effect is less – approximately rounding error at N = 30 and even smaller for larger N's regardless of ρ.

Table 3
Discrepancies of Mean Estimated Correlations from Population Value

	N = 20	N = 30	N = 50	N = 100
		Unadjusted r		
ρ = .2	-0.004	-0.004	-0.005	0.000
ρ = .5	-0.010	-0.008	-0.006	-0.003
ρ = .8	-0.009	-0.005	-0.002	-0.001
		z transformation		
ρ = .2	0.006	0.002	-0.001	0.002
ρ = .5	0.012	0.006	0.002	0.001
ρ = .8	0.007	0.005	0.004	0.002
		G(r)		
ρ = .2	0.001	-0.001	-0.003	0.001
ρ = .5	0.000	-0.001	-0.002	-0.001
ρ = .8	-0.001	0.000	0.001	0.001

Note: G(r) is a bias correction formula developed by Olkin and Pratt (1958).

While the z transformation average is slightly closer than the average r, in Table 3 it is positively biased. It would not be unreasonable to average these two estimates but that is not necessary as the Olkin and Pratt adjusted r's of Table 3 are within absolute value of .003 of the true value.

To address the issue of confidence interval distortion, two tables were developed. Table 4 gives the discrepancies of the estimated $SE\ r$ from the observed average standard deviation of the correlations given in Table 1. The estimates computed were by the formula for the $SE\ r$, equation (5), computed by assuming the observed r is the population correlation. The second was computed from Fisher's z transformation using (6). The greatest discrepancy was for the formula, which consistently gave underestimates. These were up to -.01 with $N = 20$. Both could be considered accurate within rounding error with larger Ns although, overall, the z transformation gave more accurate SE's while having a slight but consistent positive bias.

Table 4
Discrepancies of Standard Errors from Sample Standard Deviation

	$N = 20$	$N = 30$	$N = 50$	$N = 100$
		SE_r		
$\rho = .2$	-0.013	-0.006	-0.001	-0.002
$\rho = .5$	-0.013	-0.005	-0.004	-0.001
$\rho = .8$	-0.007	-0.003	-0.001	0.000
		$SE_{z \to r}$		
$\rho = .2$	0.007	0.005	0.004	0.000
$\rho = .5$	0.003	0.002	0.000	0.000
$\rho = .8$	0.002	0.001	0.000	0.001

Note: The above discrepancies are calculated as the computed value minus the standard deviation. SE_r is the mean of the standard errors of the correlations, as calculated by the formula, $SE_r = (1-r^2)/\sqrt{(n-1)}$ for each correlation coefficient. $SE_{z \to r}$ was calculated by taking the mean $z \pm 1/\sqrt{x(n-3)}$. These endpoints were then back converted into r values and the range between the endpoints was divided by 2. As a characteristic of the z transformation, the upper and lower ranges were not equal; however, the small difference between the ranges had a trivial impact.

Table 5 was computed by computing the lower and upper 95% confidence intervals, and then computing the percent of samples which actually fell below or above the lower or upper boundary. These were computed with four procedures, that from the SE_r formula using a "good enough" value of 2 and also a value from the t table, and that computed by Fisher's z transformation, again for a value of 2 and for the t. The former is a symmetrical CI procedure, with the lower boundary being the same distance below the observed r as the upper boundary is above the r. Fisher's z transformation gives asymmetrical intervals. Table 5 contains

the discrepancies from the 2.5% that each should produce.

Table 5
Discrepancies from 2.5% of Observed Correlations Outside of 95%
Confidence Intervals based on r and z Standard Errors

	N = 20		N = 30		N = 50		N = 100	
	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
CI r								
$\rho = .2$	1.36	-0.54	0.64	-0.60	0.32	-0.82	0.24	-0.32
$\rho = .5$	2.38	-1.64	1.62	-1.46	1.50	-1.20	0.86	-0.92
$\rho = .8$	2.70	-2.42	2.10	-2.40	1.70	-2.00	0.70	-1.50
CI r(t)								
$\rho = .2$	0.68	-1.20	0.42	-0.88	0.30	-0.88	0.32	-0.20
$\rho = .5$	1.72	-2.02	1.30	-1.68	1.48	-1.24	0.94	-0.84
$\rho = .8$	2.28	-2.48	1.76	-2.42	1.68	-2.02	0.76	-1.46
CI z								
$\rho = .2$	0.26	0.80	0.00	0.36	-0.22	-0.38	-0.20	0.10
$\rho = .5$	-0.14	0.94	-0.46	0.86	-0.18	0.12	-0.36	0.40
$\rho = .8$	-1.10	1.46	-0.98	0.88	-0.66	0.20	-0.60	0.06
CI z(t)								
$\rho = .2$	-0.22	0.16	-0.24	0.08	-0.28	-0.38	-0.02	0.16
$\rho = .5$	-0.68	0.30	-0.60	0.40	-0.20	0.04	0.26	0.50
$\rho = .8$	-1.44	0.76	-1.14	0.58	-0.72	0.14	-0.52	0.12

