Statistical methods for study-centric data in RevMan Web

Gert van Valkenhoef

July 30, 2021

1 Key differences from custom-input analyses

The following transformations / calculations no longer need to be applied by authors, as they are done automatically when data are pulled into an analysis:

1. Combining arms: when multiple arms in the same study match an intervention of interest in the review, these arms are now combined automatically. Previously, authors would use a calculator feature in RevMan 5 to combine the arms before being able to enter the data in RevMan 5.

2. Splitting control arms: in analyses that sub-group by the type of intervention (e.g. low-dose vs high-dose), sometimes the same control arm would be used multiple times. Previously, authors would compensate for this by "splitting" the control arm, so that the overall pooled estimate wouldn't be over-confident. This is no longer necessary as the study will automatically be counted only once for the pooled estimate when using study-centric data.

3. Calculating other contrasts: in analyses where the comparison of interest is not reported directly in the study, authors would have to calculate the missing contrast from the available contrasts (e.g. when B-C is of interest but A-B and A-C were reported). This is done automatically for analyses based on study-centric data.

4. Calculating contrasts from arm-level data: when there is a mix of contrast-level and arm-level data, e.g. some studies reported odds ratios when others reported counts, authors would have to use the calculator to convert counts to odds ratios. This is done automatically for analyses based on study-centric data.

2 Including study data in an analysis

When an analysis is based on study-centric data, RevMan automatically includes study data based on the PICO eligibility settings for the analysis.

Studies are included when:

- They have compatible result data for the selected outcome and effect measure (see below), AND

- They have arms matching both the experimental and the control intervention, AND

- If a study filter was set: the study matches the study filter

Studies are excluded when:

- The study has missing data for any of the attributes used for the study filter or the subgroup factor

Result data is compatible when:

- If the analysis is based on contrasts, and the result reports the effect measure selected for analysis

- If the analysis is based on contrasts, and the result reports arm-level data that can be used to calculate the effect measure selected for the analysis

- If the analysis is based on arm-level data, and the result reports arm-level data

- The result has complete data for the relevant arms or contrasts, or sufficient data to calculate the relevant contrasts

If more than one result is compatible for a given study, either contrast-level data or arm-level data is preferred depending on the analysis settings (currently this situation can only occur for a contrast-level analysis).

If a subgroup factor was selected, studies will be placed in subgroups according to the value assigned to the subgroup factor for each study.

3 Data transformation

A number of data transformations are applied automatically by RevMan.

3.1 Combining arms - dichotomous data

If we have n_1/N_1 and n_2/N_2 to combine, the combined arm has n_{1+2}/N_{1+2}, where $n_{1+2} = n_1 + n_2$ and $N_{1+2} = N_1 + N_2$.

3.2 Combining arms - continuous data

If we have $\mu_1 \pm \mathrm{sd}_1(N_1)$ and $\mu_2 \pm \mathrm{sd}_2(N_2)$ to combine, the combined arm has:

$$N_{(1+2)} = N_1 + N_2$$
$$\mu_{(1+2)} = \frac{N_1\mu_1 + N_2\mu_2}{N_1 + N_2}$$
$$\mathrm{sd}^2_{(1+2)} = \frac{(N_1-1)\mathrm{sd}_1^2 + (N_2-1)\mathrm{sd}_2^2}{N_1 + N_2 - 1} + \frac{\frac{N_1 N_2}{N_1+N_2}(\mu_1^2 + \mu_2^2 - 2\mu_1\mu_2)}{N_1 + N_2 - 1}$$

3.3 Calculating contrasts from arm-level data

Calculating contrasts from arm-level data is covered in "Statistical algorithms in Review Manager 5", under "Formulae for individual studies". In this document we adopt the general assumption that arm-level data can be summarized as a mean and standard error for each arm: $\mu_i \pm \sigma_i$, and so for example the contrast of arm i relative to the reference arm 1, $\mu_{1,i} \pm \sigma_{1,i}$ is given by:

$$\mu_{1,i} = \mu_i - \mu_1$$
$$\sigma_{1,i}^2 = \sigma_1^2 + \sigma_i^2$$

This applies directly for the mean difference, with $\sigma_i = \text{sd}_i/\sqrt{N_i}$, but other effect measures can also be expressed this way, for example for the log odds ratio where the arm level expressions are for the log odds and its standard error.

