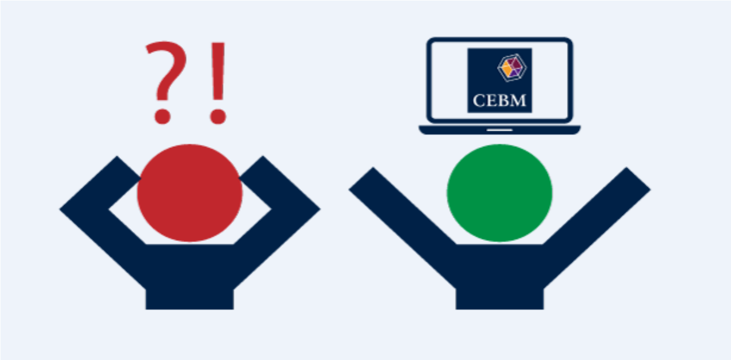
Tip for data extraction for meta-analysis - 22



What if the data I want are reported for the wrong time point?

Kathy Taylor

Previously, I highlighted a list of ways where, when extracting data for meta-analysis of continuous outcomes, you might find that a summary statistic that you want is missing. In this post I'll focus on the 1st way - ***the summary statistic you want is reported, but it's for the wrong time point.***

An example of this case is when you have the mean and standard deviations (SDs) of the baseline and final values of the outcome of interest, but you want those summary statistics of the change from baseline. The Cochrane Handbook refers to changes from baseline as change scores. Although it's possible to pool changes from baseline with final values in a meta-analysis, analysing change may be desirable because it is less skewed than final values, and it also removes the between-person variability from the analysis.

The following equations (which I shall call 'change score equations') give the mean and SD of the change from baseline:

$$mean_{Change} = mean_{Final} - mean_{Baseline}$$

$$SD_{Change} = \sqrt{\left(SD_{Baseline}^2 + SD_{Final}^2 - 2\rho_{B,F} SD_{Baseline} SD_{Final}\right)}$$

$\rho_{B,F}$ is the correlation coefficient for the correlation between the baseline and final values.

If this correlation coefficient is unknown, it may be estimated as 0.5. If there is a similar study that reports summary statistics for change from baseline, baseline and final values, a better estimate of the correlation coefficient is

$$\rho_{B,F} = \frac{SD_{Baseline}^2 + SD_{Final}^2 - SD_{Change}^2}{2\ SD_{Baseline}SD_{Final}}$$

If the correlation coefficients of the two intervention groups are similar, The Cochrane Handbook states that the average value may be used.

Summary data for final values may be derived from summary data for baseline and change with the following equations (which I shall call ‘endpoint equations’) :

$$mean_{Final} = mean_{Baseline} + \ mean_{Change}$$

$$SD_{Final} = \sqrt{\left(SD_{Baseline}^2 + SD_{Change}^2 + 2\rho_{B,C}SD_{Baseline}SD_{Change}\right)}$$

The change score SD equation cannot be simply rearranged due to data being correlated. So this SD involves a different correlation coefficient.

$\rho_{B,C}$ is the correlation coefficient for the correlation between the baseline and change from baseline values.

If this correlation coefficient is unknown, it may also be estimated as 0.5. As before, a more accurate correlation coefficient may be estimated from summary data from a similar study.

$$\rho_{B,C} = \frac{SD_{Final}^2 - SD_{Baseline}^2 - SD_{Change}^2}{2\ SD_{Baseline}SD_{Change}}$$

These estimates involve imputation (‘filling in’) of the correlation coefficients. Don’t forget to flag up studies that involve imputation to remove as part of your sensitivity analysis when you’ll want to check that your conclusions are robust to the estimates that you’ve made.

Here’s a tip…
There are equations that you can use to calculate summary data for change from baseline and the values at the end of the trial when other timepoint data are reported

In my next post, I'll focus on the 2nd way of how a summary statistic that you want may be missing: **the summary statistic you want is reported, but it's for the wrong group.**

Where did the equations come from?

The baseline, final and change from baseline are random variables.
The mean (average) value of a random variable (X) is the expected value of X. This is written as E(X). For a continuous random variable, it is calculated as

$E(X) = \int_{-\infty}^{\infty} xf(x)dx$
Also,
$E(X - Y) = \iint_{-\infty}^{\infty} (x - y)f(x, y)dxdy$
$E(X + Y) = \iint_{-\infty}^{\infty} (x + y)f(x, y)dxdy$
where
$f(x)$ is the probability density function
$f(x, y)$ is the joint probability density function
$A \Longrightarrow B$ means 'A implies B
Remember that AB means $A \times B$