Note: Percentiles are discrepancies of observed correlations from the expected value of 2.5%. Critical values were averaged across the 5000 replications within each condition. The standard error used for CI r and CI $r(t)$ was computed with the formula, $SE_r = (1 - r^2)/\sqrt{(n-1)}$. CI r was calculated as $r \pm (2 * SE_r)$. CI $r(t)$ was calculated as $r \pm (t * SE_r)$ with t having $df = N - 1$. The standard error used for CI z and CI $z(t)$ was computed with Fisher's formula, $SE_z = 1/\sqrt{(n-3)}$. CI z was calculated as $z \pm (2 * SE_z)$. CI $z(t)$ was calculated as $z \pm (t * SE_z)$. with t having $df = N - 1$

Table 5 shows that the symmetrical CI's give up to double the 2.5% in the lower region and down to almost 0.0 in the higher region. Using a t table does reduce the discrepancies in the lower region at the expense of increasing the discrepancies for the upper area with N's of 20 to 30. These discrepancies make symmetrical intervals almost unusable. While the t table was slightly better with N's of 20 to 30, the differences at large N's were small. The z transformed values have a better split, with about the same percent in the lower region as in the upper region, and occasionally gives 94% confidence intervals rather than 95% with smaller samples; with larger samples the total percentage is accurate to within one-half a percentage point.

Both procedures for confidence intervals are better behaved with larger N's. As some of the formula CI's are off by almost three percentage points with smaller N's and is still off by 1.5 points with the largest N, non-symmetrical CI's using Fisher's z transformation are the more accurate of the two. Given the similar accuracy of the SE's in Table 4, the advantage of z transformations can be attributed to the fact that it provides non-symmetrical estimates, which is particularly useful with skewed data.

Using the "good enough" procedure of multiplying the SE by 2 – one digit accuracy — works fairly well. But it is so simple to use t and t does .provide a minor improvement, it seems better to use the 3 digit accuracy for calculations, provided we remember that, as is illustrated by Table 5, the result is only accurate to plus and minus 1 of the probably, that is, the 2.5% tail is actually 2 to 3%.

Table 5 supports three conclusions. First, the upper boundary of the CI from SE r approach is too high. Second, the CI's from the z approach are more accurate. Third, unlike the mean bias corrections which are not needed with N's of 30 or more, the z approach gave better results even for N's of 100.

Discussion

The major conclusion from this study is that the seldom investigated effects of the skewed distributions on confidence intervals of r has more impact on statistics than the widely investigated impact on the mean correlation. The symmetrical confidence intervals over-estimate the frequency of cases below the lower boundary – up to 2 to 3 percentage points – while under estimating cases above the upper boundary. This problem is not eliminated by simply increasing the N. Fisher's z transformation procedure gives more accurate results, particularly for the upper boundary. In large part, this is due to the greater interval from the correlation to the lower confidence boundary as compared to the interval from the correlation to the upper confidence boundary. We recommend that confidence intervals be based on Fisher's z transformation.

The confidence intervals were based on the observed correlation for the estimated population ρ. Our results found no problem with doing so.

The mean estimates were within rounding error when N's were greater than 30. Although there are special instances in meta-analysis, which are not a topic within this study, where more precise estimates of r may be warranted, generally mean estimates can be considered accurate. There is one potential problem about means. With the negative skew, the majority of users observe a correlation that is farther from 0.0 than the population value. Whether or not this is sufficient to distort the discussion of correlations in a substantive area remains to be seen.

The results of the analysis of bias in the mean correlation replicated

previous findings that there is a small mean bias with small N's. We extend those conclusions to larger N's than previously investigated. With larger N's, the bias is minimal – generally no more than rounding error. We support others who have suggested that the bias is trivial. At the conclusion of investigating the applicability of mean bias correction procedures, Silver and Dunlap (1987) report that the correction is less than rounding error with N's of 30 or more. Zimmernman, Zumbo, and Williams (2003) also suggested that the differences between the procedures were trivial with a larger N. Our data agree with their conclusion.

