7.7.3.2 Obtaining standard deviations from standard errors and confidence intervals for group means

A standard deviation can be obtained from the standard error of a mean by multiplying by the square root of the sample size:

$$SD = SE \times \sqrt{N}$$

When making this transformation, standard errors must be of means calculated from within an intervention group and not standard errors of the difference in means computed between intervention groups.

Confidence intervals for means can also be used to calculate standard deviations. Again, the following applies to confidence intervals for mean values calculated within an intervention group and not for estimates of differences between interventions (for these, see Section 7.7.3.3). Most confidence intervals are 95% confidence intervals. If the sample size is large (say bigger than 100 in each group), the 95% confidence interval is 3.92 standard errors wide (3.92 = 2 × 1.96). The standard deviation for each group is obtained by dividing the length of the confidence interval by 3.92, and then multiplying by the square root of the sample size:

$$SD = \sqrt{N} \times (\text{upper limit} - \text{lower limit}) / 3.92$$

For 90% confidence intervals 3.92 should be replaced by 3.29, and for 99% confidence intervals it should be replaced by 5.15.

If the sample size is small (say less than 60 in each group) then confidence intervals should have been calculated using a value from a t distribution. The numbers 3.92, 3.29 and 5.15 need to be replaced with slightly larger numbers specific to the t distribution, which can be obtained from tables of the t distribution with degrees of freedom equal to the group sample size minus 1. Relevant details of the t distribution are available as appendices of many statistical textbooks, or using standard computer spreadsheet packages. For example the t value for a 95% confidence interval from a sample size of 25 can be obtained by typing =**tinv(1-0.95,25-1)** in a cell in a Microsoft Excel spreadsheet (the result is 2.0639). The divisor, 3.92, in the formula above would be replaced by 2 × 2.0639 = 4.128.

For moderate sample sizes (say between 60 and 100 in each group), either a t distribution or a standard normal distribution may have been used. Review authors should look for evidence of which one, and might use a t distribution if in doubt.

As an example, consider data presented as follows:

Group	Sample size	Mean	95% CI
Experimental intervention	25	32.1	(30.0, 34.2)
Control intervention	22	28.3	(26.5, 30.1)

The confidence intervals should have been based on t distributions with 24 and 21 degrees of freedom respectively. The divisor for the experimental intervention group is 4.128, from above. The standard deviation for this group is √25 × (34.2 – 30.0)/4.128 = 5.09. Calculations for the control group are performed in a similar way.

It is important to check that the confidence interval is symmetrical about the mean (the distance between the lower limit and the mean is the same as the distance between the mean and the upper limit). If this is not the case, the confidence interval may have been calculated on transformed values (see Section 7.7.3.4).

7.7.3.3 Obtaining standard deviations from standard errors, confidence intervals, t values and P values for differences in means

Standard deviations can be obtained from standard errors, confidence intervals, t values or P values that relate to the differences between means in two groups. The difference in means itself (MD) is required in the calculations from the t value or the P value. An assumption that the standard deviations of outcome measurements are the same in both groups is required in all cases, and the standard deviation would then be used for both intervention groups. We describe first how a t value can be obtained from a P value, then how a standard error can be obtained from a t value or a confidence interval, and finally how a standard deviation is obtained from the standard error. Review authors may select the appropriate steps in this process according to what results are available to them. Related methods can be used to derive standard deviations from certain F statistics, since taking the square root of an F value may produce the same t value. Care is often required to ensure that an appropriate F value is used, and advice of a knowledgeable statistician is recommended.

From P value to t value

Where actual P values obtained from t-tests are quoted, the corresponding t value may be obtained from a table of the t distribution. The degrees of freedom are given by $N_E + N_C - 2$, where $N_E$and N_C are the sample sizes in the experimental and control groups. We will illustrate with an example. Consider a trial of an experimental intervention ($N_E = 25$) versus a control intervention ($N_C = 22$), where the difference in means was MD = 3.8. It is noted that the P value for the comparison was P = 0.008, obtained using a two-sample t-test.

The t value that corresponds with a P value of 0.008 and 25+22–2=45 degrees of freedom is t = 2.78. This can be obtained from a table of the t distribution with 45 degrees of freedom or a computer (for example, by entering =**tinv(0.008, 45)** into any cell in a Microsoft Excel spreadsheet).

Difficulties are encountered when levels of significance are reported (such as P<0.05 or even P=NS which usually implies P>0.05) rather than exact P values. A conservative approach would be to take the P value at the upper limit (e.g. for P<0.05 take P=0.05, for P<0.01 take P=0.01 and for P<0.001 take P=0.001). However, this is not a solution for results which are reported as P=NS: see Section 7.7.3.7.

