August 2021

META-ANALYSIS: EASY TO ANSWER

Version IV

David A. Kenny

University of Connecticut

Meta-analysis, Easy to Answer, or META, was designed to assist the user in the computation of statistics during meta-analysis. META computes an effect size for each study, pools these effect sizes, and tests them for homogeneity. It tests both whether the average effect size is different from zero and computes the combined probability. Results can be weighted by sample size, study variance, or user-inputted values. The program is a compiled version of a QB64 (QuickBasic) program. Because the program was written nearly 30 years ago, it has a very old-fashioned DOS-like appearance. It is awkward to use.

META cannot be used to measure correlates of effect sizes. For instance, if a researcher seeks to know whether more recent studies had larger effect sizes than older studies, the researcher cannot answer this question with META. However, META produces a file of effect sizes that can be combined with moderators variables like study recency, and with a statistical program (e.g., SPSS or SAS) that question could be answered.

META follows the standard convention in data analysis that *p* values are two-tailed. In some presentations of meta-analysis, formulas are given that involve one-tailed *p* values. META always uses the two-tailed version of formulas.

META treats study as the basic unit in the analysis. If there are multiple outcomes from the same study, one must choose the most appropriate one. Also META cannot handle one-group studies, repeated-measures or within participants designs, or factors with more than two levels.

META has three stages of operation. In the first stage preliminary information is gained that informs META of the basic job parameters. In the second stage, information concerning each study is entered. In the third stage, the information concerning the studies is integrated. So the three stages are setup, study input, and integration.

The results of a meta-analysis is dependent on whether the studies found are representative and the *p* values from the study are accurate.

Setup

The user begins by typing META. Responses to requests from the program can be given in upper or lower case. The program initially asks the user about the choice of output file name. If the file already exists, META writes over it. So the user should never give as the name for the output file, the name of a data file.

In META there are usually a series of options that are given when questions are asked. The default answer is ordinarily listed first. Typing the ENTER key usually results in the first option being implemented.

The program then requests the type of effect size. There are three basic measures of effect size: d, correlation, the difference between proportions. Cohen's d equals $(m_1 - m_2)/s$ where the symbol m is the mean and s the standard deviation. The measure d is appropriate when there are two groups and the dependent variable is continuous. The measure r or correlation coefficient is appropriate when both the independent and dependent variables are continuous. The measure of the difference between proportions or $p_1 - p_2$ is appropriate when both the independent and dependent variables are dichotomies. If the researcher wishes to combine only probabilities, then no effect size need be computed.

For each of the three measures of effect size, there are possible transformations. If d is used as the measure of effect size, it is possible to employ the Hedges correction which controls for the bias in d as a function of degrees of freedom. If r is the measure of effect size, the correlations can be transformed by the Fisher's z that normalizes the distribution of r. There are three possible normalizing transformations of proportions: arcsin, logit, and probit. The arcsin equals $\arcsin(2p_1) - \arcsin(2p_2)$. The logit equals $\ln[p_1/(1 - p_1)] - \ln[p_2/(1 - p_2)]$ where ln is the natural logarithm. Probit involves computing the value on the abscissa of standard normal distribution for a cumulative probability. Arcsin is the least radical of these transformations, probit the most, and logit is intermediate. The user can consult Kenny (1987) for a detailed discussion concerning these three transformations.

If the effect size measure is not r, META asks if it can be assumed if group sizes are equal. If the answer is no, META requests the group sizes for the two groups from each study. If the answer is yes, META divides the number of participants by two to determine the sample size in each group, even if the total sample size is odd.

META requests whether the user has an input dataset. That is, is there a data file that will be re-analyzed in the current run of META? If there is, the user types Y and enters the name of the dataset. META looks for the data file and if it cannot find it, META notifies the user and ends the run.

The program asks if the studies should be weighted: Should some studies count more than others. For instance, studies with a larger sample size or with a more reliable estimate of effect size might be weighted more. There are three different choices for weights. First, one can weight using the sample size or degrees of freedom. If either of these is desired, the weight can be square rooted. Second, one can weight by the variance of the effect size. This option can be chosen only if an effect size measure is computed. Finally, the user may input for each study a user-determined weight. For instance, the user may have coded each study's quality. If a user-inputted weight is used, it must be a value that is greater than zero. One cannot, by setting a weight to zero, use the weighting procedure to drop a study from the meta-analysis.

