Introduction to Meta-Analysis

Michael Borenstein

Larry Hedges

Hannah Rothstein

Dedication

Dedicated in honor of Sir Iain Chalmers and to the memory of Dr. Thomas Chalmers.

Section: Effect size indices 50

Section: Fixed effect vs. random effects models

Overview

One goal of a meta-analysis will often be to estimate the overall, or combined effect.

If all studies in the analysis were equally precise we could simply compute the mean of the effect sizes. However, if some studies were more precise than others we would want to assign more weight to the studies that carried more information. This is what we do in a meta-analysis. Rather than compute a simple mean of the effect sizes we compute a weighted mean, with more weight given to some studies and less weight given to others.

The question that we need to address, then, is how the weights are assigned. It turns out that this depends on what we mean by a "combined effect". There are two models used in meta-analysis, the fixed effect model and the random effects model. The two make different assumptions about the nature of the studies, and these assumptions lead to different definitions for the combined effect, and different mechanisms for assigning weights.

Definition of the combined effect

Under the fixed effect model we assume that there is one true effect size which is shared by all the included studies. It follows that the combined effect is our estimate of this common effect size.

By contrast, under the random effects model we allow that the true effect could vary from study to study. For example, the effect size might be a little higher if the subjects are older, or more educated, or healthier, and so on. The studies included in the meta-analysis are assumed to be a random sample of the relevant distribution of effects, and the combined effect estimates the mean effect in this distribution.

Computing a combined effect

Under the fixed effect model all studies are estimating the same effect size, and so we can assign weights to all studies based entirely on the amount of information captured by that study. A large study would be given the lion's share of the weight, and a small study could be largely ignored.

By contrast, under the random effects model we are trying to estimate the mean of a distribution of true effects. Large studies may yield more precise estimates than small studies, but each study is estimating a different effect size, and we want to be sure that all of these effect sizes are included in our estimate of the mean. Therefore, as compared with the fixed effect model, the weights assigned under random effects are more balanced. Large studies are less likely to dominate the analysis and small studies are less likely to be trivialized.

Precision of the combined effect

Under the fixed effect model the only source of error in our estimate of the combined effect is the random error within studies. Therefore, with a large enough sample size the error will tend toward zero. This holds true whether the large sample size is confined to one study or distributed across many studies.

By contrast, under the random effects model there are two levels of sampling and two levels of error. First, each study is used to estimate the true effect in a specific population. Second, all of the true effects are used to estimate the mean of the true effects. Therefore, our ability to estimate the combined effect precisely will depend on both the number of subjects within studies (which addresses the first source of error) and also the total number of studies (which addresses the second).

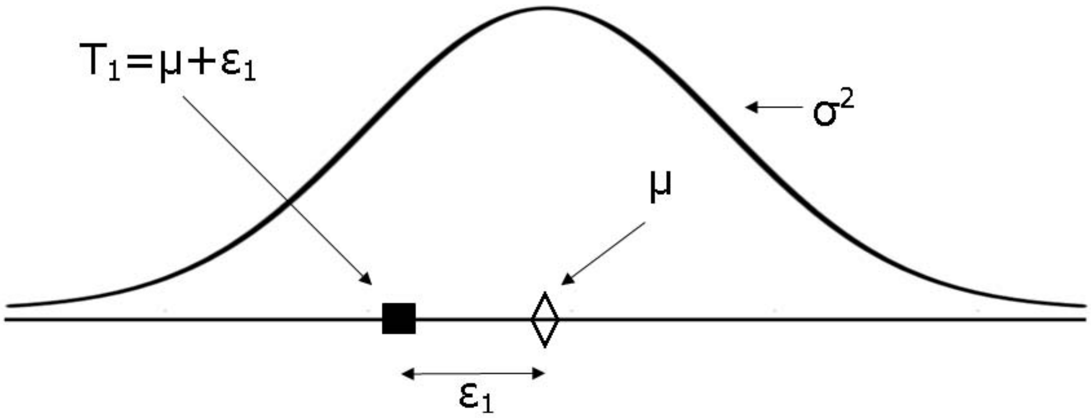
How this section is organized

The two chapters that follow provide detail on the fixed effect model and the random effects model. These chapters include computational details and worked examples for each model. Then, a chapter highlights the differences between the two.

Fixed effect model

Definition of a combined effect

In a fixed effect analysis we assume that all the included studies share a common effect size, μ. The observed effects will be distributed about μ, with a variance σ2 that depends primarily on the sample size for each study.



Fixed effect model. The observed effects are sampled from a distribution with true effect μ, and variance σ^2. The observed effect T_1 is equal to $\mu+\varepsilon_i$.

In this schematic the observed effect in Study 1, T_1, is a determined by the common effect μ plus the within-study error ε_1. More generally, for any observed effect T_i,

$$T_i = \mu + e_i \tag{2.2}$$

Assigning weights to the studies

In the fixed effect model there is only one level of sampling, since all studies are sampled from a population with effect size μ. Therefore, we need to deal with only one source of sampling error – within studies (*e*).

Since our goal is to assign more weight to the studies that carry more information, we might propose to weight each study by its sample size, so that a study with 1000 subjects would get 10 times the weight of a study with 100 subjects. This is basically the approach used, except that we assign weights based on the inverse of the variance rather than sample size. The inverse variance is roughly proportional to sample size, but is a more nuanced measure (see notes), and serves to minimize the variance of the combined effect.

Concretely, the weight assigned to each study is

$$w_i = \frac{1}{v_i} \tag{2.3}$$

where v_i is the within-study variance for study (i). The weighted mean ($\overline{T}_\bullet$) is then computed as

$$\overline{T}_\bullet = \frac{\sum_{i=1}^{k} w_i T_i}{\sum_{i=1}^{k} w_i} \tag{2.4}$$

that is, the sum of the products $w_i T_i$ (effect size multiplied by weight) divided by the sum of the weights. The variance of the combined effect is defined as the reciprocal of the sum of the weights, or

$$v_\bullet = \frac{1}{\sum_{i=1}^{k} w_i} \tag{2.5}$$

and the standard error of the combined effect is then the square root of the variance,

$$SE(\overline{T}_\bullet) = \sqrt{v_\bullet} \tag{2.6}$$

The 95% confidence interval for the combined effect would be computed as

$$Lower\ Limit = \overline{T}_\bullet - 1.96 * SE(\overline{T}_\bullet) \tag{2.7}$$

$$Upper\ Limit = \overline{T}_\bullet + 1.96 * SE(\overline{T}_\bullet) \tag{2.8}$$

