A Bluffer's guide to Meta Analysis I: Correlations

Andy Field

Royal Holloway, University of London,

Meta-analysis is an increasingly popular tool in modern statistics. Put simply, meta-analysis is a way in which data from several different studies can be assimilated in an objective way. The traditional approach to literature reviews of a particular scientific area has been to collate findings in a subjective manner, which can often lead to bias representations of the literature and unjustified speculations by the reviewer. The main advantage of meta-analysis is that, in theory, it provides a framework for a scientifically rigorous accumulation of research findings (but see Wolf, 1986 for some pitfalls). In summary, by doing meta-analysis we hope to obtain some idea of what conclusions would have been reached had the data from lots of independnet studies been collected in one big study.

This article introduces various techniques for combining correlation coefficients from different studies (collected with different sample sizes) to produce an overall measure of the true relationship. Some data are presented to investigate which method is best.

1.1. The Simple Average

1.1.1. Theory

The easiest way to combine correlation coefficients from different studies is to take an average of the coefficients. Equation 1 shows how this average is calculated. Put simply, you add the correlation coefficients together and then divide by the number of coefficients.

$$\bar{r} = \frac{\sum r_i}{n} \tag{1}$$

1.1.2. Example

Imagine we are interested in combining the correlation coefficients from four independent studies. In three out of the four studies the correlation is significant. However, in study four no significant correlation was found and in study 2, the correlation coefficient was only just significant. Given the inconsistency of these results, it might be prudent to conduct a meta-analysis on these correlation coefficients to see whether the overall size of the relationship. Table 1 shows the correlation coefficients from the four studies, we will use these values throughout this article.

Table 1

	Study 1	Study 2	Study 3	Study 4
r	0.453	0.321	0.301	0.075
Sig.	<0.001	0.049	0.033	0.722
N	87	38	50	25

To calculate the simple average, we add the coefficients together and divide by the number of coefficients (in this case 4):

$$\bar{r} = \frac{\sum r_i}{n} = \frac{0.453 + 0.321 + 0.301 + 0.075}{4} = 0.288$$

Cohen (1988) termed 0.10 as a small effect size, 0.30 as a medium effect size, and 0.50 as a large effect size. By Cohen's criteria this averaged correlation represents a medium effect size. This average correlation is well below three of the four observed correlation coefficients which demonstrates how this measure can be biased. In this example, study 4, in which a very small correlation coefficient was found, used only a very small sample, yet this coefficient is treated equally to the other studies in which larger samples were used. The end result is that the average correlation has been suppressed by one, relatively small, study. One solution to this problem is the weighted average.

1.2. The Weighted Average

1.2.1. Theory

The weighted average is similar to the simple average, except that each correlation coefficient is 'weighted' by the sample size on which it is based. Equation 2 shows how this weighted average is calculated. First, each correlation coefficient is multiplied by the sample size on which it is based. Then, these products are added together (this gives us the top half of equation 2). Finally, the total sample size is calculated by adding together the individual sample sizes (this is the bottom half of equation 2).

$$\bar{r}_w = \frac{\sum (N_i r_i)}{\sum N_i} \tag{2}$$

1.2.2. Example

Using the same four example studies. We can replace the various components of equation 2 to give us:

$$\bar{r}_w = \frac{\sum (N_i r_i)}{\sum N_i} = \frac{(87 \times 0.453) + (38 \times 0.321) + (50 \times 0.301) + (25 \times 0.075)}{87 + 38 + 50 + 25} = \frac{68.53}{200} = 0.343$$

The weighted average is substantially larger than the unweighted value of 0.288: this is because studies with larger samples are given more emphasis. A word of warning should be made here. That is, if there is one study (or a small minority of studies) that use substantially larger samples than the others, then this procedure will bias heavily in favour of the large sample. In this example, the weighted average is actually greater than three of the observed correlation coefficients and this is because the correlation observed in the largest sample (study 1) was much greater than the other three studies. If the sample size of study 1 had been even greater (say 500 rather than 87) then the weighted average correlation would have reached 0.417 (work it out for yourself!).