Table 6 was developed to put the mean bias into perspective. It contains the underestimate of ρ when, first, the reliability of a scale decreases, based upon the attenuation due to low reliability formula. Second, Table 6 contains the attenuation of the observed correlation when the range of scores in the sample is restricted. These results show that the problem of mean bias is probably one of the least important considerations when reviewing or designing a study. Attenuation can generally be offset by using a slightly more reliable measure, by using a broader sample, or by increasing the sample size.

Table 6
Comparison of the Effect of Attenuating Artifacts

Adjustment	$\rho = .2$	$\rho = .5$	$\rho = .8$
Reliability			
$r_{xx} = .7$	-0.033	-0.082	-0.131
$r_{xx} = .8$	-0.021	-0.053	-0.084
$r_{xx} = .9$	-0.010	-0.026	-0.041
Range Restriction			
RR = 80%	-0.042	-0.129	-0.288
RR = 90%	-0.021	-0.063	-0.136
RR = 95%	-0.010	-0.031	-0.067

In Zimmerman, et al. (2003) analysis of the impact of errors in correlations on significance levels, the most discrepant p's were generally within 1 point of the first non-zero digit, except in the case of outliers or distributions mixing in log normal distributions. In this study, the CI's p's based on Fisher's z were also accurate within 1 point of the first non-zero digit. Nothing in this study would dispute the conclusion that the effects of mean bias and confidence interval distortions are not important so long as one remembers that the reported p's are accurate to the first non-zero digit only. Reporting p's with more digits implies more accuracy than our data generally warranted, and should be discouraged.

References

Alexander, R. A., Scozzaro, M. J., & Borodkin, L. J. (1989). Statistical and empirical examination of the Chi-Square Test for homogeneity of correlations in meta-analysis. *Psychological Bulletin*, *160*, 329-331. doi:10.1037/0033-2909.106.2.329

Callender, J. C., & Osburn, H. G. (1988). Unbiased estimation of sampling variance of correlations. *Journal of Applied Psychology*, *73*, 312-315. doi:10.1037/0021-9010.73.2.312

Cohen, J., Cohen, P., West, S.G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Corey, D. M., Dunlap, W. P., & Burke, M. J. (1998). Averaging correlations: Expected values and bias in combined Pearson rs and Fisher's z transformations. *The Journal of General Psychology*, *125*, 245-261. doi:10.1080/00221309809595548

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, *10*, 507-521.

Fisher, R. A. (1921). On the "probable error" of a coefficient of correlation deduced from a small sample. *Metron*, *1*, 3-32.

Fisher, R. A. (1924). The distribution of the partial correlation coefficient. *Metron*, *3*, 329-332.

Guilford, J. P. (1956). *Fundamental statistics in psychology and education.* (3rd ed.). New York: McGraw-Hill.

Hafdahl, A. R. (2010). Random-effects meta-analysis of correlations: Monte Carlo evaluation of mean estimators. *British Journal of Mathematical and Statistical Psychology*, *63*, 227-254. doi:10.1348/000711009X431914

Hafdahl, A. R., & Williams, M. A. (2009). Meta-analysis of correlations revisited: Attempted replication and extension of Field's (2001) simulation studies. *Psychological Methods*, *14*, 24-42. doi:10.1037/a0014697

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks, CA: Sage.

Meng, X. L., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin*, *111*, 172-175. doi:10.1037/0033-2909.111.1.172

Olkin, I., & Pratt, J. W. (1958). Unbiased estimation of certain correlation coefficients. *The Annals of Mathematical Statistics*, *29*, 201-211. doi:10.1214/aoms/1177706717

Romano, J. L., Kromrey, J. D., & Hibbard, S. T. (2010). A Monte Carlo study of eight confidence interval methods for coefficient alpha. *Educational and Psychological Measurement*, *70*, 376-393. doi:10.1177/0013164409355690

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. doi:10.1037/0021-9010.73.4.665

Silver, N. C., & Dunlap, W. P. (1987). Averaging correlation coefficients: Should Fisher's z transformation be used? *Journal of Applied Psychology, 72*, 146-148. doi:10.1037/0021-9010.72.1.146

Strube, M. J. (1988). Averaging correlation coefficients: Influence of Heterogeneity and set size. *Journal of Applied Psychology, 73*, 559-568. doi:10.1037/0021-9010.73.3.559

Zimmerman, D. W., Zumbo, B. D., & Williams, R. H. (2003). Bias in estimation and hypothesis testing of correlation. *Psicologica, 24*, 133-158.