# Annals of Internal Medicine Statistical Guidelines

## Presentation

<table>
<thead>
<tr>
<th>Issue</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentages</td>
<td>Report percentages to one decimal place (i.e., xx.x%) when sample size is $\geq$200.</td>
</tr>
<tr>
<td>To avoid the appearance of a level of precision that is not present with small samples, do not use decimal places (i.e., xx%, not xx.xx%) when sample size is $&lt;200$.</td>
<td></td>
</tr>
<tr>
<td>Standard deviations</td>
<td>Use “mean (SD)” rather than “mean ± SD” notation. The ± symbol is ambiguous and can represent standard deviation or standard error.</td>
</tr>
<tr>
<td>Standard errors</td>
<td>Report confidence intervals, rather than standard errors, when possible.</td>
</tr>
<tr>
<td>P values</td>
<td>Report exact p-values to two decimal places except when p$&lt;0.001$, in which case “p$&lt;0.001$” is sufficient.</td>
</tr>
<tr>
<td>Use the word trend when describing a test for trend or dose-response.</td>
<td></td>
</tr>
<tr>
<td>“Trend”</td>
<td>Avoid the term “trend” when referring to p-values near but not below 0.05. In such instances, simply report a difference and the confidence interval of the difference (if appropriate) with or without the p-value.</td>
</tr>
<tr>
<td>Statistical software</td>
<td>Specify in the statistical analysis section the statistical software—version, manufacturer, and manufacturer’s location—used for analyses.</td>
</tr>
</tbody>
</table>

When reporting the findings from Cox proportional hazards models:

- Do not describe hazard ratios as relative risks.
- Do report how the assumption of proportional hazards was tested, and what the test showed.

In tables that simply describe characteristics of two or more groups (e.g., Table 1 of a clinical trial):

- Report averages with standard deviations, not standard errors, when data are normally distributed.
- Report median (minimum, maximum) or median (25th, 75th percentile [interquartile range, or IQR]) when data are not normally distributed.
- Avoid reporting p values as there can be imbalance when p’s are not significant (because of small sample size) and balance.
Tables Reporting Multivariable Analyses

Authors sometimes present tables that compare one by one an outcome with multiple individual factors followed by a multivariable analysis that adjusts for confounding. If confounding is present, as is often the case, the one-way comparisons are simply intermediate steps that offer little useful information for the reader. In general, omit presenting these intermediate steps in the manuscript and do not focus on them in the Results or Discussion.

The following references give useful information about the design and format of informative tables and figures:


Also, follow a few simple rules of thumb:

1. Avoid pie charts.
2. Avoid simple bar plots or histograms that do not present measures of variability.
3. Provide raw data (numerators and denominators) in the margins of meta-analysis forest plots.
4. Depict numbers of people at risk at different times in survival plots. (see Pocock et al. above).

Multivariable Analysis TOP

Screening covariates

Approaches that select factors for inclusion in a multivariable model only if the factors are “statistically significant” in “bivariate screening” are not optimal. A factor can be a confounder even it is not statistically significant by itself because it changes the effect of
the exposure of interest when it is included in the model, or because it is a confounder only when included with other covariates.

Reference


Model building

Authors should avoid stepwise methods of model building, except for the narrow application of hypothesis generation for subsequent studies. Stepwise methods include forward, backward, or combined procedures for the inclusion and exclusion of variables in a statistical model based on predetermined p value criteria. Better strategies than p-value driven approaches for selecting variables are those that use external clinical judgment. Authors might use a bootstrap procedure to determine which variables, under repeated sampling, would end up in the model using stepwise variable selection procedures. Regardless, authors should tell readers how model fit was assessed, how and which interactions were explored, and the results of those assessments.

References


Measurement Error

If several risk factors for disease are considered in a logistic regression model and some of these risk factors are measured with error, the point and interval estimates of relative risk corresponding to any of these factors may be biased either toward or away from the null value; the direction of bias is never certain. In addition to potentially biased estimates, confidence intervals of correctly adjusted estimates will be wider, sometime substantially, than naïve confidence intervals. Authors are encouraged to consult the references below for strategies to address this problem.