If the arcsin transformation is used, then the study variance weights are the sample sizes. If one weights and uses a Fisher's z transformation of correlations, the weights are the sample sizes less three.

The final setup question is the following: For each study does the degrees of freedom for the test of significance equal the sample size less two. For most studies this would likely be true. However, if F tests from factorial designs are used or partial correlations are computed, the answer would be no.

Within META, the user is presented with the review of the choices made and is asked whether to proceed with the analysis.

Study Information

For each study the user is presented with a basic screen. It lists the different types of study parameters that are possible. If the user has already entered the last study and so there are no more studies, then the user types the symbol E to tell META that one is finished.

The following are the options for study statistics to use to compute an effect size: t, F, r, chi square, p value, means and variances, proportions, and Z. For each of these statistics, one needs to include additional information. Usually the sample size or degrees of freedom for the study must also be inputted.

```
      Type of Statistic for Study 1
          T (t test)            M (Means and Variances)
          F (F test)            P (p Value)
          R (Correlation)       S (Proportions)
          X (Chi Square)        Z (Z Test)
                 E (No More Studies)
       Your Choice?
```

If the p value is inputted, META expects a two-tailed p value. Also if the reported p value is "not significant," then the two-tailed p value is 1.000 and not .5. Whatever the p value, the user is asked whether the p value is from a t or Z statistic.

If the user knows the d for the study, then the user should ask for means and variances (M). Input the d for the mean of the first group and zero for the mean of the second group. Input one as the variance or the standard deviation.

META always computes the effect size as a positive value, even if the information is given in the other direction. It then asks the user if the result is in the expected direction. The expected direction is determined by the user. Studies that are in the unexpected direction are treated as having negative effect sizes. For instance, in a dataset to be discussed later the effect of the intervention is to lower the probability that children are likely to have problems in school. Technically the effect size should be negative. However, the user should probably define the expected direction as showing a decline. It is important to be consistent across the entire set of studies.

After the information has been inputted for a given study, META computes the basic statistics and displays them for the user. Particular attention should be given to the direction of the effect. If there were any errors in the entry, then the study can be deleted. Below META gives the basic statistics where the effect size is the difference between proportions and the two proportions that were entered were .5 and .6. The sample size is 100.

```
                         STUDY 1
          N:  100                   df:  98
          d: .1990                   r: .1000
          Z:  1.0000               t: .9949
   Variance: .0098               BESD: .4500 To .5500
                   Effect size: .1000

     Do you want to include these results in the meta-analysis?
         Yes (Y) Or No (N)?
```

The study variance is a measure of how reliable the measure of effect size is. The BESD is the binomial effect size (Rosenthal & Rubin, 1982) and measures the estimated difference between two groups in terms of a proportions. If one wanted to delete this study, one would type N.

META stores the results of its computations in an output data file. This output data file can be used as an input data file in another run of META. There are two reasons for saving a data file. First, one may find new studies that one could then add to the file. Less frequently one might decide to delete a study from the data file. Second, one might decide to alter the weighting procedure or change the effect size measure. Of course, it is improper to make changes only to produce a desired result. The output data file stores the following variables in free format: study number, degrees of freedom, number of participants, effect size, d, r, t, Z, p_1, p_2, n_1, n_2, and weight. The record may extend beyond column 80. Ordinarily if one were to integrate this data file with study characteristics, the two key values are the effect size and the sample size. The file can be read into spreadsheet programs (e.g., EXCEL) or statistical packages (e.g., SPSS). The format for META is free format (i.e., spaces between variables).

If there is an input dataset, the user is asked if he or she wants to review the results. Such a review would be necessary if the user wanted to input new weights or to delete studies. To delete a study, the user asks to review the individual studies and when the study to be deleted is presented, chose the option not to include it in the meta-analysis. If the user only wishes to add studies or to change the effect size measure, it is not necessary to review each study. The program will process the studies and ask the user if there are any more studies. There is no practical upper limit to the number of studies that META can process.

After the last study in inputted, type E and META moves onto the third stage.