Finally, if one were so inclined, the *Z*-value could be computed using

$$Z = \frac{\bar{T}_\bullet}{SE(\bar{T}_\bullet)} \tag{2.9}$$

For a one-tailed test the p-value would be given by

$$p = 1 - \Phi(|Z|) \tag{2.10}$$

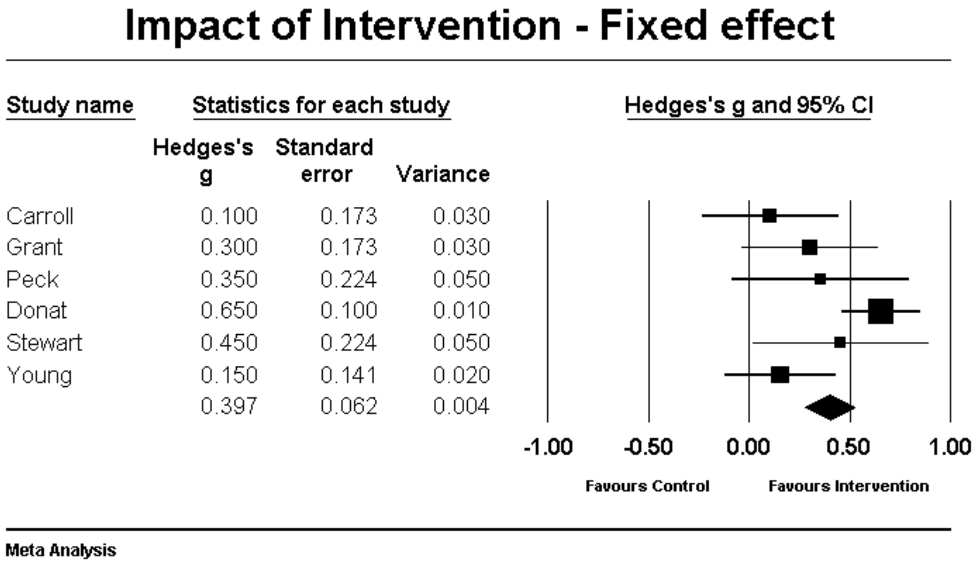
(assuming that the effect is in the hypothesized direction), and for a two-tailed test by

$$p = 2\left[1 - \left(\Phi\left(|Z|\right)\right)\right] \tag{2.11}$$

Where Φ is the standard normal cumulative distribution function.

Illustrative example

The following figure is the forest plot of a fictional meta-analysis that looked at the impact of an intervention on reading scores in children.



Impact of Intervention - Fixed effect

Study name	Statistics for each study		
	Hedges's g	Standard error	Variance
Carroll	0.100	0.173	0.030
Grant	0.300	0.173	0.030
Peck	0.350	0.224	0.050
Donat	0.650	0.100	0.010
Stewart	0.450	0.224	0.050
Young	0.150	0.141	0.020
	0.397	0.062	0.004

In this example the Carroll study has a variance of 0.03. The weight for that study would computed as

$$w_1 = \frac{1}{(0.03)} = 33.333$$

and so on for the other studies. Then,

$$\bar{T}_{\bullet} = \frac{101.833}{256.667} = 0.3968$$

$$v_{\bullet} = \frac{1}{256.667} = 0.0039$$

$$SE(\bar{T}_{\bullet}) = \sqrt{0.0039} = 0.0624$$

$$\mathit{Lower\ Limit} = 0.3968 - 1.96 * 0.0624 = 0.2744$$

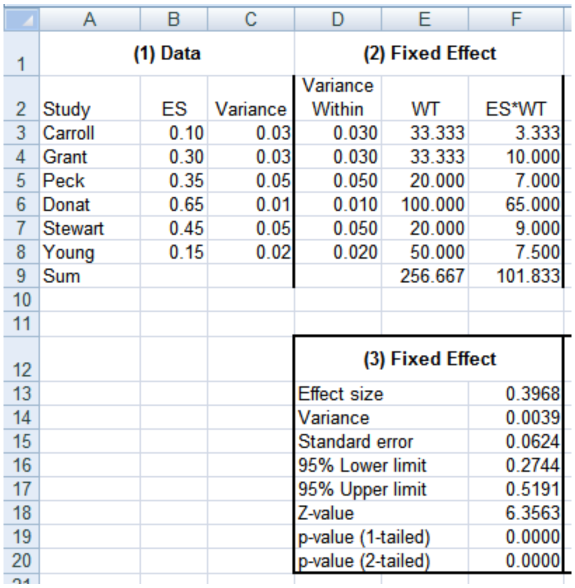
$$\mathit{Upper\ Limit} = 0.3968 - 1.96 * 0.0624 = 0.5191$$

$$Z = \frac{0.3968}{0.0624} = 6.3563$$

$$p_{1T} = 1 - \left(\Phi\left(|\,6.3563\,|\right)\right) < .0001$$

$$p_{2T} = 2\left[1 - \left(\Phi\left(|\,6.3563\,|\right)\right)\right] < .0001$$

The fixed effect computations are shown in this spreadsheet



	A	B	C	D	E	F
1		(1) Data			(2) Fixed Effect	
2	Study	ES	Variance	Variance Within	WT	ES*WT
3	Carroll	0.10	0.03	0.030	33.333	3.333
4	Grant	0.30	0.03	0.030	33.333	10.000
5	Peck	0.35	0.05	0.050	20.000	7.000
6	Donat	0.65	0.01	0.010	100.000	65.000
7	Stewart	0.45	0.05	0.050	20.000	9.000
8	Young	0.15	0.02	0.020	50.000	7.500
9	Sum				256.667	101.833
10						
11						
12					(3) Fixed Effect	
13				Effect size		0.3968
14				Variance		0.0039
15				Standard error		0.0624
16				95% Lower limit		0.2744
17				95% Upper limit		0.5191
18				Z-value		6.3563
19				p-value (1-tailed)		0.0000
20				p-value (2-tailed)		0.0000

Column (Cell)	Label	Content	Excel Formula*	See formula
(Section 1) Effect size and weights for each study				
A	Study name		Entered	
B	Effect size		Entered	
C	Variance		Entered	
(Section 2) Compute WT and WT*ES for each study				
D	Variance within study		=$C3	
E	Weight		=1/D3	(2.3)
F	ES*WT		=$B3*E3	
Sum the columns				
E9	Sum of WT		=SUM(E3:E8)	
F9	Sum of WT*ES		=SUM(F3:F8)	
(Section 3) Compute combined effect and related statistics				
F13	Effect size		=F9/E9	(2.4)
F14	Variance		=1/E9	(2.5)
F15	Standard error		=SQRT(F14)	(2.6)
F16	95% lower limit		=F13-1.96*F15	(2.7)
F17	95% upper limit		=F13+1.96*F15	(2.8)
F18	Z-value		=F13/F15	(2.9)
F19	p-value (1-tailed)		=(1-(NORMDIST(ABS(F18),0,1,TRUE)))	(2.10)
F20	p-value (2-tailed)		=(1-(NORMDIST(ABS(F18),0,1,TRUE)))*2	(2.11)

Comments

Some formulas include a “$”. In Excel this means that the reference is to a specific column. These are not needed here, but will be needed when we expand this spreadsheet in the next chapter to allow for other computational models.

Inverse variance vs. sample size.

As noted, weights are based on the inverse variance rather than the sample size. The inverse variance is determined primarily by the sample size, but it is a more nuanced measure. For example, the variance of a mean difference takes account not only of the total N, but also the sample size in each group. Similarly, the variance of an odds ratio is based not only on the total N but also on the number of subjects in each cell.

Random effects model

The fixed effect model, discussed above, starts with the assumption that the true effect is the same in all studies. However, this is a difficult assumption to make in many (or most) systematic reviews. When we decide to incorporate a group of studies in a meta-analysis we assume that the studies have enough in common that it makes sense to synthesize the information. However, there is generally no reason to assume that they are “identical” in the sense that the true effect size is exactly the same in all the studies.

