HJMPH
Hawai'i Journal of Medicine & Public Health

General Recommendations on Data Presentation and Statistical Reporting (Biostatistical Guideline for HJMPH)

[Adapted from Annals of Internal Medicine & American Journal of Public Health]

The following guidelines are developed based on many common errors we see in manuscripts submitted to HJMPH. They are not meant to be all encompassing, or be restrictive to authors who feel that their data must be presented differently for legitimate reasons. We hope they are helpful to you; in turn, following these guidelines will reduce or eliminate the common errors we address with authors later in the publication process.

Percentages: Report percentages to one decimal place (eg, 26.7%) when sample size is >=200. For smaller samples (<200), do not use decimal places (eg, 26%, not 26.7%), to avoid the appearance of a level of precision that is not present.

Standard deviations (SD)/standard errors (SE): Please specify the measures used: using "mean (SD)" for data summary and description; to show sampling variability, consider reporting confidence intervals, rather than standard errors, when possible to avoid confusion.

Population parameters versus sample statistics: Using Greek letters to represent population parameters and Roman letters to represent estimates of those parameters in tables and text. For example, when reporting regression analysis results, Greek symbol (β), or Beta (b) should only be used in the text when describing the equations or parameters being estimated, never in reference to the results based on sample data. Instead, one can use "b" or β for unstandardized regression parameter estimates, and "B" or β for standardized regression parameter estimates.

***P* values:** Using *P* values to present statistical significance, the actual observed *P* value should be presented. For *P* values between .001 and .20, please report the value to the nearest thousandth (eg, *P*=.123). For *P* values greater than .20, please report the value to the nearest hundredth (eg, *P*=.34). If the observed *P* value is great than *.999*, it should be expressed as "*P*>*.99*". For a *P* value less than .001, report as "*P*<.001". Under no circumstance should the symbol "NS" or "ns" (for not significant) be used in place of actual *P* values.

"Trend": Use the word trend when describing a test for trend or dose-response. Avoid using it to refer to *P* values near but not below .05. In such instances, simply report a difference and the confidence interval of the difference (if appropriate), with or without the *P* value.

One-sided tests: There are very rare circumstances where a "one-sided" significance test is appropriate, eg, non-inferiority trials. Therefore, "two-sided" significance tests are the rule, not the exception. Do not report one-sided significance test unless it can be justified and presented in the experimental design section.

Statistical software: Specify in the statistical analysis section the statistical software used for analysis (version, manufacturer, and manufacturer's location), eg, SAS software, version 9.2 (SAS Institute Inc., Cary, NC).

Comparisons of interventions: Focus on between-group differences, with 95% confidence intervals of the differences, and not on within-group differences.

***Post-hoc* pairwise comparisons**: It is important to first test the overall hypothesis. One should conduct *post-hoc* analysis if and only if the overall hypothesis is rejected.

Clinically meaningful estimates: Report results using meaningful metrics rather than reporting raw results. For example, instead of the

log odds ratio from a logistic regression, authors should transform coefficients into the appropriate measure of effect size, eg, odds ratio. Avoid using an estimate, such as an odds ratio or relative risk, for a one unit change in the factor of interest when a 1-unit change lacks clinical meaning (age, mm Hg of blood pressure, or any other continuous or interval measurement with small units). Instead, reporting effort for a clinically meaningful change (eg, for every 10 years of increase of age, for an increase of one standard deviation (or interquartile range) of blood pressure), along with 95% confidence intervals.

Risk ratios: Describe the risk ratio accurately. For instance, an odds ratio of 3.94 indicates that the outcome is almost 4 times as likely to occur, compared with the reference group, and indicates a nearly 3-fold increase in risk, not a nearly 4-fold increase in risk.

Longitudinal data: Consider appropriate longitudinal data analyses if the outcome variables were measured at multiple time points, such as mixed-effects models or generalized estimating equation approaches, which can address the within-subject variability.

Sample size, response rate, attrition rate: Please clearly indicate in the methods section: the total number of participants, the time period of the study, response rate (if any), and attrition rate (if any).

Tables (general): Avoid the presentation of raw parameter estimates, if such parameters have no clear interpretation. For instance, the results from Cox proportional hazard models should be presented as the exponentiated parameter estimates, (ie, the hazard ratios) and their corresponding 95% confidence intervals, rather than the raw estimates. The inclusion of *P*-values in tables is unnecessary in the presence of 95% confidence intervals.

Descriptive tables: In tables that simply describe characteristics of 2 or more groups (eg, 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.

Figures (general): Avoid using pie charts; avoid using simple bar plots or histograms without measures of variability; provide raw data (numerators and denominators) in the margins of meta-analysis forest plots; provide numbers of subjects at risk at different times in survival plots.

Missing values: Always report the frequency of missing variables and how missing data was handled in the analysis. Consider adding a column to tables or a footnote that makes clear the amount of missing data.

Removal of data points: Unless fully justifiable, all subjects included in the study should be analyzed. Any exclusion of values or subjects should be reported and justified. When influential observations exist, it is suggested that the data is analyzed both with and without such influential observations, and the difference in results discussed.