Integration

After all the data are entered, META computes the summary statistics. A re-analysis of a meta-analysis done by Burger (1981) is considered. Burger examined a phenomenon of defensive attribution which is the supposition that people attribute less blame to people who are similar to them. Using the *t* values and the sample sizes presented in Burger's Table 1, the following summary was obtained:

```
    Effect size measure is Cohen's d
       Transformation: Hedges

       Study number:     22
       Subject N:      2202
          Average effect size:  .1210
          Effect size sd: .2799
          t test of effect size:  2.0274  p value: .05559  df: 21
             Average d: .1219
             Average r: .0589
             BESD: .4705 to .5295

        Homogeneity of effect sizes
                Chi square: 32.4486   p value: .05270 df:  21

         Average Z:  3.1403    p value: .00169
                Fail-safe N:  35
         Average t:  3.1052    p value: .00163
                Fail-safe N:  35
```

The measure of effect size is Cohen's d and the Hedges correction has been used. The total number of studies is 22 and the number of participants in the studies is 2,202. The average effect size across the 22 studies is .1210. Cohen defines a d of .2 as small, and so this effect size is about half as small as small. The standard deviation of the effect size is .2799 and the t test is not significant. Thus, the value of .1219 is not significantly different from zero. The average effect size, its variance, and the test that it is statistically significant are presented. If the average effect size is significant ($p < .05$, two-tailed), the fail-safe N is outputted. The fail-safe N represents the number of studies with zero effect size that there would have to be make the result no longer statistically significant ($p < .05$, two tailed).

If the studies are weighted, the estimate of the average effect size and its test are weighted. If the data were transformed, META also produces the untransformed estimate of average effect size. (For the Hedges transformation, this is not done because it is not really a transformation but rather a sample size correction factor.) So, for instance, if the effect size was the difference between proportions and the logit transformation was used, the untransformed effect size would be in the units of the difference between proportions.

The test of homogeneity evaluates the whether the effect sizes significantly vary from study to study. It is computed only if there is an effect size. For the Burger meta-analysis, this value is not statistically significant. The program employs six different tests of homogeneity. Which one that is employed depends on the choice of effect size and transformation. If d is the measure of effect size, the Hedges (1982) test of homogeneity is employed. If r is used, the Fisher test of homogeneity of correlations is used. For the difference between proportions one of four different tests is used which depend on the choice of transformation. So the test of homogeneity depends on the measure of effect size.

For each of these tests of homogeneity, a chi square test is used. If the chi square test of homogeneity is significant, the studies are not homogeneous, and so the measures of effect sizes differ because of factors other than sampling error. If the effect sizes are not homogeneous, it is sensible to search for correlates of the effect sizes. For the Burger meta-analysis, the test of homogeneity is not statistically significant. If the user decides to weight studies, those weights are not used in the test of homogeneity.

The next set of statistics is the combined probability values. These are the only statistics that are outputted if no effect size is computed. META computes the adding ts and Zs, both of which are statistically significant for the Burger meta-analysis. These methods in effect treat person as the unit of analysis in the meta-analysis, whereas the earlier presented t test treats study as the unit of analysis. When the chi square test of homogeneity is not significant, it is sensible to employ the combined probability method. If weighting is used, only the adding Zs is computed. For each statistic, the fail-safe N is outputted. As can be seen, the fail-safe N for the Burger meta-analysis is 35 for both tests. So there would have to be 35 null studies for this test to be not significant.

The final set of results are the optimal weighted estimate of the effect size, the estimated standard deviation of the effect sizes with sampling error removed and the estimated sampling error standard deviation. These estimates are only presented if an output file is created. For instance, for the Burger study these values are:

```
   Optimally weighted estimate: .1675
            Estimated study sd: .2959
            Estimated error sd: .1456
```

The optimally weighted effect sizes are weighted by computing the weights that require knowledge of the effect size. Thus, these estimates are iterative. As can be seen, the optimally weighted estimate is larger than the unweighted estimate. It is also seen that there is considerable variation in the effect sizes, called tau or τ. Note the test of heterogeneity of effect sizes is given in the prior screen in "Homogeneity of Effect Sizes."

Disclaimer and Purchase Information

META is an experimental program. The user should carefully check its computations. To address for the following files are as follows:

program: davidakenny.net/progs/meta.exe
this documentation: davidakenny.net/progs/meta.pdf
Burger META data file: davidakenny.net/progs/darl.dat
Darlington META data file: davidakenny.net/progs/darl.dat

Inquiries concerning the program can be sent to:

David A. Kenny
Department of Psychology
University of Connecticut
Storrs, CT 06269-1020
email: david.kenny AT uconn.edu

There is no charge for the program.