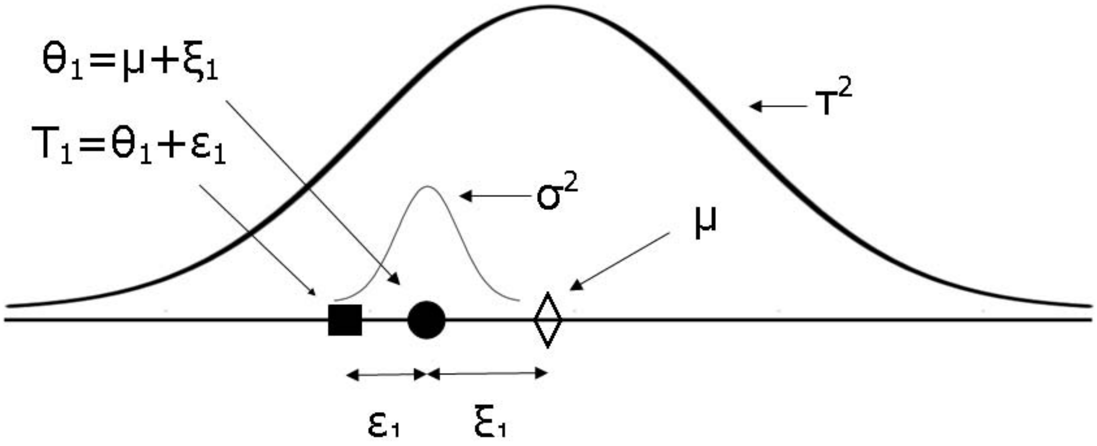
For example, assume that we’re working with studies that compare the proportion of patients developing a disease in two groups (vaccination vs. placebo). If the treatment works we would expect the effect size (say, the risk ratio) to be similar but not identical across studies. The impact of the treatment impact might be more pronounced in studies where the patients were older, or where they had less natural immunity.

Or, assume that we’re working with studies that assess the impact of an educational intervention. The magnitude of the impact might vary depending on the other resources available to the children, the class size, the age, and other factors, which are likely to vary from study to study.

We might not have assessed these covariates in each study. Indeed, we might not even know what covariates actually are related to the size of the effect. Nevertheless, experience says that such factors exist and may lead to variations in the magnitude of the effect.

Definition of a combined effect

Rather than assume that there is one true effect, we allow that there is a distribution of true effect sizes. The combined effect therefore cannot represent the one common effect, but instead represents the mean of the population of true effects.



Random effects model. The observed effect T_1 (box) is sampled from a distribution with true effect θ_1, and variance σ^2. The true effect θ_1, in turn, is sampled from a distribution with mean μ and variance τ^2.

In this schematic the observed effect in Study 1, T_1, is a determined by the true effect θ_1 plus the within-study error ε_1. In turn, θ_1, is determined by the mean of all true effects, μ and the between-study error ξ_1. More generally, for any observed effect T_i,

$$T_i = \theta_i + e_i = \mu + \varepsilon_i + e_i \tag{3.1}$$

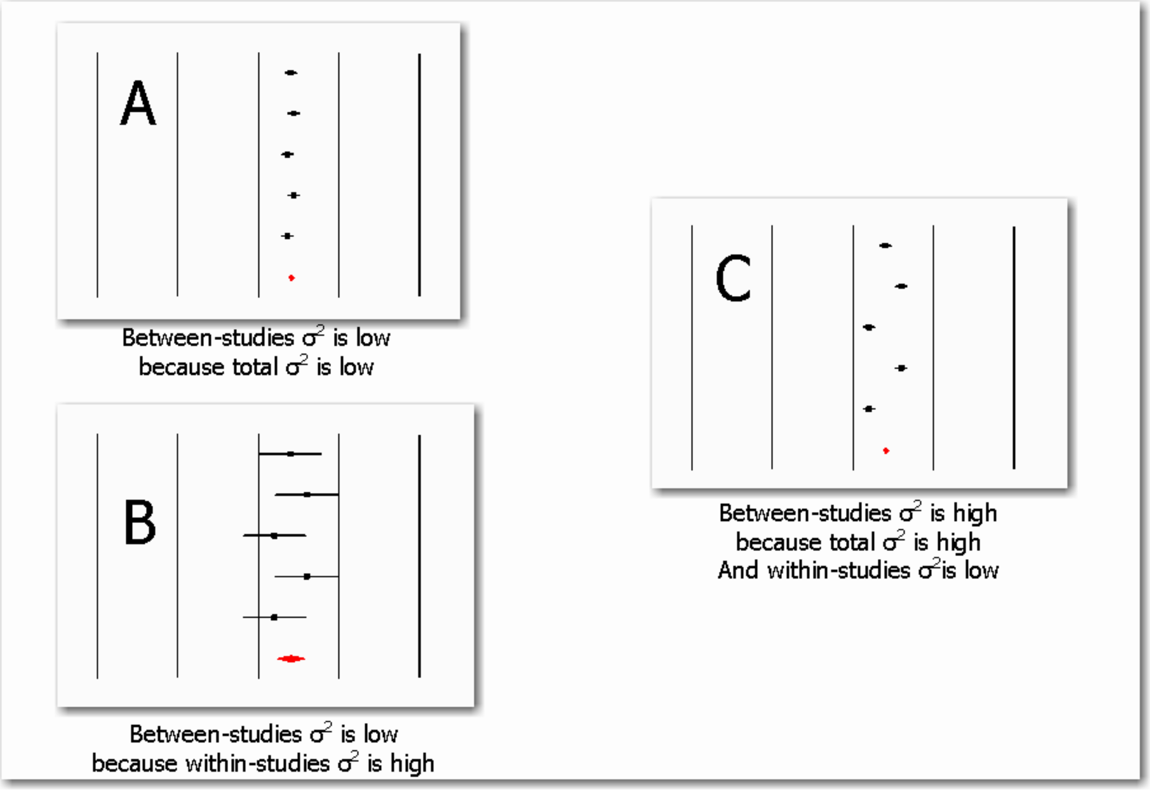
Assigning weights to the studies

Under the random effects model we need to take account of two levels of sampling, and two source of error. First, the true effect sizes θ are distributed about μ with a variance τ^2 that reflects the actual distribution of the true effects about their mean. Second, the observed effect T for any given θ will be distributed about that θ with a variance σ^2 that depends primarily on the sample size for that study. Therefore, in assigning weights to estimate *μ*, we need to deal with both sources of sampling error – within studies (*e*), and between studies (*ε*).

Decomposing the variance

The approach of a random effects analysis is to decompose the observed variance into its two component parts, within-studies and between-studies, and then use both parts when assigning the weights. The goal will be to reduce both sources of imprecision.

The mechanism used to decompose the variance is to compute the total variance (which is observed) and then to isolate the within-studies variance. The difference between these two values will give us the variance between-studies, which is called Tau-squared (τ^2). Consider the three graphs in the following figure.



Between-studies σ^2 is low
because total σ^2 is low

Between-studies σ^2 is low
because within-studies σ^2 is high

Between-studies σ^2 is high
because total σ^2 is high
And within-studies $\sigma^2$is low

In (A), the studies all line up pretty much in a row. There is no variance between studies, and therefore tau-squared is low (or zero).

In (B) there is variance between studies, but it is fully explained by the variance within studies. Put another way, given the imprecision of the studies, we would expect the effect size to vary somewhat from one study to the next. Therefore, the between-studies variance is again low (or zero).