1.3. Fisher's Transformed Correlation (Z_r)

1.3.1. Theory 1: Transforming the correlation coefficient

As values of the correlation coefficient in the population depart from zero, the distribution of coefficients sampled from that population becomes skewed. In short, this introduces a small bias when comparing correlation coefficients from different studies: especially as we are always likely to compare correlation coefficients that are different from zero (zero correlations are unlikely to be published in journals!). Fisher devised a transformation of the correlation coefficient that ensures that sample distributions are normal. Many authors (Rosenthal, 1991; Wolf, 1986) recommend that when combining correlation coefficients from different studies, these coefficients should first be transformed using Fisher's method. Therefore, when calculating the simple or weighted average of several correlation coefficients we should first transform each one.

Fisher's transformation is achieved using Equation 3. Simply take your existing correlation coefficient and replace it in the equation. Imagine a correlation coefficient of 0.5. This value can be transformed into Z_r by first dividing the correlation coefficient plus one, by that same coefficient minus 1 (1.5/0.5 = 3). You then need to find the natural logarithm of this value. The natural Log of 3 is 1.0986 (this value can be obtained using the *ln* button on most scientific calculators). Once this value is obtained, simply divide it by 2. The end result for a correlation coefficient of 0.5 is, therefore, (1.0986/2) = 0.549. To save you the trouble of calculating Fisher's transformed *r* many statistics textbooks (such as Howell, 1997) include tables of the transformed values of *r*. These transformed values can then be used to calculate the simple average or weighted average. The Appendix shows SPSS syntax for this conversion also.

$$Z_r = \tfrac{1}{2} Log_e \left(\frac{1+r}{1-r} \right) \tag{3}$$

1.3.2. Theory 2: Converting the average back to a correlation coefficient

Assuming we have transformed the correlation coefficients of interest to find the respective values of Z_r and then calculated the average or weighted average, this resulting average is in transformed form ($\overline{Z}_r$). In other words, it is not a correlation coefficient. Therefore, the averaged value has to be converted back. This conversion is achieved by re-arranging equation 3. Equation 4 shows this re-arranged form of the equation in which *e* is the base of natural Logarithms and *z* is the Fisher transformed version of *r*. The Appendix shows SPSS syntax for this conversion also.

$$r = \frac{e^{\frac{z}{0.5}} - 1}{1 + e^{\frac{z}{0.5}}} \quad (4)$$

1.3.3. Example: Transforming *r* and calculating the average.

Equation 3 can be broken down systematically to calculate the value of Z_r for each correlation coefficient. Table 2 shows how this can be done for the four studies used in previous examples. Remember that to find the value of Log_e you look for the *ln* button on your calculator.

Table 2

	Study 1	**Study 2**	**Study 3**	**Study 4**
r	0.453	0.321	0.301	0.075
$1+r$	1.453	1.321	1.301	1.075
$1-r$	0.547	0.679	0.699	0.925
$\frac{1+r}{1-r}$	2.656	1.946	1.861	1.162
$Log_e\left(\frac{1+r}{1-r}\right)$	0.976	0.666	0.621	0.150
$\frac{1}{2}Log_e\left(\frac{1+r}{1-r}\right)$	0.488	0.333	0.311	0.075

Once the correlation coefficients have been calculated, an average can be calculated as follows:

$$\overline{Z}_r = \frac{\sum Z_{r_i}}{n} = \frac{0.488 + 0.333 + 0.311 + 0.075}{4} = 0.302$$

Alternatively, the weighted average can be calculated:

$$\overline{Z}_{wr} = \frac{\sum\left(N_i Z_{r_i}\right)}{\sum N_i} = \frac{(87 \times 0.488) + (38 \times 0.333) + (50 \times 0.311) + (25 \times 0.075)}{87 + 38 + 50 + 25} = \frac{72.54}{200} = 0.363$$

1.3.4. Example II: Converting the average back

To convert the average and weighted average back into a correlation coefficient, we use equation 4. The average transformed value is converted back as follows:

$$\bar{r} = \frac{e^{\frac{\bar{z}_r}{0.5}} - 1}{1 + e^{\frac{\bar{z}_r}{0.5}}} = \frac{e^{0.604} - 1}{1 + e^{0.604}} = \frac{1.829 - 1}{1 + 1.829} = 0.293$$

Likewise, the weighted average is transformed back using the same equation:

$$\bar{r}_w = \frac{e^{\frac{\bar{z}_{wr}}{0.5}} - 1}{1 + e^{\frac{\bar{z}_{wr}}{0.5}}} = \frac{e^{0.726} - 1}{1 + e^{0.726}} = \frac{2.067 - 1}{1 + 2.067} = 0.348$$

These values of 0.293 and 0.348 can be compared to the earlier averages calculated without first transforming the correlation coefficients. In both cases, the average using the transformed correlation coefficients are greater.