3.4 Correlation of contrast-level data

When a study has three or more arms, there are two or more contrasts to any one given reference arm. These contrasts are correlated due to the fact that they were calculated to the same reference (`https://doi.org/10.1002/jrsm.1049`). This has two important consequences:

- Collecting the mean and standard error for each contrast is not sufficient when more than one of the contrasts is being considered simultaneously, as this would ignore their covariance

- The covariance between any two contrasts relative to the same reference arm are equal to the standard error of the reference arm

In this document, arm 1 is the reference arm by definition, unless stated otherwise. Let us denote the contrast of arm i relative to the reference arm 1 as $\mu_{1,i} \pm \sigma_{1,i}$. Then the contrasts $\mu_{1,i} \pm \sigma_{1,i}$ and $\mu_{1,j} \pm \sigma_{1,j}$ have a covariance σ_1^2.

The vector of contrasts $x = (\mu_{12}, \ldots \mu_{1n})$ therefore has an associated covariance matrix given by:

$$\Sigma = \begin{pmatrix} \sigma_{1,2}^2 & \sigma_1^2 & \cdots & \sigma_1^2 \\ \sigma_1^2 & \sigma_{1,3}^2 & \ddots & \vdots \\ \sigma_1^2 & \ddots & \ddots & \sigma_1^2 \\ \sigma_1^2 & \cdots & \sigma_1^2 & \sigma_{1,n}^2 \end{pmatrix}$$

i.e., the diagonal elements are the standard errors of the contrasts squared, and the off-diagonal elements are the (common) covariance term.

3.5 Combining arms - contrast-level data

To combine $\mu_{1,2} \pm \sigma_{1,2}$ and $\mu_{1,3} \pm \sigma_{1,3}$ into $\mu_{1,(2+3)} \pm \sigma_{1,(2+3)}$, an inverse-variance pooling of the arm-level effects is used:

$$\sigma_2^2 = \sigma_{1,2}^2 - \sigma_1^2$$
$$\sigma_3^2 = \sigma_{1,3}^2 - \sigma_1^2$$
$$\sigma_{(2+3)}^2 = \left(\frac{1}{\sigma_2^2} + \frac{1}{\sigma_3^2}\right)^{-1}$$
$$\sigma_{1,(2+3)}^2 = \sigma_{(2+3)}^2 + \sigma_1^2$$
$$\mu_{1,(2+3)} = \sigma_{(2+3)}^2 \left(\frac{1}{\sigma_2^2}\mu_{1,2} + \frac{1}{\sigma_3^2}\mu_{1,3}\right)$$

This generalises to a summation if more than two contrasts are to be combined:

$$\sigma_A^2 = \left(\sum_{i \in A} \frac{1}{\sigma_i^2}\right)^{-1}$$
$$\sigma_{1,A}^2 = \sigma_1^2 + \sigma_A^2$$
$$\mu_{1,A} = \sigma_A^2 \sum_{i \in A} \frac{1}{\sigma_i^2} \mu_{1,i}$$

where A is the set of arms to be combined.