In (C) there is variance between studies. And, it cannot be fully explained by the variance within studies, since the within-study variance is minimal. The excess variation (between-studies variance), will be reflected in the value of tau-squared.

It follows that tau-squared will increase as either the variance within-studies decreases and/or the observed variance increases.

This logic is operationalized in a series of formulas. We will compute *Q*, which represents the total variance, and *df*, which represents the expected variance if all studies have the same true effect. The difference, *Q* - *df*, will give us the excess variance. Finally, this value will be transformed, to put it into the same scale as the within-study variance. This last value is called Tau-squared.

The *Q* statistic represents the total variance and is defined as

$$Q = \sum_{i=1}^{k} w_i \left(T_i - \overline{T}_{.} \right)^2 \tag{3.2}$$

that is, the sum of the squared deviations of each study (T_i) from the combined mean ($\overline{T}.$). Note the "w_i" in the formula, which indicates that each of the squared deviations is weighted by the study's inverse variance. A large study that falls far from the mean will have more impact on *Q* than would a small study in the same location. An equivalent formula, useful for computations, is

$$Q = \sum_{i=1}^{k} w_i T_i^2 - \frac{\left(\sum_{i=1}^{k} w_i T_i \right)^2}{\sum_{i=1}^{k} w_i} \tag{3.3}$$

Since *Q* reflects the total variance, it must now be broken down into its component parts. If the only source of variance was within-study error, then the expected value of *Q* would be the degrees of freedom for the meta-analysis (*df*) where

$$df = (Number\ Studies) - 1 \tag{3.4}$$

This allows us to compute the between-studies variance, τ^2, as

$$T^2 = \begin{cases} \dfrac{Q - df}{C} & \text{if Q > df} \\ 0 & \text{if Q } \le \text{ df} \end{cases} \tag{3.5}$$

where

$$C = \sum w_i - \frac{\sum w_i^2}{\sum w_i} \tag{3.6}$$

The numerator, *Q - df*, is the excess (observed minus expected) variance. The denominator, *C*, is a scaling factor that has to do with the fact that *Q* is a weighted sum of squares. By applying this scaling factor we ensure that tau-squared is in the same metric as the variance within-studies.

In the running example,

$$Q = 53.208 - \left(\frac{101.833^2}{256.667}\right) = 12.8056$$

$$df = (6-1) = 5$$

$$C = 256.667 - \left(\frac{15522.222}{256.667}\right) = 196.1905$$

$$\boldsymbol{T}^2 = \frac{12.8056 - 5}{196.1905} = 0.0398$$

<u>Assigning weights under the random effects model</u>

In the fixed effect analysis each study was weighted by the inverse of its variance. In the random effects analysis, too, each study will be weighted by the inverse of its variance. The difference is that the variance now includes the original (within-studies) variance plus the between-studies variance, tau-squared.

Note the correspondence between the formulas here and those in the previous chapter. We use the same notations, but add a (*) to represent the random effects version. Concretely, under the random effects model the weight assigned to each study is

$$w_i^* = \frac{1}{v_i^*} \tag{3.7}$$

where v^*_i is the within-study variance for study (i) plus the between-studies variance, tau-squared. The weighted mean ($\bar{T}_{\bullet}*$) is then computed as

$$\bar{T}_{\bullet}* = \frac{\sum_{i=1}^{k} w_i^* T_i}{\sum_{i=1}^{k} w_i^*} \tag{3.8}$$

that is, the sum of the products (effect size multiplied by weight) divided by the sum of the weights.

The variance of the combined effect is defined as the reciprocal of the sum of the weights, or

$$v_{.}^* = \frac{1}{\sum_{i=1}^{k} w_i^*} \tag{3.9}$$

and the standard error of the combined effect is then the square root of the variance,

$$SE(\bar{T}_{\bullet}*) = \sqrt{v_{\bullet}*} \tag{3.10}$$

The 95% confidence interval for the combined effect would be computed as

$$Lower\ Limit* = \bar{T}_{\bullet}* - 1.96 * SE(\bar{T}_{\bullet}*) \tag{3.11}$$

$$Upper\ Limit* = \bar{T}_{\bullet}* + 1.96 * SE(\bar{T}_{\bullet}*) \tag{3.12}$$

Finally, if one were so inclined, the Z-value could be computed using

$$Z^* = \frac{\bar{T}_{\bullet}*}{SE(\bar{T}_{\bullet}*)} \tag{3.13}$$

The one-tailed *p*-value (assuming an effect in the hypothesized direction) is given by

$$p^{*} = 1 - \Phi\left(| Z^{*} |\right) \tag{3.14}$$

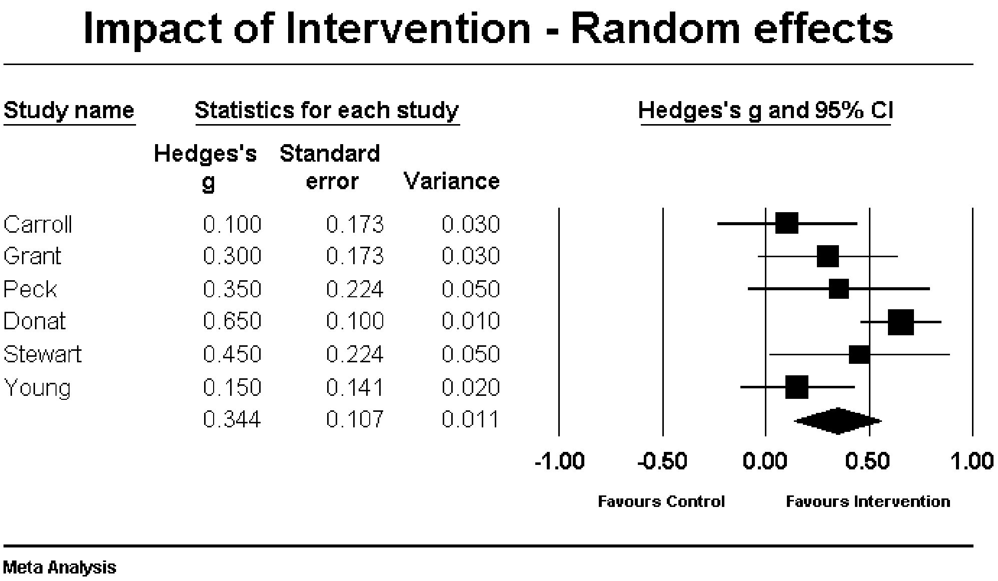
and the two-tailed *p*-value by

$$p^{*} = 2\left[1 - \Phi\left(| Z^{*} |\right)\right] \tag{3.15}$$

Where Φ is the standard normal cumulative distribution function.

Illustrative example

The following figure is based on the same studies we used for the fixed effect example.



Impact of Intervention - Random effects

Study name	Hedges's g	Standard error	Variance
Carroll	0.100	0.173	0.030
Grant	0.300	0.173	0.030
Peck	0.350	0.224	0.050
Donat	0.650	0.100	0.010
Stewart	0.450	0.224	0.050
Young	0.150	0.141	0.020
	0.344	0.107	0.011

Note the differences from the fixed effect model.

- The weights are more balanced. The boxes for the large studies such as Donat have decreased in size while those for the small studies such as Peck have increase in size.

