Appendix 4. Statistical analyses methodology

One step multilevel model

In the primary analysis average RORs were estimated by using a one-step multi-level weighted regression model with random effects as described by Siersma et al (2007). This modelling framework allowed for random intervention effects (between-trial heterogeneity) within meta-analyses as well as random variation in the effect of mITT effect between meta-analyses.

Let us assume that we collected a sample of *M* meta-analyses comprising a total of *T* trials and compare. We are interested in comparing mITT versus noITT or ITT.

We setup mITT as the reference level in the model and compare it with noITT and ITT.

Each t^{th} trial contributes one observation, the observed effect size is expressed as log(OR), with weight equalling the inverse of its standard error SE[log(OR)]. The weighing is to account for the precision of the estimate in each trial: a high precision, i.e. low standard error of the log(OR), means that the trial has a small model error.

The random influence of the mITT vs ITT comparison in the m^{th} review is specified as $\xi_{mITT\ vs\ ITT,m} \sim i.i.d\ N\left(0, \sigma_{\xi_{mITT\ vs\ ITT}}\right)$ and mITT vs noITT comparison is specified as $\xi_{mITT\ vs\ noITT,m} \sim i.i.d\ N\left(0, \sigma_{\xi_{mITT\ vs\ noITT}}\right)$.

Correspondingly, the trial outcomes are modelled as follows:

$$\log(OR)_t = \beta_{mITT\ vs\ ITT} c_{mITT\ vs\ ITT} + \beta_{mITT\ vs\ noTT} c_{mITT\ vs\ noITT} + \sum_{m=1}^{M} \gamma_m I_m + \sum_{m=1}^{M} \xi_{mITT\ vs\ ITT,m} c_{mITT\ vs\ ITT} I_m + \sum_{m=1}^{M} \xi_{mITT\ vs\ noITT,m} c_{mITT\ vs\ noITT} I_m + \varepsilon_t.$$

The bias coefficient $\beta_{mITT\ vs\ ITT}$ measures the influence of the mITT vs ITT contrast on the estimated treatment effect where $\exp(\beta_{mITT\ vs\ ITT})$ is the average ROR. The bias due to the trial characteristic in a meta-analysis *m* is described by $\beta_{mITT\ vs\ ITT}+\xi_{mITT\ vs\ ITT}$). I_m, a dummy variable for the study being in meta-analysis $m\ and\ \gamma_m$ is the true treatment effect in meta-analysis *m*. Finally, $\varepsilon_t \sim i.i.d.N(0, \psi v_t)$, a residual term proportional to the apparent standard error of the log(OR)s.

As noted, the approach already accommodates between-trial heterogeneity and increases the power of the study because the smallest reviews in our data have two trials and many reviews, therefore, cannot contribute to an analysis using a literal version of an approach where we stratify on review (Sterne 2002).

If a trial was included in more than one meta-analysis we inflated standard errors to avoid double counting of patients (e.g.: if the control group of a trial with three arms was included in two

different meta-analyses, we inflated the standard error of the estimate for the control group by $\sqrt{2}$) (Nuesch 2009).

One step multilevel model - Adjustment on potential confounders

The approach can also easily incorporates other source of bias with respect to our main research interest (mITT vs ITT and mITT vs noITT), such as active treatment vs placebo, single centre vs multicenter, study size, post-randomization exclusion, sequence generation, allocation concealment, blinding and funding.

Adjusted analyses were performed incorporating one of these at time in the model allowing for bias heterogeneity across reviews. Bias estimates are plotted in one graph.

To control for publication bias, we adjusted the models for the variance of the log odds ratio for each randomised trial (Moreno 2009). We used this regression-based method for preventing loss of power, in fact this analysis permits to include all the trials we analysed in the unadjusted analysis.

Two steps model

In a secondary analysis average RORs were estimated by using a two-step approach where separate meta-regression is conducted within each review and the estimated RORs from each review are then pooled using the DerSimonian and Laird (1986) random-effects principle (Sterne et al. 2002). The two step approach was used for sensitivity analysis. We provide separates model for mITT vs ITT, mITT vs noITT and mITT vs noITT+ITT comparisons.

The heterogeneity across meta-analyses was quantified with the between–meta-analyses variance τ^2.

For mITT vs ITT comparison, at least one trial with mITT and one with ITT is required to perform such stratified analyses. Therefore, all individual meta-analyses did not contribute to the analysis. The same applies for mITT vs noITT and mITT vs noITT + ITT comparisons.

Investigating heterogeneity and influence with Baujat plots

We used Baujat (2002)approach for plotting, on the same graph, the heterogeneity of the bias effects between meta-analyses and the influence of each meta-analysis on the overall result. Each meta-analysis is represented by a dot on a 2D graph and the most heterogeneous and influential ones will appear in the upper right corner of the graph.

Two plots were presented, one for the mITT vs ITT comparison and one for the mITT vs noITT comparison.

The X-axis represents the contribution of the meta-analysis to the bias heterogeneity across reviews, that is, the squared difference between the bias estimates of the meta-analysis under consideration

and the mean bias estimate, weighted by the inverse of the estimated variance of the bias in the meta-analysis:

$$x_i = \frac{\left(\hat{\beta}_i - \hat{\beta}\right)^2}{\hat{\sigma}_i^2}$$

The Y-axis represents the influence of the trial on the overall bias estimate, defined as the squared difference between the mean bias estimate removing the meta-analysis under consideration and the mean bias estimate, weighted by the inverse of the estimated variance of the bias removing the meta-analysis:

$$y_i = \frac{\left(\hat{\beta}_{-i} - \hat{\beta}\right)^2}{\hat{\sigma}_{-i}^2}$$

The estimated quantities derived from one-step multilevel model both for mITT vs ITT and mITT vs noITT comparisons.

The approach is fully exploratory, thus cut-off is not provided for excluding meta-analyses in subsequent sensitivity analysis. In the sensitivity analysis we excluded the most heterogeneous and influential meta-analyses until no significant heterogeneity was present.