Appendix 2: Details of the simulation analysis to correct incomplete reporting in the whole sample of trials and re-estimate the avoidable waste

We performed a simulation analysis to correct incomplete reporting in the whole sample of trials and re-assess the proportion of avoidable waste related to inadequate methods. The simulation was performed in three successive steps.

Step 1: We started by extrapolating our reassessment of risk of bias to the remaining 1 086 trials (1 286 minus the 200 trials already reassessed).We considered the domains of the non-reassessed trials as missing data and used multiple imputation by chained equations to simulate the risk of bias that would have been obtained if the trials had been reassessed. Only one imputation was performed at this step, with 10 iterations of the Gibbs sampler. The imputation model used the risk of bias of the same domain as evaluated by the review authors, the trials medical fields and the type of outcome (objective or subjective). This implies making use of the observed agreement by domain between the review authors' initial assessment and our reassessment for the 200 trials for which we had both evaluations. This simulation relied on the assumption that data were "missing at random" (MAR) (and not completely at random), that is, that the missingness was independent of what the reassessed risk of bias of a domain would have been, given the risk of bias assessed by the review authors, the medical field and the type of outcome. In our case, this was verified by design for all trials with at least 1 domain at high risk of bias, since the 200 reassessed trials were selected at random among them. For the other trials, our model assumed that the reassessment of a domain at low or unclear risk of bias would be independent of the fact that the review authors had initially rated another domain of the same trial at high risk.

Step 2: We then considered each domain rated as unclear as missing data and used multiple imputations to attribute a risk-of-bias assessment of high or low to these domains. Data were

considered MAR and were imputed by using chained equations. Variables used for multiple imputations were the risk of bias of all domains, the year of publication of the trial (by 5 years categories), the medical field and the type of outcome. Again, one imputed dataset was obtained, with 10 iterations of the Gibbs sampler. At the end of this step, we had a set of trials with no domain rated at unclear risk of bias, for only high- or low-risk trials. In our case, the MAR assumption stated that the rating of a domain of a given trial as "unclear" could be related to other characteristics of the trial and in particular to the risk of bias of other domains but did not depend on the true risk of bias (low or high) of this domain given the other variables.

Step 3: We imputed the risk of bias for each domain obtained after applying methodological adjustments for the domains that were missing because the initial rating was unclear or because the trial was not part of the sample of trials that were reassessed and corrected. Both situations were imputed simultaneously using chained equations. Again, one imputed dataset was obtained, with 10 iterations of the Gibbs sampler. Variables used for imputation were the risk of bias of each domain before and after methodological adjustments, the medical field, the type of outcome and the year of publication. As for the preceding steps, we assumed the missing data mechanism would be MAR and that the imputation model would be correctly specified. In our case this meant that the probability to implement methodological adjustments leading to a low risk of bias specific domain did not depend on our ability to implement such adjustments; given the risk of bias of this and other domains and our ability to implement methodological adjustments to these other domains.

Steps 2 and 3 took advantage of the potential multiple correlations among domains (eg, random sequence generation and allocation concealment could always be corrected easily and at no cost, and modifying the risk of bias for both blinding domains appeared to be correlated).

The three steps were repeated 999 times, thus yielding 999 imputed datasets with no domain rated as unclear risk (all domains being at high or low risk of bias). Among these datasets, domains at high risk of bias were potentially corrected to low risk because of easy adjustments at no or minor extra cost.

The proportion of trials with at least 1 domain at high risk of bias that could change to low risk was computed for each dataset, and a pooled estimate was obtained by Rubin's rule[1]. Multiple imputations involved the "mice" package[2] for R software[3].

References

1. Rubin DB. Multiple Imputation for Nonresponse in Surveys. John Wiley & Sons, New York, 1987.
2. Van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. *J Stat Soft.* 2011;45(3):1-67.
3. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2013. URL http://www.R-project.org/.