From t value to standard error

The t value is the ratio of the difference in means to the standard error of the difference in means. The standard error of the difference in means can therefore be obtained by dividing the difference in means (MD) by the t value:

$$SE = \frac{MD}{t}.$$

In the example, the standard error of the difference in means is obtained by dividing 3.8 by 2.78, which gives 1.37.

From confidence interval to standard error

If a 95% confidence interval is available for the difference in means, then the same standard error can be calculated as:

$$SE = (\text{upper limit} - \text{lower limit})/3.92$$

as long as the trial is large. For 90% confidence intervals 3.92 should be replaced by 3.29, and for 99% confidence intervals it should be replaced by 5.15. If the sample size is small then confidence intervals should have been calculated using a t distribution. The numbers 3.92, 3.29 and 5.15 need to be replaced with larger numbers specific to both the t distribution and the sample size, and can be obtained from tables of the t distribution with degrees of freedom equal to $N_E + N_C - 2$, where N_E and

N_C are the sample sizes in the two groups. Relevant details of the t distribution are available as appendices of many statistical textbooks, or using standard computer spreadsheet packages. For example, the t value for a 95% confidence interval from a comparison of a sample size of 25 with a sample size of 22 can be obtained by typing =**tinv(1-0.95,25+22-2)** in a cell in a Microsoft Excel spreadsheet.

From standard error to standard deviation

The within-group standard deviation can be obtained from the standard error of the difference in means using the following formula:

$$SD = \frac{SE}{\sqrt{\frac{1}{N_E} + \frac{1}{N_C}}}$$

In the example,

$$SD = \frac{1.37}{\sqrt{\frac{1}{25} + \frac{1}{22}}} = 4.69.$$

Note that this standard deviation is the average of the standard deviations of the experimental and control arms, and should be entered into RevMan twice (once for each intervention group).

16.1.3.1 Imputing standard deviations

Missing standard deviations are a common feature of meta-analyses of continuous outcome data. One approach to this problem is to impute standard deviations. Before imputing missing standard deviations however, authors should look carefully for statistics that allow calculation or estimation of the standard deviation (e.g. confidence intervals, standard errors, t values, P values, F values), as discussed in Chapter 7 (Section 7.7.3).

The simplest imputation is of a particular value borrowed from one or more other studies. Furukawa et al. found that imputing standard deviations either from other studies in the same meta-analysis, or from studies in another meta-analysis, yielded approximately correct results in two case studies (Furukawa 2006). If several candidate standard deviations are available, review authors would have to decide whether to use their average, the highest, a 'reasonably high' value, or some other strategy. For meta-analyses of mean differences, choosing a higher standard deviation down-weights a study and yields a wider confidence interval. However, for standardized mean difference meta-analyses, choice of an overly large standard deviation will bias the result towards a lack of effect. More complicated alternatives are available for making use of multiple candidate standard deviations. For example, Marinho et al. implemented a linear regression of log(standard deviation) on log(mean), because of a strong linear relationship between the two (Marinho 2003).

All imputation techniques involve making assumptions about unknown statistics, and it is best to avoid using them wherever possible. If the majority of studies in a meta-analysis have missing standard deviations, these values should not be imputed. However, imputation may be reasonable for a small proportion of studies comprising a small proportion of the data if it enables them to be combined with other studies for which full data are available. Sensitivity analyses should be used to assess the impact of changing the assumptions made.

16.1.3.2 Imputing standard deviations for changes from baseline

A special case of missing standard deviations is for changes from baseline. Often, only the following information is available:

	Baseline	Final	Change
Experimental intervention (sample size)	mean, SD	mean, SD	mean
Control intervention (sample size)	mean, SD	mean, SD	mean

Note that the mean change in each group can always be obtained by subtracting the final mean from the baseline mean even if it is not presented explicitly. However, the information in this table does *not* allow us to calculate the standard deviation of the changes. We cannot know whether the changes were very consistent or very variable. Some other information in a paper may help us determine the standard deviation of the changes. If statistical analyses comparing the changes themselves are presented (e.g. confidence intervals, standard errors, t values, P values, F values) then the techniques described in Chapter 7 (Section 7.7.3) may be used.

When there is not enough information available to calculate the standard deviations for the changes, they can be imputed. When change-from-baseline standard deviations for the same outcome measure are available from other studies in the review, it may be reasonable to use these in place of the missing standard deviations. However, the appropriateness of using a standard deviation from another study relies on whether the studies used the same measurement scale, had the same degree of measurement error and had the same time periods (between baseline and final value measurement).