1.4. Which Method is Best?

1.4.1. To weight or not to weight?

So far I have described four different methods of obtaining combined correlation coefficients from several studies. However, a pertinent question is which method provides the best results. Hunter and Schmidt (1990) argue that it would be a very rare case in which an unweighted analysis proved better than a weighted one. However, the weighted method gives greatest weight to large studies and as such, in situations in which one study has a much greater sample than the others being compared, this could create a bias. Nevertheless, if the population correlations from which the samples were taken are the same then the weighted version will always be better. In addition, if population correlations differ by only a little then the weighted version will still be superior, and even when population correlations differ considerably, the weighted version is still better provided that the sample size does not correlate with the population correlation (See Hunter & Schmidt, 1990).

1.4.2. To transform or not to transform?

The next question is whether the Fisher transformation yields any benefit. Rosenthal (1994) certainly believes so. However, Hunter and Schmidt (1990) insist that the Fisher transformed average is lass accurate than the untransformed version. I conducted a small experiment to test (under limited conditions) the accuracy of each of the four methods described. This work was a pilot study for a systematic Monte Carlo investigation of the factors influencing the various methods. Assume that the actual relationship of interest is a medium sized effect by Cohen's criterion (the actual population correlation coefficient was set at 0.347). On each occasion I combined the correlation coefficients arising from only four studies. Therefore, four correlation coefficients were combined to obtain the average, weighted average, transformed average and transformed weighted average. The averages based on the transformed correlation coefficients were then converted back to correlation coefficients (using equation 4). The combined sample size across studies was constant (N = 200), but the sample size of the four studies differed according to five set ratios. For each of these sample size ratio combinations, 10 random samples were taken. Therefore, 50 different correlation coefficients were derived for each study, which combined to make 50 of each type of average correlation.

Table 3

Sample Ratio	Average		Weighted Average	
	Not Transformed	Transformed	Not Transformed	Transformed
87:38:50:25	0.358	0.365	0.348	0.355
50:50:50:50	0.346	0.354	0.346	0.354
60:40:60:40	0.346	0.352	0.346	0.352
60:40:75:25	0.337	0.343	0.343	0.349
80:20:70:30	0.354	0.369	0.349	0.358
Total	**0.348**	**0.357**	**0.347**	**0.354**

Table 3 shows the resulting averaged correlation coefficients for each of the five sample size ratios. The ratios represent the samples on which the four studies are based, so 50:50:50:50 means that all four studies used samples of 50, whereas 60:40:60:40 means that studies 1 and 3 used samples of 60 whereas studies 2 and 4 used samples of 40.

This manipulation was to see whether small variations in relative sample sizes would influence the accuracy of the resulting averages. For each set of ratios there were ten sets of samples taken and the correlation coefficients in the table represent the averages of these 10 trials. The total correlation coefficients represent the correlation coefficients collapsed with respect to the sample size ratios.

A three way 2 (Weighted: weighted or not) × 2 (Transformed: transformed or not) × 5 (ratio: sample size ratios) ANOVA was conducted on the resulting coefficients. This analysis revealed no significant effects involving the sample size ratio manipulation. However, there was a significant effect of whether the averages were based on transformed or non-transformed correlation coefficients [$F(1, 45) = 64.40$, $p < 0.001$]. This showed that averages based on Fisher transformed coefficients were significantly larger than those based on ordinary coefficients. There was no main effect of weighting the average [$F < 1$], but there was an interaction between whether the mean was weighted and whether it was based on transformed values [$F(1, 45) = 4.36$, $p < 0.05$]. This interaction seemed to suggest that weighting the average had a benefit only when that average was based on transformed scores (see Figure 1).