3.6 Changing the reference arm - constrast-level data

Given a series of contrasts with arm 1 as the reference arm, we want to calculate a series of contrasts with arm 2 as the reference arm (the contrasts can always be arranged in this way). The given contrast are $x = (\mu_{1,2}, \ldots \mu_{1,n})$ with covariance matrix Σ. The contrasts to be calculated are $x' = (\mu_{2,1}, \mu_{2,3}, \ldots, \mu_{2,n})$, where

$$\mu_{2,1} = -\mu_{1,2}$$
$$\mu_{2,i} = \mu_{1,i} - \mu_{1,2} \ (i > 2)$$

To calculate the covariance matrix Σ' that corresponds to x', find the matrix B such that $x' = Bx$, then $\Sigma' = B\Sigma B^T$ (https://statproofbook.github.io/P/mvn-ltt.html). The matrix B is given by:

$$B = \left(\begin{array}{c|ccc} -1 & 0 & \cdots & 0 \\ \vdots & & D & \\ -1 & & & \end{array}\right)$$

Where D is the standard diagonal matrix.

Due to the specific structures of B and Σ, the expressions for the new variance terms and covariance simplify to:

$$\sigma_2^2 = \sigma_{1,2}^2 - \sigma_1^2$$
$$\sigma_{2,1}^2 = \sigma_{1,2}^2$$
$$\sigma_{2,i}^2 = \sigma_{1,i}^2 + \sigma_{1,2}^2 - 2\sigma_1^2$$

3.7 Calculating the contrast of interest

In any given analysis, multiple arms on a study result may map to the control and/or experimental intervention and/or the reference arm of the study may not match the control intervention. The operations of combining arms and changing the reference arm can be composed to calculate the contrast of interest, as shown in this diagram listing all the qualitatively different scenarios:
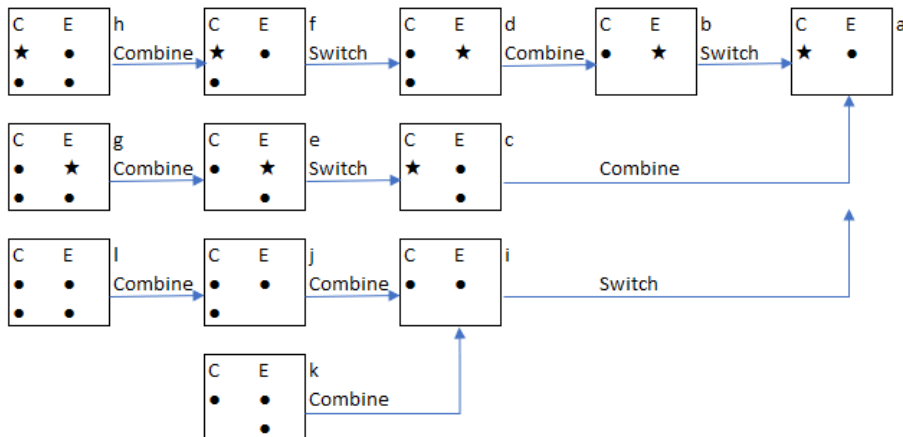
Scenarios are labeled a-l. Arms mapping to the control intervention are listed under "C", and those mapping to the experimental intervention under "E". A star denotes the reference arm, and a circle a non-reference arm. Arms that do not map to either the control or experimental intervention are not shown. "Combine" denotes the operation of combining arms, and "Switch" the operation of changing the reference arm.

3.8 Estimating the covariance of contrast-level data

Three alternative methods for estimating the covariance are available. If variances for three contrasts are known:

$$\sigma_1^2 = \frac{1}{2}(\sigma_{1,2}^2 + \sigma_{1,3}^2 - \sigma_{2,3}^2)$$

If the correlation $\rho_{1,2;1,3}$ is known:

$$\sigma_1^2 = \rho_{1,2;1,3}\sigma_{1,2}\sigma_{1,3}$$

If neither are known, the covariance can be estimated under the assumption that the arm-level variances are similar:

$$\sigma_1^2 \approx \frac{1}{2(n-1)}\sum_{i=2}^{n}\sigma_{1,i}^2$$

3.9 Subgrouping by type of intervention

When subgrouping by type of intervention, some contrasts may be separated between subgroups, but combined for the overall total. In such cases, the overall estimate will be calculated based on the method for "combining arms" above, rather than using the individual contrasts from within the subgroups.