Item 7b. When applicable, explanation of any interim analyses and stopping guidelines

Examples—“Two interim analyses were performed during the trial. The levels of significance maintained an overall P value of 0.05 and were calculated according to the O’BrienFleming stopping boundaries. This final analysis used a Z score of 1.985 with an associated P value of 0.0471.”126 “An independent data and safety monitoring board periodically reviewed the efficacy and safety data. Stopping rules were based on modified Haybittle Peto boundaries of 4 SD in the first half of the study and 3 SD in the second half for efficacy data, and 3 SD in the first half of the study and 2 SD in the second half for safety data. Two formal interim analyses of efficacy were performed when 50% and 75% of the expected number of primary events had accrued; no correction of the reported P value for these interim tests was performed.”

Explanation—

Many trials recruit participants over a long period. If an intervention is working particularly well or badly, the study may need to be ended early for ethical reasons. This concern can be addressed by examining results as the data accumulate, preferably by an independent data monitoring committee. However, performing multiple statistical examinations of accumulating data without appropriate correction can lead to erroneous results and interpretations. If the accumulating data from a trial are examined at five interim analyses that use a P value of 0.05, the overall false positive rate is nearer to 19% than to the nominal 5%. Several group sequential statistical methods are avail‑ able to adjust for multiple analyses, and their use should be pre-specified in the trial protocol. With these methods, data are compared at each interim analysis, and a P value less than the critical value specified by the group sequential method indicates statistical significance. Some trialists use group sequential methods as an aid to decision making, whereas others treat them as a formal stop‑ ping rule (with the intention that the trial will cease if the observed P value is smaller than the critical value). Authors should report whether they or a data monitoring committee took multiple “looks” at the data and, if so, how many there were, what triggered them, the statistical methods used (including any formal stopping rule), and whether they were planned before the start of the trial, before the data monitoring committee saw any interim data by allocation, or some time thereafter. This information is often not included in published trial reports, even in trials that report stopping earlier than planned.