Bearing in mind that we know that the correlation that we are trying to estimate from these averages is actually 0.347, Figure 1 shows that the weighted average (untransformed) provides the correct value, on average. All other averages over-estimate the correlation. The significant over-estimation caused by using Fisher-transformed coefficients supports Hunter and Schmidt's (1990) work. There are limitations to this simulation study, for one thing, sample sizes although varying were relatively similar throughout (which probably accounts for the absence of an effect of weighting the average). In addition, these results are based on relatively few trials, on a situation in which only four coefficients are averaged, and on only one population coefficient (what would happen if the correlation in the population was large or small?). Nevertheless, it supports other work that suggests that the Fisher transformation is un-necessary. However, for the reasons suggested by Hunter and Schmidt (1990) a weighted average is preferred.

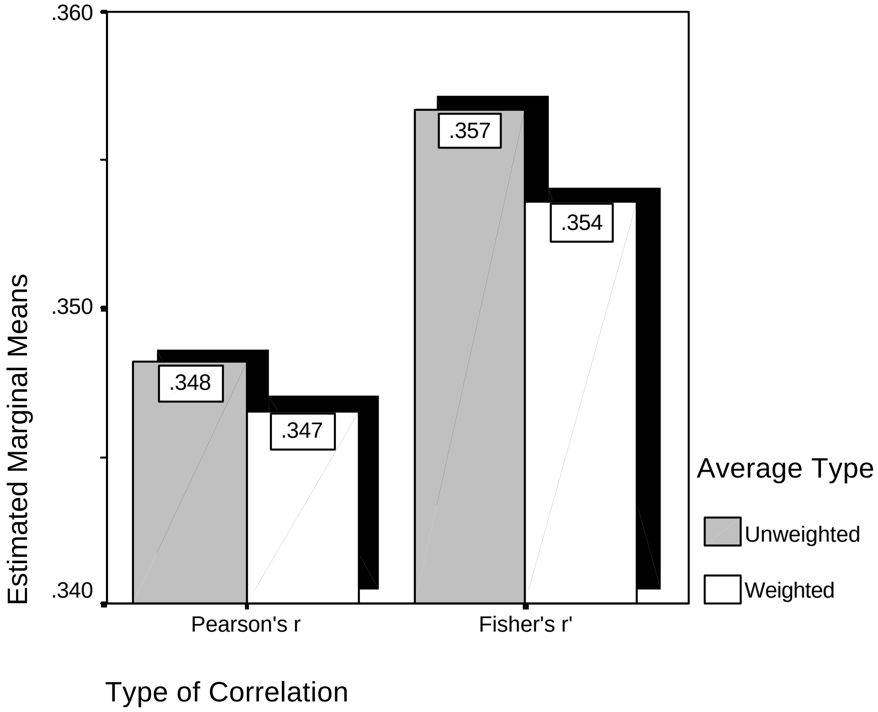

Figure 1

1.5. Summary

Four methods of combining correlation coefficients from several studies are described. Of the four, the weighted average appears best for theoretical reasons and data presented supports the previous suggestion that Fisher-transformed averages significantly overestimate the actual correlation.

1.6. Appendix

SPSS syntax to transform a column of correlation coefficients (labelled *r*) into a column of Fisher-transformed coefficients (labelled *z*):

```
COMPUTE z = 0.5 * (ln((1 + r)/(1 - r))) .
EXECUTE .
```

SPSS syntax to transform a column of Fisher-transformed coefficients (labelled *z*) into a column of correlation coefficients (labelled *r*):

```
COMPUTE r = ((EXP(z/0.5))-1)/(1 + EXP(z/0.5)).
EXECUTE.
```

References

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2^{nd} Ed.). New York: Academic Press.

Field, A. P. (in press*). Discovering statistics using SPSS for windows: advanced topics for beginners*. London: Sage.

Howell, D. C. (1997). *Statistical methods for psychology* (4^{th} Edition). Belmont, CA: Duxbury.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of Meta-analysis: correcting error and bias in research findings*. Newbury Park: Sage.

Rosenthal, R. (1991). *Meta-analytic procedures for social research* (revised). Newbury Park: Sage.

Wolf, (1986). *Meta-Analysis*. Sage university paper series on quantitative applications in the social sciences, 07–061. Newbury Park, CA: Sage.