- The combined effect has moved toward the left, from 0.40 to 0.34. This reflects the fact that the impact of Donat (on the right) has been reduced.
- The confidence interval for the combined effect has increased in width.

In the running example the weight for the Carroll study would be computed as

$$w_i^* = \frac{1}{(0.030 + 0.040)} = \frac{1}{(0.070)} = 14.330$$

and so on for the other studies. Then,

$$\bar{T}_\bullet * = \frac{30.207}{87.747} = 0.3442$$

$$v_\bullet * \frac{1}{87.747} = 0.0114$$

$$SE(\bar{T}_\bullet *) = \sqrt{0.0114} = 0.1068$$

$$Lower\ Limit * = 0.3442 - 1.96 * 0.1068 = 0.1350$$

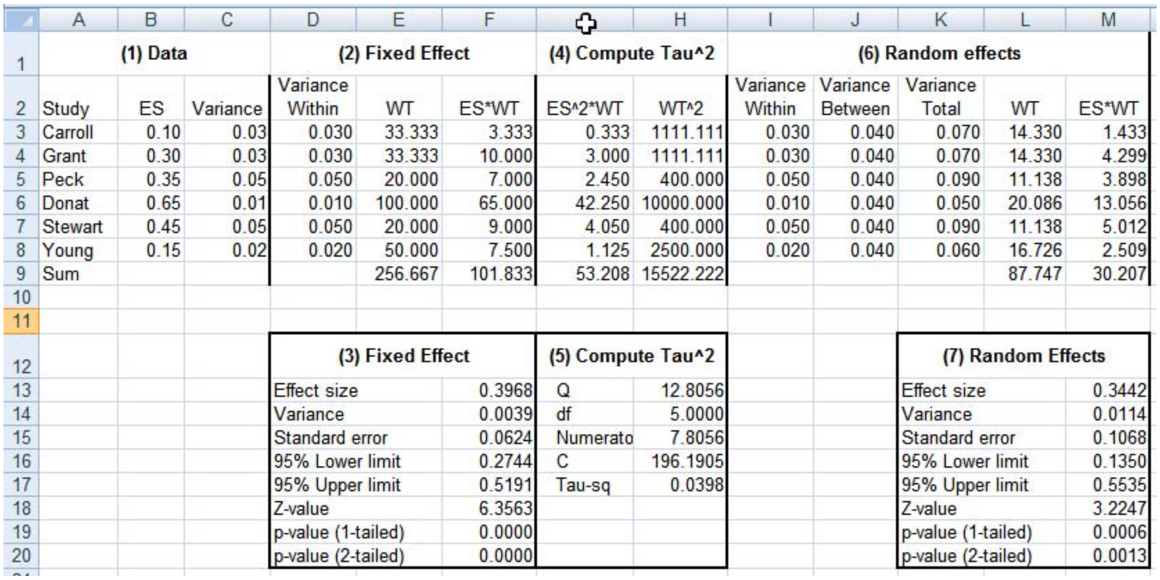
$$Upper\ Limit * = 0.3968 - 1.96 * 0.1068 = 0.5535$$

$$Z* = \frac{0.3442}{0.1068} = 3.2247$$

$$P_{1T} = 1 - \left(\Phi\left(ABS(3.2247)\right)\right) = 0.0006$$

$$P_{2T} = \left[1 - \left(\Phi\left(ABS(3.2247)\right)\right)\right] * 2 = 0.0013$$

These formulas are incorporated in the following spreadsheet



	A	B	C	D	E	F	G	H	I	J	K	L	M
1		(1) Data		(2) Fixed Effect			(4) Compute Tau^2		(6) Random effects				
2	Study	ES	Variance	Variance Within	WT	ES*WT	ES^2*WT	WT^2	Variance Within	Variance Between	Variance Total	WT	ES*WT
3	Carroll	0.10	0.03	0.030	33.333	3.333	0.333	1111.111	0.030	0.040	0.070	14.330	1.433
4	Grant	0.30	0.03	0.030	33.333	10.000	3.000	1111.111	0.030	0.040	0.070	14.330	4.299
5	Peck	0.35	0.05	0.050	20.000	7.000	2.450	400.000	0.050	0.040	0.090	11.138	3.898
6	Donat	0.65	0.01	0.010	100.000	65.000	42.250	10000.000	0.010	0.040	0.050	20.086	13.056
7	Stewart	0.45	0.05	0.050	20.000	9.000	4.050	400.000	0.050	0.040	0.090	11.138	5.012
8	Young	0.15	0.02	0.020	50.000	7.500	1.125	2500.000	0.020	0.040	0.060	16.726	2.509
9	Sum				256.667	101.833	53.208	15522.222				87.747	30.207
10													
11													
12				(3) Fixed Effect			(5) Compute Tau^2				(7) Random Effects		
13				Effect size		0.3968	Q	12.8056			Effect size		0.3442
14				Variance		0.0039	df	5.0000			Variance		0.0114
15				Standard error		0.0624	Numerato	7.8056			Standard error		0.1068
16				95% Lower limit		0.2744	C	196.1905			95% Lower limit		0.1350
17				95% Upper limit		0.5191	Tau-sq	0.0398			95% Upper limit		0.5535
18				Z-value		6.3563					Z-value		3.2247
19				p-value (1-tailed)		0.0000					p-value (1-tailed)		0.0006
20				p-value (2-tailed)		0.0000					p-value (2-tailed)		0.0013

This spreadsheet builds on the spreadsheet for a fixed effect analysis. Columns A-F are identical to those in that spreadsheet. Here, we add columns for tau-squared (columns G-H) and random effects analysis (columns I-M).

Note that the formulas for fixed effect and random effects analyses are identical, the only difference being the definition of the variance. For the fixed effect analysis the variance (Column D) is defined as the variance within-studies (for example D3=C3). For the random effects analysis the variance is defined as the variance within-studies plus the variance between-studies (for example, K3=I3+J3).

Column (Cell)	Label	Content	Excel Formula*	See formula
(Section 1) Effect size and weights for each study				
A	Study name		Entered	
B	Effect size		Entered	
C	Variance		Entered	
(Section 2) Compute fixed effect WT and WT*ES for each study				
D	Variance within study		=$C3	
E	Weight		=1/D3	(2.3)
F	ES*WT		=$B3*E3	
Sum the columns				
E9	Sum of WT		=SUM(E3:E8)	
F9	Sum of WT*ES		=SUM(F3:F8)	
(Section 3) Compute combined effect and related statistics for fixed effect model				
F13	Effect size		=F9/E9	(2.4)
F14	Variance		=1/E9	(2.5)
F15	Standard error		=SQRT(F14)	(2.6)
F16	95% lower limit		=F13-1.96*F15	(2.7)
F17	95% upper limit		=F13+1.96*F15	(2.8)
F18	Z-value		=F13/F15	(2.9)
F19	p-value (1-tailed)		=(1-(NORMDIST(ABS(F18),0,1,TRUE)))	(2.10)
F20	p-value (2-tailed)		=(1-(NORMDIST(ABS(F18),0,1,TRUE)))*2	(2.11)
(Section 4) Compute values needed for Tau-squared				
G3	ES^2*WT		=B3^2*E3	
H3	WT^2		=E3^2	
Sum the columns				
G9	Sum of ES^2*WT		=SUM(G3:G8)	
H9	Sum of WT^2		=SUM(H3:H8)	
(Section 5) Compute Tau-squared				
H13	Q		=G9-F9^2/E9	(3.3)
H14	Df		=COUNT(B3:B8)-1	(3.4)
H15	Numerator		=MAX(H13-H14,0)	
H16	C		=E9-H9/E9	(3.6)
H17	Tau-sq		=H15/H16	(3.5)
(Section 6) Compute random effects WT and WT*ES for each study				
I3	Variance within		=$C3	
J3	Variance between		=$H$17	
K3	Variance total		=I3+J3	
L3	WT		=1/K3	(3.7)
M3	ES*WT		=$B3*L3	
Sum the columns				
L9	Sum of WT		=SUM(L3:L8)	
M9	Sum of ES*WT		=SUM(M3:M8)	
(Section 7) Compute combined effect and related statistics for random effects model				