References


Measures of Effect and Risk **TOP**

Clinically meaningful estimates

Authors should report results for meaningful metrics rather than reporting raw results. For example, rather than reporting the log odds ratio from a logistic regression, authors should transform coefficients into the appropriate measure of effect size, odds ratio, relative risk, or risk difference. Don’t give readers an estimate, such as an odds ratio or relative risk, for a one unit change in the factor of interest when a one unit change lacks clinical meaning (age, mmHg of blood pressure, or any other continuous or interval measurement with small units) All estimates should reflect a clinically meaningful change, along with 95% confidence bounds.

Between group differences

For comparisons of interventions (e.g., trials), focus on between-group differences, with 95% confidence intervals of the differences, and not on within-group differences. State the results using absolute numbers (numerator/denominator) when feasible. When discussing effects, refer to the confidence intervals rather than p values and point out for readers if the confidence intervals exclude the possibility of significant clinical benefit or harm.

Odds ratios and predicted probabilities

Authors often report odds ratios for multivariable results when the odds ratio is difficult to interpret or not meaningful. First, the odds ratio might overstate the effect size when the reference risk is high. For example, if the reference risk is 25% (odds = 0.33) and the odds ratio is 3.0, the relative risk is only 2.0. Statements such as “threefold increased
“risk” or “three times the risk” are incorrect. Second, readers want an easily understood measure of the level of risk (and the confidence intervals) for different groups of patients as defined by treatment, exposure, and covariates. Consider providing them a table of predicted probabilities for each of the factors of interest, and confidence intervals of those predicted probabilities. Moreover, a multiway table that cross classifies predicted probabilities by the most important factor and then adjusts for the remaining factors will often be more meaningful than a table of adjusted odds ratios. Standard commercial software can produce predicted probabilities and confidence bounds.

Reference

Altman DG, Deeks JJ, Sackett DL. Odds ratios should be avoided when events are common. BMJ. 1998;317:1318. PMID: 9804732

Missing Data

Missing variables

Always report the frequency of missing variables and how the analysis handled missing data. Consider adding a column to tables or a row under figures that makes clear the amount of missing data. Avoid using a simple indicator or dummy variable to represent a missing value. Replacing missing predictors with dummy variables or missing indicators generally leads to biased estimates.

References

Vach W, Blettner M. Biased estimation of the odds ratio in case-control studies due to the use of ad hoc methods or correcting for missing values of confounding variables. Am J Epidemiol. 1991;134:895-907. PMID: 1670320


Missing Outcomes

Always report the frequency of missing outcomes and follow-up data; reasons and any patterns for the missing data; and how you handled missing data in the analyses. Do not use a last observation carried forward approach (LOCF) to address incomplete follow-up even if the original protocol pre-specified that approach for handling missing data. LOCF approaches understate variability and result in bias. The direction of the bias is not
predictable. Although the method of addressing missing data may have little import on findings when the proportion of missing data is small (e.g. <5%), authors should avoid using out-dated or biased methods to address incomplete follow-up. Appropriate methods for handling missing data include imputation, pattern-mixture (mixed) models, and selection models. Application of these methods requires consideration of the patterns and potential mechanisms behind the missing data.

References


Longitudinal Analyses TOP

Consider using longitudinal analyses if outcome data were collected at more than one time point. With an appropriate model for longitudinal analysis, you can report differences within groups over time, differences between groups, and differences across groups of their within-group changes over time (usually the key contrast of interest). You can control for any confounding that might emerge, such as a difference in a variable (e.g., body weight) among those who remained in the study until completion. Longitudinal analysis options include a population averaged analysis (GEE, for example) that estimates the time by treatment interaction and adjusts variance for the repeated measures within individuals over time. Another option is a mixed effects model, with random effects for patient, and the estimate of interest being the time by treatment interaction. In choosing a model, consider whether any missing data are missing at random (i.e. “ignorable” missing data) or missing dependent on the observed data (i.e. informative missing data). In GEE analyses, missing data are assumed to be missing completely at random independent of both observed and unobserved data. In random coefficient analysis, missing data are assumed missing at random dependent on observed data but not on unobserved data.

Reference