Item 18. Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing prespecified from exploratory

Example—“On the basis of a study that suggested periop‑ erative β-blocker efficacy might vary across baseline risk, we prespecified our primary subgroup analysis on the basis of the revised cardiac risk index scoring system. We also did prespecified secondary subgroup analyses based on sex, type of surgery, and use of an epidural or spinal anaesthetic. For all subgroup analyses, we used Cox pro‑ portional hazard models that incorporated tests for inter‑ actions, designated to be significant at p <0.05 Figure 3 shows the results of our prespecified subgroup analyses and indicates consistency of effects … Our subgroup analyses were underpowered to detect the modest differences in subgroup effects that one might expect to detect if there was a true subgroup effect.”

Explanation

Multiple analyses of the same data create a risk for false positive findings. Authors should resist the temptation to perform many subgroup analyses. Analyses that were prespecified in the trial protocol (see item 24) are much more reliable than those suggested by the data, and therefore authors should report which analyses were prespecified. If subgroup analyses were undertaken, authors should report which subgroups were examined, why, if they were prespecified, and how many were prespecified. Selective reporting of subgroup analyses could lead to bias. When evaluating a subgroup the question is not whether the subgroup shows a statistically significant result but whether the subgroup treatment effects are significantly different from each other. To determine this, a test of interaction is helpful, although the power for such tests is typically low. If formal evaluations of interaction are undertaken (see item 12b) they should be reported as the estimated difference in the intervention effect in each subgroup (with a confidence interval), not just as P values. In one survey, 35 of 50 trial reports included subgroup analyses, of which only 42% used tests of interaction. It was often difficult to determine whether subgroup analyses had been specified in the protocol. In another survey of surgical trials published in high impact journals, 27 of 72 trials reported 54 subgroup analyses, of which 91% were post hoc and only 6% of subgroup analyses used a test of interaction to assess whether a subgroup effect existed. Similar recommendations apply to analyses in which adjustment was made for baseline variables. If done, both unadjusted and adjusted analyses should be reported. Authors should indicate whether adjusted analyses, including the choice of variables to adjust for, were planned. Ideally, the trial protocol should state whether adjustment is made for nominated baseline variables by using analysis of covariance. Adjustment for variables because they differ significantly at baseline is likely to bias the estimated treatment effect. A survey found that unacknowledged discrepancies between protocols and publications were found for all 25 trials reporting subgroup analyses and for 23 of 28 trials reporting adjusted analyses.