M13	Effect size	=M9/L9	(3.8)
M14	Variance	=1/L9	(3.9)
M15	Standard error	=SQRT(M14)	(3.10)
M16	95% lower limit	=M13-1.96*M15	(3.11)
M17	95% upper limit	=M13+1.96*M15	(3.12)
M18	Z-value	=M13/M15	(3.13)
M19	p-value (1-tailed)	=(1-(NORMDIST(ABS(M18),0,1,TRUE)))	(3.14)
M20	p-value (2-tailed)	=(1-(NORMDIST(ABS(M18),0,1,TRUE)))*2	(3.15)

Fixed effect vs. random effects models

In the previous two chapters we outlined the two basic approaches to meta-analysis – the Fixed effect model and the Random effects model. This chapter will discuss the differences between the two.

The concept

The fixed effect and random effects models represent two conceptually different approaches.

Fixed effect

The fixed effect model assumes that all studies in the meta-analysis are drawn from a common population. Put another way, all factors which could influence the effect size are the same in all the study populations, and therefore the effect size is the same in all the study populations. It follows that the observed effect size varies from one study to the next only because of the random error inherent in each study.

Random effects

By contrast, the random effects model assumes that the studies were drawn from populations that differ from each other in ways that could impact on the treatment effect. For example, the intensity of the intervention or the age of the subjects may have varied from one study to the next. It follows that the effect size will vary from one study to the next for two reasons. The first is random error within studies, as in the fixed effect model. The second is true variation in effect size from one study to the next.

Definition of a combined effect

The meaning of the "combined effect" is different for fixed effect vs. random effects analyses.

Fixed effect

Under the fixed effect model there is one true effect size. It follows that the combined effect is our estimate of this value.

Random effects

Under the random effects model there is not one true effect size, but a distribution of effect sizes. It follows that the combined estimate is not an estimate of one value, but rather is meant to be the average of a distribution of values.

Computing the combined effect

These differences in the definition of the combined effect lead to differences in the way the combined effect is computed.

Fixed effect

Under the fixed effect model we assume that the true effect size for all studies is identical, and the only reason the effect size varies between studies is random error. Therefore, when assigning weights to the different studies we can largely ignore the information in the smaller studies since we have better information about the same effect size in the larger studies.

Random effects

By contrast, under the random effects model the goal is not to estimate one true effect, but to estimate the mean of a distribution of effects. Since each study provides information about an effect size <u>in a different population</u>, we want to be sure that all the populations captured by the various studies are represented in the combined estimate.

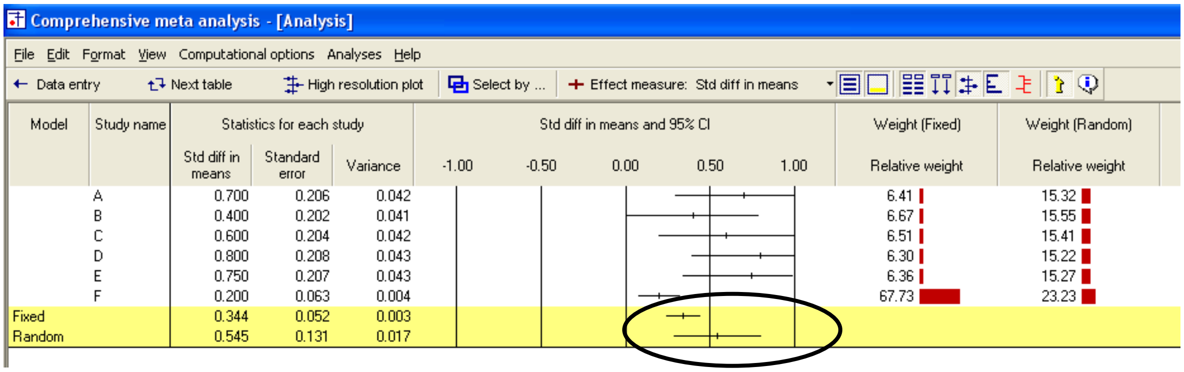
This means that we cannot discount a small study by giving it a very small weight (the way we would in a fixed effect analysis). The estimate provided by that study may be imprecise, but it is information about a population that no other study has captured. By the same logic we cannot give too much weight to a very large study (the way we might in a fixed effect analysis). Our goal is to estimate the effects in a range of populations, and we do not want that overall estimate to be overly influenced by any one population.

Extreme effect size in large study

How will the selection of a model influence the overall effect size? Consider the case where there is an extreme effect in a large study. Here, we have five small studies (Studies A-E, with 100 subjects per study) and one large study (Study F, with 1000 subjects). The confidence interval for each of the studies A-E is wide, reflecting relatively poor precision, while the confidence interval for Study F is narrow, indicating greater precision. In this example the small studies all have relatively large effects (in the range of 0.40 to 0.80) while the large study has a relatively small effect (0.20).

Fixed effect

Under the fixed effect model these studies are all estimating the same effect size, and the large study (F) provides a more precise estimate of this effect size. Therefore, this study is assigned 68% of the weight in the combined effect, with each of the remaining studies being assigned about 6% of the weight (see the column labeled “Relative weight” under fixed effects.


Comprehensive meta analysis - [Analysis]
File Edit Format View Computational options Analyses Help
Data entry
Next table
High resolution plot
Select by ...
Effect measure: Std diff in means
-1.00
-0.50
0.00
0.50
1.00

Model	Study name	Statistics for each study			Std diff in means and 95% CI	Weight (Fixed)	Weight (Random)
		Std diff in means	Standard error	Variance		Relative weight	Relative weight
	A	0.700	0.206	0.042		6.41	15.32
	B	0.400	0.202	0.041		6.67	15.55
	C	0.600	0.204	0.042		6.51	15.41
	D	0.800	0.208	0.043		6.30	15.22
	E	0.750	0.207	0.043		6.36	15.27
	F	0.200	0.063	0.004		67.73	23.23
Fixed		0.344	0.052	0.003			
Random		0.545	0.131	0.017			

Because Study F is assigned so much of the weight it “pulls” the combined estimate toward itself. Study F had a smaller effect than the other studies and so it pulls the combined estimate toward the left. On the graph, note the point estimate for the large study (Study F, with d=.2), and how it has “pulled” the fixed effect estimate down to 0.34 (see the shaded row marked “Fixed” at the bottom of the plot).

Random effects

By contrast, under the random effects model these studies are drawn from a range of populations in which the effect size varies and our goal is to summarize this range of effects. Each study is estimating an effect size for its unique population, and so each must be given appropriate weight in the analysis. Now, Study F is assigned only 23% of the weight (rather than 68%), and each of the small studies is given about 15% of the weight (rather than 6%) (see the column labeled “Relative weights” under random effects.