Extended Example

As an example, consider a meta-analysis conducted by Darlington, Royce, Snipper, Murray, and Lazar (1980) who examined the effects of preschool intervention strategies on children's problems once they do enter school. Seven different studies investigated the hypothesis that preschool interventions lowered the problems that children have in school (e.g., being held back a grade). The basic statistics from the seven studies are:

Study	N	Treatment	Control	χ^2
1	82	39.1	61.5	2.25
2	55	52.8	68.4	1.25
3	221	24.1	44.7	7.66
4	123	17.2	38.5	6.78
5	69	45.9	50.0	0.11
6	127	22.1	43.5	4.47
7	125	20.6	11.1	0.89

Because the basic statistic is a percentage, the appropriate effect size is the difference between proportions that is transformed by the logit transformation. No information was given concerning group size and so it will be assumed that they are equal. Degrees of freedom plus two are assumed to equal sample size. Notice that for the seventh study, the result is not in the expected direction. The children who received the intervention had more problems than those in the control group. So this group's effect size is negative. For each study, the user would tell META that the statistic is proportions (option S). One would enter the proportions and then the sample size. One will be prompted concerning whether there is a chi square result. One would answer yes and then enter the chi square. The basic set of results is as follows:

```
Effect size measure is difference between proportions
    Transformation: Logit

     Study number:      7
     Subject N:       802
        Average effect size: .5772
        Effect size sd: .6561
        t test of effect size:  2.3276  p value: .05995 df: 6
          Average d: .2539
          Average r: .1258
          BESD: .4283 to .5717
          Transformed effect size: .1433
```

```
          Homogeneity of effect sizes
                  Chi square:  11.4363   p value: .07579 df:  6
          Average Z:  3.5877   p value: .00033
                Fail-safe N:  17
          Average t:  3.5879   p value: .00033
                Fail-safe N:  17
```

The average effect size is .5772 which corresponds to a percentage difference of .1433. So children who were in the intervention were 14% less likely to have difficulties when they entered school. This value is not quite significant. The test of homogeneity indicates no significant difference between studies, the p value being .076. The combined two-tailed probability is .0003 and the fail-safe N is 17.

The next page for META for this dataset is:

```
Effect size option:   Difference between proportions
                      Logit transformation
Study weight:         Not weighted

Study number:         7
Subject N:            802

     Average effect size: 0.5772
          Effect size sd: 0.6561;  t test of effect size: 2.3276;
          p value: 0.0600;  df: 6

     Average d: 0.2539
     Average r: 0.1258
          BESD: 0.4283 to 0.5717
     Average z: 3.5877;  p value: 0.0003
          Fail-safe N: 17
     Average t: 3.5879;  p value: 0.0003
          Fail-safe N: 17

Transformed effect size:     0.1433
Homogeneity of effect sizes: Chi square: 11.4363;
p value: 0.0758;  df: 6
```

If one were to weight the effect sizes by study variances, the transformed effect size would increase to .6985 but its p value would hardly change and equals .057.

The optimally weighted effect size estimates are as follows:

Optimally weighted estimate: .7066
Estimated study sd: .6928
Estimated error sd: .3760

The optimally weighted effect size is somewhat larger than the unweighted estimate. It is also seen that there is considerable variation in the effect size estimates.

References

Burger, J. M. (1981). Motivational biases in the attribution of responsibility for an accident: A meta-analysis of the defensive-attribution hypothesis. *Psychological Bulletin, 90*, 496-512.

Darlington, R. B., Royce, J. M., Snipper, A. S., Murray, H. W., & Lazar, I. (1980). Preschool programs and later school competence of children from low-income families. *Science*, 208, 202-204.

Hedges, L. V. (1982). Estimation of effect size from a series of independent studies. *Psychological Bulletin, 92*, 490-499.

Kenny, D. A. (1987). *Statistics for the social and behavioral sciences*. Boston: Little Brown.

Rosenthal, R., & Rubin, D. B. (1982). A simple purpose display of magnitude of experimental effect. *Journal of Educational Psychology, 74*, 166-169.