Item 12a. Statistical methods used to compare groups for primary and secondary outcomes

Example—“The primary endpoint was change in body‑ weight during the 20 weeks of the study in the intention-to treat population … Secondary efficacy endpoints included change in waist circumference, systolic and diastolic blood pressure, prevalence of metabolic syndrome … We used an analysis of covariance (ANCOVA) for the primary endpoint and for secondary endpoints waist circumference, blood pressure, and patient-reported outcome scores; this was supplemented by a repeated measures analysis. The ANCOVA model included treatment, country, and sex as fixed effects, and bodyweight at randomisation as covariate. We aimed to assess whether data provided evidence of superiority of each liraglutide dose to placebo (primary objective) and to orlistat (secondary objective).”

Explanation—Data can be analysed in many ways, some of which may not be strictly appropriate in a particular situation. It is essential to specify which statistical procedure was used for each analysis, and further clarification may be necessary in the results section of the report. The principle to follow is to, “Describe statistical methods with enough detail to enable a knowledgeable reader with access to the original data to verify the reported results” (www.icmje.org). It is also important to describe details of the statistical analysis such as intention-to-treat analysis (see box 6). Almost all methods of analysis yield an estimate of the treatment effect, which is a contrast between the outcomes in the comparison groups. Authors should accompany this by a confidence interval for the estimated effect, which indicates a central range of uncertainty for the true treatment effect. The confidence interval may be interpreted as the range of values for the treatment effect that is compatible with the observed data. It is customary to present a 95% confidence interval, which gives the range expected to include the true value in 95 of 100 similar studies. Study findings can also be assessed in terms of their statistical significance. The P value represents the probability that the observed data (or a more extreme result) could have arisen by chance when the interventions did not truly differ. Actual P values (for example, P=0.003) are strongly prefer‑ able to imprecise threshold reports such as P<0.05.Standard methods of analysis assume that the data are “independent.” For controlled trials, this usually means that there is one observation per participant. Treating multiple observations from one participant as independent data is a serious error; such data are produced when outcomes can be measured on different parts of the body, as in dentistry or rheumatology. Data analysis should be based on counting each participant onceor should be done by using more complex statistical procedures. Incorrect analysis of multiple observations per individual was seen in 123 (63%) of 196 trials in rheumatoid arthritis.