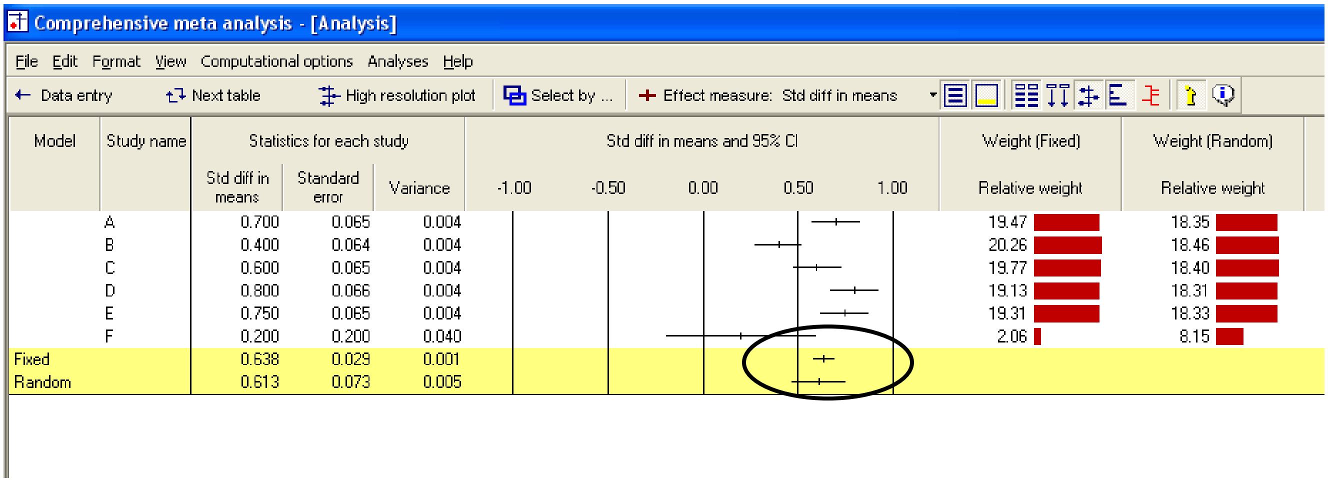
What happens to our estimate of the combined effect when we weight the studies this way? The overall effect is still being pulled by the large study, but not as much as before. In the plot, the bottom two lines reflect the fixed effect and random effect estimates, respectively. Compare the point estimate for “Random” (the last line) with the one for “Fixed” just above it. The overall effect is now 0.55 (which is much closer to the range of the small studies) rather than 0.34 (as it was for the fixed effect model). The impact of the large study is now less pronounced.

Extreme effect size in small study

Now, let’s consider the reverse situation: The effect sizes for each study is the same as in the prior example, but this time the first 5 studies are large while the sixth study is small. Concretely, we have five large studies (A-E, with 1000 subjects per study) and one small study (F, with 100 subjects). On the graphic, the confidence intervals for studies A-E are each relatively narrow, indicating high precision, while that for Study F is relatively wide, indicating less precision. The large studies all have relatively large effects (in the range of 0.40 to 0.80) while the small study has a relatively small effect (0.20).

Fixed effect

Under the fixed effect model the large studies (A-E) are each given is given about 20% of the weight, while the small study (F) is given only about 2% of the weight (see column labeled “Relative weights” under Fixed effect). This follows from the logic of the fixed effect model. The larger studies provide a good estimate of the common effect, and the small study offers a less reliable estimate of that same effect, so it is assigned a small (in this case trivial) weight. With only 2% of the weight Study F has little impact on the combined value, which is computed as 0.64.


Comprehensive meta analysis - [Analysis]
File Edit Format View Computational options Analyses Help
Data entry
Next table
High resolution plot
Select by ...
Effect measure: Std diff in means
Std diff in means and 95% CI
-1.00
-0.50
0.00
0.50
1.00

Model	Study name	Statistics for each study			Weight (Fixed)	Weight (Random)
		Std diff in means	Standard error	Variance	Relative weight	Relative weight
	A	0.700	0.065	0.004	19.47	18.35
	B	0.400	0.064	0.004	20.26	18.46
	C	0.600	0.065	0.004	19.77	18.40
	D	0.800	0.066	0.004	19.13	18.31
	E	0.750	0.065	0.004	19.31	18.33
	F	0.200	0.200	0.040	2.06	8.15
Fixed		0.638	0.029	0.001		
Random		0.613	0.073	0.005		

Random effects

By contrast, under the random effects model each study is estimating an effect size for its unique population, and so each must be given appropriate weight in the analysis. As shown in the column "Relative weights" under random effects each of the large studies (A-E) is now given about 18% of the weight (rather than 20%) while the small study (F) receives 8% of the weight (rather than 2%).

What happens to our estimate of the combined effect when we weight the studies this way? Where the small study has almost no impact under the fixed effect model, it now has a substantially larger impact effect. Concretely, it gets 8% of the weight, which is nearly half the weight assigned to any of the larger studies (18%).

The small study therefore has more of an impact now than it did under the fixed effect model. Where it was assigned only 2% of the weight before, it is now assigned 8% of the weight. This is 50% of the weight assigned to studies A-E, and as such is no longer a trivial amount. Compare the two lines labeled "Fixed" and "Random" at the bottom of the plot. The overall effect is now 0.61, which is .03 points closer to study F than it had been under the fixed effect model (0.64).

Summary

The operating premise, as illustrated in these examples, is that the relative weights assigned under random effects will be more balanced than those assigned under fixed effects. As we move from fixed effect to random effects, extreme studies will lose influence if they are large, and will gain influence if they are small.

In these two examples we included a single study with an extreme size and an extreme effect, to highlight the difference between the two weighting schemes. In most analyses, of course, there will be a range of sample sizes within studies and the larger (or smaller) studies could fall anywhere in this range. Nevertheless, the same principle will hold.

Confidence interval width

Above, we considered the impact of the model (fixed vs. random effects) on the combined effect size. Now, let's consider the impact on the width of the confidence interval.

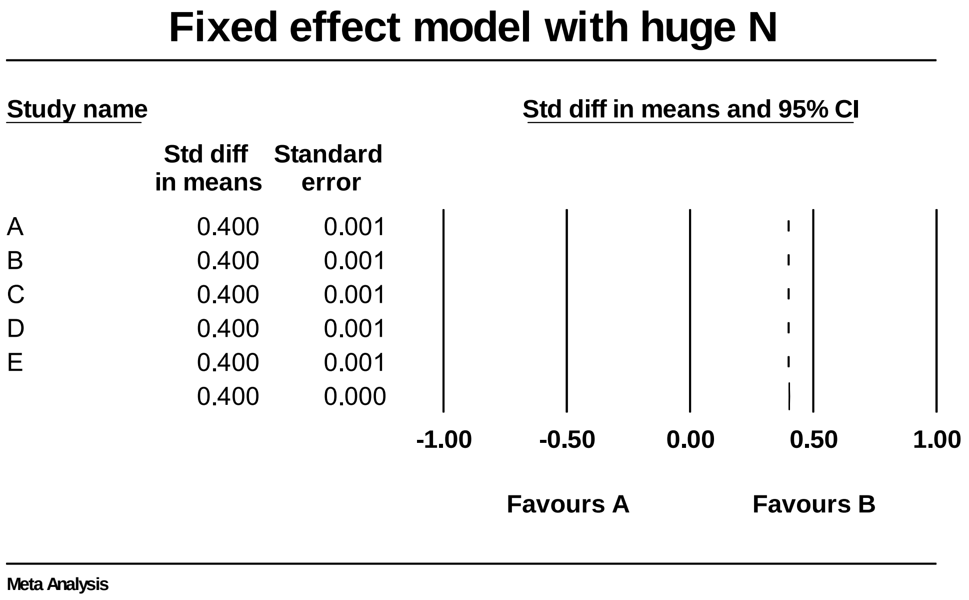
Recall that the fixed effect model defines “variance” as the variance within a study, while the random effects model defines it as variance within a study plus variance between studies. To understand how this difference will affect the width of the confidence interval, let’s consider what would happen if all studies in the meta-analysis were of infinite size, which means that the within-study error is effectively zero.