The following alternative technique may be used for imputing missing standard deviations for changes from baseline (Follmann 1992, Abrams 2005). A typically unreported number known as the correlation coefficient describes how similar the baseline and final measurements were across participants. Here we describe (1) how to calculate the correlation coefficient from a study that is reported in considerable detail and (2) how to impute a change-from-baseline standard deviation in another study, making use of an imputed correlation coefficient. Note that the methods in (2) are applicable both to correlation coefficients obtained using (1) and to correlation coefficients obtained in other ways (for example, by reasoned argument). These methods should be used sparingly, because one can never be sure that an imputed correlation is appropriate (correlations between baseline and final values will, for example, decrease with increasing time between baseline and final measurements, as well as depending on the outcomes and characteristics of the participants). An alternative to these methods is simply to use a comparison of final measurements, which in a randomized trial in theory estimates the same quantity as the comparison of changes from baseline.

(1) Calculating a correlation coefficient from a study reported in considerable detail

Suppose a study is available that presents means and standard deviations for change as well as for baseline and final measurements, for example:

	Baseline	Final	Change
Experimental intervention (sample size 129)	mean=15.2 SD=6.4	mean=16.2 SD=7.1	mean=1.0 SD=4.5
Control intervention (sample size 135)	mean=15.7 SD=7.0	mean=17.2 SD=6.9	mean=1.5 SD=4.2

An analysis of change from baseline is available from this study, using only the data in the final column. However, we can use the other data from the study to calculate two correlation coefficients, one for each intervention group. Let us use the following notation:

	Baseline	Final	Change
Experimental intervention (sample size N_E)	$M_{E,baseline}$, $SD_{E,baseline}$	$M_{E,final}$, $SD_{E,final}$	$M_{E,change}$, $SD_{E,change}$
Control intervention (sample size N_C)	$M_{C,baseline}$, $SD_{C,baseline}$	$M_{C,final}$, $SD_{C,final}$	$M_{C,change}$, $SD_{C,change}$

The correlation coefficient in the experimental group, $Corr_E$, can be calculated as:

$$\text{Corr}_E = \frac{SD^2_{E,baseline} + SD^2_{E,final} - SD^2_{E,change}}{2 \times SD_{E,baseline} \times SD_{E,final}};$$

and similarly for the control intervention, to obtain $Corr_C$. In the example, these turn out to be

$$\text{Corr}_E = \frac{6.4^2 + 7.1^2 - 4.5^2}{2 \times 6.4 \times 7.1} = 0.78,$$

$$\text{Corr}_C = \frac{7.0^2 + 6.9^2 - 4.2^2}{2 \times 7.0 \times 6.9} = 0.82.$$

Where either the baseline or final standard deviation is unavailable, then it may be substituted by the other, providing it is reasonable to assume that the intervention does not alter the variability of the outcome measure. Correlation coefficients lie between –1 and 1. If a value less than 0.5 is obtained, then there is no value in using change from baseline and an analysis of final values will be more precise. Assuming the correlation coefficients from the two intervention groups are similar, a simple average will provide a reasonable measure of the similarity of baseline and final measurements across all individuals in the study (the average of 0.78 and 0.82 for the example is 0.80). If the correlation coefficients differ, then either the sample sizes are too small for reliable estimation, the intervention is affecting the variability in outcome measures, or the intervention effect depends on baseline level, and the use of imputation is best avoided. Before imputation is undertaken it is recommended that correlation coefficients are computed for many (if not all) studies in the meta-analysis and it is noted whether or not they are consistent. Imputation should be done only as a very tentative analysis if correlations are inconsistent.

(2) Imputing a change-from-baseline standard deviation using a correlation coefficient

Now consider a study for which the standard deviation of changes from baseline is missing. When baseline and final standard deviations are known, we can impute the missing standard deviation using an imputed value, Corr, for the correlation coefficient. The value Corr might be imputed from another study in the meta-analysis (using the method in (1) above), it might be imputed from elsewhere, or it might be hypothesized based on reasoned argument. In all of these situations, a sensitivity analysis should be undertaken, trying different values of Corr, to determine whether the overall result of the analysis is robust to the use of imputed correlation coefficients.

To impute a standard deviation of the change from baseline for the experimental intervention, use

$$SD_{E,change} = \sqrt{SD^2_{E,baseline} + SD^2_{E,final} - (2 \times \text{Corr} \times SD_{E,baseline} \times SD_{E,final})},$$

and similarly for the control intervention. Again, if either of the standard deviations (at baseline and final) are unavailable, then one may be substituted by the other if it is reasonable to assume that the intervention does not alter the variability of the outcome measure.

As an example, given the following data:

	Baseline	Final	Change
Experimental intervention (sample	mean=12.4 SD=4.2	mean=15.2 SD=3.8	mean=2.8

size 35)			
Control intervention (sample size 38)	mean=10.7 SD=4.0	mean=13.8 SD=4.4	mean=3.1

and using an imputed correlation coefficient of 0.80, we can impute the change-from-baseline standard deviation in the control group as:

$$SD_{C,change} = \sqrt{4.0^2 + 4.4^2 - (2 \times 0.80 \times 4.0 \times 4.4)} = 2.68 .$$