Useful Resources

**Uniform Requirements for Manuscripts Submitted to Medical Journals**
International Committee of Medical Journal Editors
http://www.icjme.org

**Basic Guide for Writing Articles and Communicating with Editors**
Morgan P. An Insider's Guide for Medical Authors and Editors. ISI Press 1986

**Suggestions for Reporting Statistics in Journals Articles**


**Reporting Recommendations for Specific Types of Studies/Manuscripts**

- **Randomized Controlled Trials:**

- **Cost-Effectiveness Analyses:**

- **Systematic Reviews:**

- **Meta-Analyses:**


- **Studies of Diagnostic Tests:**
  http://www.consort-statement.org/stardstatement.htm
Non-randomized Behavioral Interventions
Common Statistical Pitfalls and their Prevention

These are statistical pitfalls that we encounter often as editors. The first set includes issues of presentation that are easy to avoid and simple to correct. The second set includes issues that are more complex and often require statistical assistance for a proper solution.

I. Issues of Presentation: Easy to Avoid and Simple to Correct

1. The Descriptive Table (“Table 1”)

Avoid making comparisons using t-tests and chisq. These initial tables are for description only. There can be imbalance in groups while p>0.05 because of small sample sizes and good balance when p<0.05 because of large sample sizes.

Use median, minimum and maximum (or 25th and 75th percentile) in all instances of data that are not normally distributed. Mean is often not a good measure of center. Standard deviation is not a good measure of spread. Standard errors are not relevant for descriptive tables.

2. One-way Comparisons of Independent Variables with Outcome (“Table 2”)

This table often compares the outcome with each factor one by one. The next step in the analysis is often multivariate analyses to adjust for confounding. If confounding is present, as is often the case, then these one-way comparisons are simply intermediate steps in the analysis and they offer little useful information for the reader. You do not have to present them and should not focus on them in the Results or Discussion.

3. Use of Ambiguous ± Notation

Often in the abstract appears the ± symbol to represent either standard deviation or standard error. Authors often neglect to explain what it means.

4. Beta-coefficients Instead of Estimates of Effect Size

Avoid giving the reader raw results. Transform results into a meaningful metric. Neither text nor tables should report the log odds ratio from a logistic regression. All coefficients should be transformed in the appropriate measure of effect size, odds ratio, relative risk, and risk difference.

5. Standard errors instead of confidence intervals
Compute confidence intervals for the reader. Standard errors are intermediate results that the author should interpret for the reader. Authors should not assume that readers will have calculators handy with which to compute the 95% confidence bounds.

6. Effect size for a meaningless increment in the factor of interest

Readers should never be given an estimate, such as an odds ratio or relative risk, for a one unit change in the factor of interest, such as age, mmHg of blood pressure, or any other continuous or interval measurement with small units, where the one unit change has no clinical meaning. The result is odds ratios of relative risks that are very close to 1.0, although they might be important. All estimates should reflect a clinically meaningful change, such as 10 years of age, along with their 95% confidence bounds.

7. Trend

Authors often use the term “trend” often to refer to a p-value that is small but not below 0.05. “Trend” should be reserved to a test for trend or a dose response. It is better simply to report a difference and the confidence interval of the difference and then let the data speak for themselves. The traditional use of p=0.05 as a demarcation of statistical significance is arbitrary.

II. More Complex Issues: Model Building for Multivariable Regression

1. Screening Covariates Based on Bivariate Relationships

Approaches that select factors for inclusion in a model only if they are “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.

2. Stepwise (p-value driven) Model Building

Stepwise methods for variable selection are those that include forward, backward, or combined procedures for the inclusion and exclusion of factors in a statistical model based on predetermined criteria of statistical significance. Authors should avoid stepwise methods except for the narrow application of hypothesis generation for subsequent studies. The reasons for avoiding stepwise methods are now well documented in the statistical and epidemiologic literature. (Collett D, Stepniewska K. Some practical issues in binary data analysis. Statist Med. 1999;18:2209-21.).

Annals routinely asks authors to remove their stepwise regressions and reanalyze the data for several reasons:
- Stepwise methods might fail to select important confounders that should be in the model because they change the coefficient of the covariate of interest (Mickey RM, Greenland S. The impact of confounder selection criteria on effect estimation. Am J Epidemiol. 1989;129:125-37.).
- Stepwise methods involve multiple comparisons and thus the resulting model will have p-values that are meaningless because of risk of including factors that are not actually significant.

- The stepwise selection results in bias in the estimated regression coefficients (Steyerberg EW, Eijkemans MJC, Habbema DF. Stepwise selection in small data sets: a simulation study of bias in logistic regression analysis. J Clin Epidemiol. 1999;52:935-42). The direction of bias is usually away from the null — in other words the estimated effect sizes are too large. The performance of stepwise methods worsens as the sample size decreases relative to the number of candidate covariates in the model. By candidate covariable, we mean the number of factors that are considered for inclusion in a statistical model. In the case of logistic regression, there should be at least 50 events per candidate covariates before one might consider stepwise methods. For example, if one has a 10% event rate, and there are 10 candidate covariates, then there needs to be a sample size of 5000 to support the use of stepwise methods. (50/0.1)*10). Below this number, stepwise methods do not perform as well as other strategies for model selection (Steyerberg EW, Eijkemans MJC, Harrell FE, Jr., Habbema JDF. Prognostic modeling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. Statist Med. 2000;19:1059-1079.).


3. Missing Data

Missing Xs

(a) Reporting frequency of missing data
The reader should always be informed of the frequency of missing factors (Xs) and how the authors handled missing data in analysis. For table 1, the descriptive table, the authors might use a column to report the percentage of missing data for each covariate. This might be the column that was previously used for reporting p-values, which are unnecessary for the descriptive table. This information would be especially important if you decided to impute missing data, or if you chose to use a complete case analysis (using only those observations with complete data on all covariates.)
(b) Handling missing data in the analysis

A common, yet thoroughly unsupported method of handling missing data is the use of an indicator to represent a missing value. The shortcomings of this method of handling missing data are sufficiently serious that it should be avoided.


**Missing Ys: Avoid “Last Observation Carried Forward”**


**4. Report Effect Sizes that are Clinically Useful**

Authors often report odds ratios for multivariable results when the odds ratio is difficult to interpret or not meaningful to the reader. 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.

In addition, the typical reader has no need for an odds ratio as a report of the difference in risks across exposure or treatment groups. Rather, a table of predicted probabilities for each of the factors of interest, and confidence intervals of those predicted probabilities. The reader wants to know the level of risk (and the confidence intervals) for different groups of patients as defined by treatment, exposure, and covariates. A multiway table that cross classified 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.

5. Measurement Error

Measurement error, especially with regard to covariates (as contrasted with exposures of interest) is an especially difficult problem. Contrary to simple rules of thumb, in the multivariable setting, the direction of bias is never certain. “If several risk factors for disease are considered in the same multiple 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.” (Rosner B, et al. Correction of logistic regression relative risk estimates and confidence intervals for measurement error. ... Am J Epidemiol. 1990;132:734-45.) Measurement error has two effects. In addition to the potential for substantial bias, confidence intervals of correctly adjusted estimates will be wider, sometime substantially, than naive confidence intervals. This problem is now well described in the particular field of nutrition and food frequency issues in Willett C. Nutritional Epidemiology Second Edition. New York: Oxford; 1998 (chap 12). There are now many articles on this subject, and an especially good review by Ray Carroll “Measurement Error in Epidemiologic Studies” in the Encyclopedia of Biostatistics.