Fixed effect

Since we’ve started with the assumption that all variation is due to random error, and this error has now been removed, it follows that

- The observed effects would all be identical.
- The combined effect would be exactly the same as each of the individual studies.
- The width of the confidence interval for the combined effect would approach zero.

All of these points can be seen in the figure. In particular, note that the diamond representing the combined effect, has a width of zero, since the width of the confidence interval is zero.



Fixed effect model with huge N

Study name	Std diff in means	Standard error
A	0.400	0.001
B	0.400	0.001
C	0.400	0.001
D	0.400	0.001
E	0.400	0.001
	0.400	0.000

Generally, we are concerned with the precision of the combined effect rather than the precision of the individual studies. For this purpose it doesn't matter whether the sample is concentrated in one study or dispersed among many studies. In either case, as the total N approaches infinity the errors will cancel out and the standard error will approach zero.

Random effects

Under the random effects model the effect size for each study would still be known precisely. However, the effects would not line up in a row since the true treatment effect is assumed to vary from one study to the next. It follows that –

- The within-study error would approach zero, and the confidence interval for each study would approach zero.

- Since the studies are all drawn from different populations, even though the effects are now being estimated without error, the observed effects would not be identical to each other.

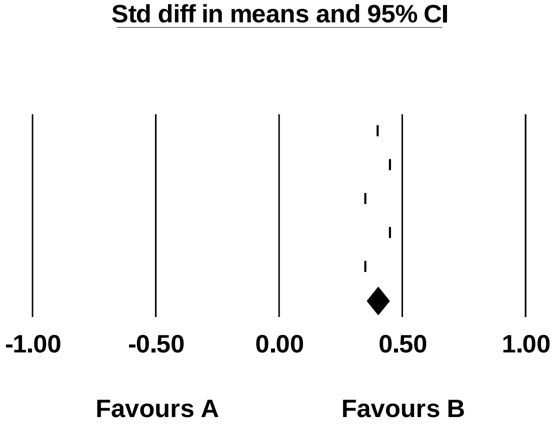
- The width of the confidence interval for the combined effect would not approach zero unless the number of studies approached infinitity.

Generally, we are concerned with the precision of the combined effect rather than the precision of the individual studies. Unlike the situation for the fixed effect model, here it does matter whether the sample is concentrated in one study or dispersed among many studies. We need an infinite N within each study for the standard error of that study to approach zero. Additionally, we need an infinite number of studies for the standard error in estimating μ from θ_i to approach zero. In our example we know the value of the five effects precisely, but these are only a random sample of all possible effects, and so there is substantial error in our estimate of the combined effect.

Note. While the distribution of the θ_i about μ represents a real distribution of effect sizes we refer to this as error, since it introduces error into our estimate of the combined effect. If the studies that we do observe tend to cluster closely together and/or our meta-analysis includes large number of studies, this source of error will tend to be small. If the studies that we do observe show much dispersion and/or we have only a small sample of studies, then this source of error will tend to be large.

Random effects model with huge N

Study name	Std diff in means	Standard error
A	0.400	0.001
B	0.450	0.001
C	0.350	0.001
D	0.450	0.001
E	0.350	0.001
	0.400	0.022



Meta Analysis

Summary

Since the variation under random effects incorporates the same error as fixed effects plus an additional component, it cannot be less than the variation under fixed effect model. As long as the between-studies variation is non-zero, the variance, standard error, and confidence interval will always be larger under random effects.

The standard error of the combined effect in both models is inversely proportional to the number of studies. Therefore, in both models, the width of the confidence interval tends toward zero as the number of studies increases. In the case of the fixed effect model the standard error and the width of the confidence interval can tend toward zero even with a finite number of studies if any of the studies is sufficiently large. By contrast, for the random effects model, the confidence interval can tend toward zero only with an infinite number of studies (unless the between-study variation is zero).

Which model should we use?

The selection of a computational model should be based on the nature of the studies and our goals.

Fixed effect

The fixed effect model makes sense if (a) there is reason to believe that all the studies are functionally identical, and (b) our goal is to compute the common effect size, which would then be generalized to other examples of this same population.

For example, assume that a drug company has run five studies to assess the effect of a drug. All studies recruited patients in the same way, used the same researchers, dose, and so on, so all are expected to have the identical effect (as though this were one large study, conducted with a series of cohorts). Also, the regulatory agency wants to see if the drug works in this one population. In this example, a fixed effect model makes sense.

Random effects

By contrast, when the researcher is accumulating data from a series of studies that had been performed by other people, it would be very unlikely that all the studies were functionally equivalent. Almost invariably, the subjects or interventions in these studies would have differed in ways that would have impacted on the results, and therefore we should not assume a common effect size.

Additionally, the goal of this analysis is usually to generalize to a range of populations. Therefore, if one did make the argument that all the studies used an identical, narrowly defined population, then it would not be possible to extrapolate from this population to others, and the utility of the analysis would be limited.

Therefore, the random effects model is more easily justified in most common cases.

Note

If the number of studies is very small, then it may be impossible to estimate the between-studies variance (tau-squared) with any precision. In this case, the fixed effect model may be the only viable option. Alternatively, one could do a sensitivity analysis by plugging in several values of tau-square.

Mistakes to avoid in selecting a model

Some have adopted the practice of starting with the fixed effect model and then moving to a random effects model if *Q* is statistically significant. This practice should be discouraged for the following reasons.

- If the logic of the analysis says that we are trying to estimate a range of effects, then the random effects formula, which addresses this goal, is the logical formula to use.

- If the actual dispersion turns out to be trivial (that is, less than expected under the hypothesis of homogeneity), then the random effects model will reduce to the fixed effect model. Therefore, there is no "cost" to using the random effects model in this case.

- If the actual dispersion turns out to be non-trivial, then this dispersion should be incorporated in the analysis, which the random effects model does, and the fixed effect model does not. That the *Q* statistic meets or does not meet a criterion for significance is simply not relevant.

The last statement above would be true even if the *Q* test did a good job of identifying dispersion. In fact, though, if the number of studies is small and the within-studies variance is large, the test based on the *Q* statistic may have low power even if the between-study variance is substantial. In this case, using the *Q* test as a criterion for selecting the model is problematic not only from a conceptual perspective, but could allow the use of a fixed effect analysis in cases with substantial dispersion.