The NEW ENGLAND JOURNAL of MEDICINE

Statistical Reporting Guidelines

Our Statistical Consultants recommend the following best statistical practices in manuscripts submitted to the *Journal*. We recommend that you follow them in the design and reporting of research studies.

For all studies:

- The Methods section of all manuscripts should contain a brief description of sample size and power considerations for the study, as well as a brief description of the methods for primary and secondary analyses.
- The Methods section of all manuscripts should include a description of how missing data have been handled. Unless missingness is rare, a complete case analysis is generally not acceptable as the primary analysis and should be replaced by methods that are appropriate, given the missingness mechanism. Multiple imputation or inverse probability case weights can be used when data are missing at random; model-based methods may be more appropriate when missingness may be informative. For the *Journal's* general approach to the handling of missing data in clinical trials please see Ware et al (N Engl J Med 2012;367:1353–1354).
- Significance tests should be accompanied by confidence intervals for estimated effect sizes, measures of association, or other parameters of interest. The confidence intervals should be adjusted to match any adjustment made to significance levels in the corresponding test.
- Unless one-sided tests are required by study design, such as in noninferiority clinical trials, all reported P values should be two-sided. In general, P values larger than 0.01 should be reported to two decimal places, and those between 0.01 and 0.001 to three decimal places; P values smaller than 0.001 should be reported as P<0.001. Notable exceptions to this policy include P values arising from tests associated with stopping rules in clinical trials or from genome-wide association studies.
- Results should be presented with no more precision than is of scientific value and is meaningful given the available sample size. For example, measures of association, such as odds ratios, should ordinarily be reported to two significant digits. Results derived from models should be limited to the appropriate number of significant digits.

For clinical trials:

- Original and final protocols and statistical analysis plans (SAPs) should be submitted along with the manuscript, as well as a table of amendments made to the protocol and SAP indicating the date of the change and its content.
- The analyses of the primary outcome in manuscripts reporting results of clinical trials should match the analyses prespecified in the original protocol, except in unusual circumstances. Analyses that do not conform to the protocol should be justified in the Methods section of the manuscript. The editors may ask for additional analyses that are not specified in the protocol.
- When comparing outcomes in two or more groups in confirmatory analyses, investigators should use the testing procedures specified in the protocol and SAP to control overall type I error — for example, Bonferroni adjustments or prespecified hierarchical procedures. P values adjusted for multiplicity should be reported when appropriate and labeled as such in the manuscript. In hierarchical testing procedures, P values should be reported only until the last comparison for which the P value was statistically significant. P values for the first nonsignificant comparison and for all comparisons thereafter should not be reported. For prespecified exploratory analyses, investigators should use methods for controlling false discovery rate described in the SAP — for example, Benjamini–Hochberg procedures.
- When no method to adjust for multiplicity of inferences or controlling false discovery rate was specified in the protocol or SAP of a clinical trial, the report of all secondary and exploratory endpoints should be limited to point estimates of treatment effects with 95% confidence intervals. In such cases, the Methods section should note that the widths of the intervals have not been adjusted for multiplicity and that the inferences drawn may not be reproducible. No P values should be reported for these analyses.
- Please see Wang et al (N Engl J Med 2007;357:2189–2194) on recommended methods for analyzing subgroups. When the SAP prespecifies an analysis of certain subgroups, that analysis should conform to the method described in the SAP. If the study team believes a post hoc analysis of subgroups is important, the rationale for conducting that analysis should be stated. Post hoc analyses should be clearly labeled as post hoc in the manuscript.
- Forest plots are often used to present results from an analysis of the consistency of a treatment effect across subgroups of factors of interest. Such plots can be a useful display of estimated treatment effects across subgroups, and the editors recommend that they be included for important subgroups. If subgroups are small, however, formal inferences about the homogeneity of treatment effects may not be feasible. A list of P values for treatment by subgroup interactions is subject to the problems of multiplicity and has limited value for inference. Therefore, in most cases, no P values for interaction should be provided in the forest plots.
- If significance tests of safety outcomes (when not primary outcomes) are reported along with the treatment-specific estimates, no adjustment for multiplicity is necessary. Because information contained in the safety endpoints may signal problems within specific organ classes, the editors believe that the type I error rates larger than 0.05 are

acceptable. Editors may request that P values be reported for comparisons of the frequency of adverse events among treatment groups, regardless of whether such comparisons were prespecified in the SAP.

- When possible, the editors prefer that absolute event counts or rates be reported before relative risks or hazard ratios. The goal is to provide the reader with both the actual event frequency and the relative frequency. Odds ratios should be avoided, as they may overestimate the relative risks in many settings and be misinterpreted.
- Authors should provide a flow diagram in CONSORT format. The editors also encourage authors to submit all the relevant information included in the CONSORT checklist. Although all of this information may not be published with the manuscript, it should be provided in either the manuscript or a supplementary appendix at the time of submission. The CONSORT statement, checklist, and flow diagram are available on the CONSORT website.

For observational studies:

The validity of findings from observational studies depends on several important assumptions, including those relating to sample selection, measured and unmeasured confounding, and the adequacy of methods used to control for confounding. The Methods section of observational studies should describe how these and other relevant issues were managed in the design and analysis.

- If an observational study included a prespecified SAP with a description of hypotheses to be tested, a signed and dated version of that plan should be included with the manuscript submission. The *Journal* encourages authors to deposit SAPs for observational studies in one of the online repositories designed for this purpose.
- When appropriate, observational studies should use prespecified accepted methods for controlling family-wise error rate or false discovery rate when multiple tests are conducted. In manuscripts reporting observational studies without a prespecified method for error control, summary statistics should be limited to point estimates and 95% confidence intervals. In such cases, the Methods section should note that the widths of the intervals have not been adjusted for multiplicity and that the inferences drawn from the inferences may not be reproducible. No P values should be reported for these analyses.
- If no prespecified analysis plan exists, the Methods section should provide an outline for the planned method of analysis, including
  - Eligibility criteria for the selection of cases and method of sampling from the data, with a diagram as appropriate.
  - A description of the association or causal effect to be estimated and the rationale for this choice.
  - The prespecified method of analysis to draw inference about treatment or exposure effect or association.
- Studies reporting the effect of a treatment or exposure should show the distribution of potential confounders and other variables, stratified by exposure or intervention group.

When the analysis depends on the confounders being balanced by exposure group, differences between groups should be summarized with point estimates and 95% confidence intervals when appropriate.

- Complex models and their diagnostics can often be best described in a supplementary appendix. Authors are encouraged to conduct an analysis that quantifies potential sensitivity to bias from unmeasured confounding; absent that, authors must provide a discussion of potential biases induced by unmeasured confounders.
- Authors are encouraged to retest findings in a similar but independent study or studies to assess the robustness of their findings.