Expanding the integral equations
$E(X - Y) = \iint_{-\infty}^{\infty} (x - y)f(x, y)dxdy$
$\Longrightarrow E(X - Y) = \iint_{-\infty}^{\infty} xf(x, y)dxdy - \iint_{-\infty}^{\infty} yf(x, y)dxdy$
$\Longrightarrow E(X - Y) = \int_{-\infty}^{\infty} xf(x)dx - \int_{-\infty}^{\infty} yf(y)dy$ since $\int_{-\infty}^{\infty} f(x)dx = 1$
$\Longrightarrow E(X - Y) = E(X) - E(Y)$ **equation 1**

$E(X + Y) = \iint_{-\infty}^{\infty} (x + y)f(x, y)dxdy$
$\Longrightarrow E(X + Y) = \iint_{-\infty}^{\infty} xf(x, y)dxdy + \iint_{-\infty}^{\infty} yf(x, y)dxdy$
$\Longrightarrow E(X + Y) = \int_{-\infty}^{\infty} xf(x)dx + \int_{-\infty}^{\infty} yf(y)dy$
$\Longrightarrow E(X + Y) = E(X) + E(Y)$ **equation 2**

$Change = Final - Baseline$
equation 1 $\Longrightarrow mean_{Change} = mean_{Final} - mean_{Baseline}$

$Final = Baseline + Change$
equation 2 $\Longrightarrow mean_{Final} = mean_{Baseline} + mean_{Change}$

The variance is the expected value of the squared deviation from the mean
$Var(X) = E((X - \mu_x)^2)$ where $\mu_x = E(X)$
$Var(Y) = E((Y - \mu_y)^2)$ where $\mu_y = E(Y)$
$\Longrightarrow Var(X - Y) = E(((X - Y) - E(X - Y))^2)$ **equation 3**

equation 1 $\Rightarrow E(X - Y) = \mu_x - \mu_y$ where $\mu_x = E(X)$ and $\mu_y = E(Y)$ **equation 4**
Substitute equation 4 into equation 3 and rearrange to variables
$Var(X - Y) = E\left(((X - \mu_x) - (Y - \mu_y))^2\right)$
$\Longrightarrow Var(X - Y) = E((X - \mu_x)^2) + (\left(Y - \mu_y\right)^2) - 2E((X - \mu_x)\left(Y - \mu_y\right))$
$\Longrightarrow Var(X - Y) = Var(X) + Var(Y) - 2Cov(X, Y)$ **equation 5**
where $Cov(X, Y)$ is the covariance of X and Y.

$Correlation\ coefficient = \rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{Cov(X,Y)}{\sqrt{SD_x^2 SD_y^2}}$
$\Longrightarrow Cov(X, Y) = \rho_{X,Y} SD_x SD_y$ **equation 6**
Substitute equation 6 into equation 5 and also using $Var(X) = SD_x^2$
$SD_{x-y}^2 = SD_x^2 + SD_y^2 - 2\rho_{X,Y} SD_x SD_y$ **equation 7**

equation 7 $\Longrightarrow SD_{Change}^2 = SD_{Baseline}^2 + SD_{Final}^2 - 2\rho_{B,F} SD_{Baseline} SD_{Final}$

$$\Longrightarrow \rho_{B,F} = \frac{SD_{Baseline}^2 + SD_{Final}^2 - SD_{Change}^2}{2\, SD_{Baseline} SD_{Final}}$$

$Var(X+Y) = E(((X+Y)-E(X+Y))^2)$ where μ=E(X+Y)   **equation 8**

equation 2 $\Rightarrow E(X+Y) = \mu_x + \mu_y$

Substitute into equation 8

$\Longrightarrow Var(X+Y) = E\big(((X-\mu_x)+E(Y-\mu_y))^2\big)$

$\Longrightarrow Var(X+Y) = E((X-\mu_x)^2 + E\big((Y-\mu_y)^2\big) + 2E((X-\mu_x)(Y-\mu_y))$

$\Longrightarrow Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$

$\Longrightarrow SD_{x+y}^2 = SD_x^2 + SD_y^2 + 2\rho_{X,Y}SD_xSD_y$ **equation 9**

equation 9 $\Longrightarrow SD_{Final}^2 = SD_{Baseline}^2 + SD_{Change}^2 + 2\rho_{B,C}SD_{Baseline}SD_{Change}$

$$\Longrightarrow \rho_{B,C} = \frac{SD_{Final}^2 - SD_{Baseline}^2 - SD_{Change}^2}{2\,SD_{Baseline}SD_{Change}}$$

Dr Kathy Taylor teaches data extraction in Meta-analysis. This is a short course that is also available as part of our MSc in Evidence-Based Health Care, MSc in EBHC Medical Statistics, and MSc in EBHC Systematic Reviews.

Follow updates on this blog, related news, and to find out about other examples of statistics being made more broadly accessible on Twitter @dataextips