JVIR
Journal of Vascular and Interventional Radiology
Society of Interventional Radiology

Guidelines for statistical methods for *JVIR*

Formatting and style of reporting statistical methods and results

- *JVIR* recommends preparing statistical methods and results sections of the manuscript in accordance with ICMJE (International Committee on Medical Journal Editors) and SAMPL (Statistical Analysis and Methods in the Published Literature) guidelines. The authors are strongly recommended to review these two brief documents and assure abiding by the guidelines prior to submission of the manuscript. The links to both documents are provided below:

  ICMJE guideline:
  http://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html#d

  SAMPL guidelines:
  https://www.equator-network.org/wp-content/uploads/2013/07/SAMPL-Guidelines-6-27-13.pdf

- The numbers in the manuscript should be checked for integrity by the authors prior submission of the manuscript:
  - Total percentages should add up to 100%.
  - At-risk numbers should match with the presented Kaplan-Meier graph.
  - Reporting of summarization data should be consistent throughout the manuscript (Choose either mean ± SD or median (range) or any other summarization method and keep it consistent throughout the manuscript). Any deviation from the chosen method needs to be justified.
- Manuscripts with data integrity issues will be returned to the authors for additional review and correction prior to further review by the journal.

Data Quality Assessment

We strongly encourage the authors to provide a *statement of data integrity assessment* in the "Statistical methods" section, as appropriate.

If the manuscript involves any kind of statistical data analysis, integrity of study data should be evaluated by the authors prior to analysis to assure appropriate validity of collected data. To that

end, we recommend the authors report the following in their methods and results sections as appropriate:

- *General "Methods and materials" section:*
  - Method of data collection (e.g. copying from printed sheets or other resources such as electronic medical records (EMR), data entry software applications such as RedCap, automated query from EMR, etc.)
  - Data collection Software application (Microsoft Excel, Microsoft Access, etc.).
  - Data entry quality assurance (e.g. double checking the entered data by a second observer, etc.)
- *Statistical methods section:*
  - Data integrity assessment methods prior to analysis (e.g. tabulating or visually exploring range and outliers for numerical and tabulation for categorical variables; bivariate cross-tabulation to assess reasonable integrity of related variables).
  - Treatment of missing values (e.g. missing value imputation, etc.) if applied.
  - Mitigation methods of detected inconsistencies (e.g. negative survival time, or unusually large or small values, etc.)
- *Results section to reflect:*
  - Number of missing values for each reported variable.
  - Number of subjects included in each analysis (e.g. each regression model).

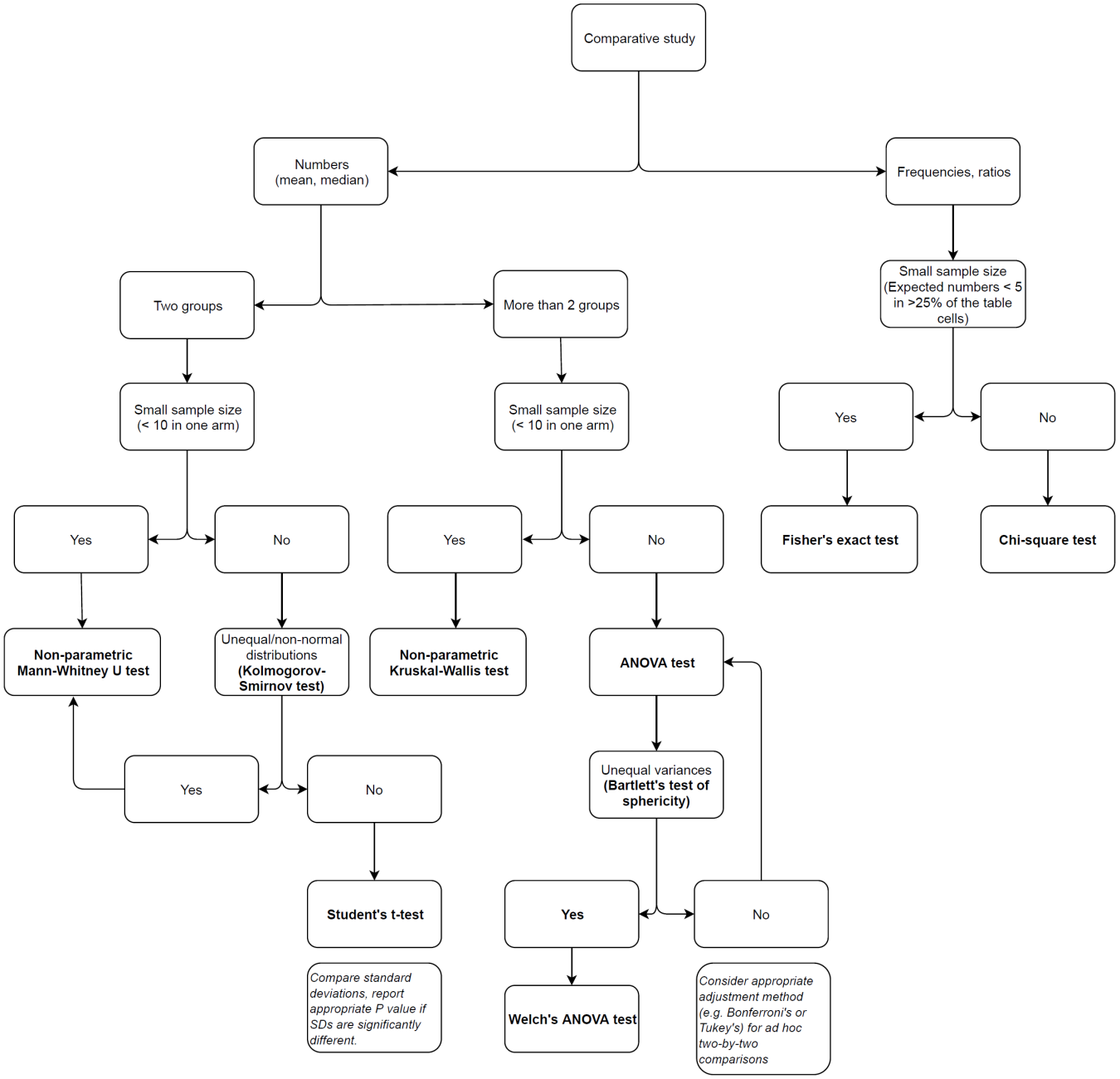
Choice of Statistical Tests; Assumption Checks and Other Considerations

- **Compatibility of tests with summarization:**
  - The applied statistical tests should be compatible with the data summarization method: e.g. Student's t-test, ANOVA for mean (SD), Mann-Whitney U test, Kruskal-Wallis test for median (range).
- **Comparative studies:**
  - For comparative analysis, please review the comparative study data flow chart at the end of this guideline for a general hint on the appropriate tests and considerations that would fit your data best before submitting your manuscript to *JVIR*.
- **Regression analysis:**
  - Considerations for multivariable models:
    - Independence of the covariates: two-by-two associations between variables entered in the model be checked using a separate regression or correlation analysis model. In case of significant interaction between the variables, their interaction term added to the model.
    - Results of univariate models need to be reported either in the same table or in a separate table before multivariable results for comparison.
  - Reporting of regression results:
    - Units of measurements for numerical variables as well as the reference category for the categorical variables need to be reported in the regression table.

        - Each regression table should be readable independently from the body text. The title of the regression table should implicate the outcome variable.
- **Survival analysis:**
    - Univariate comparison of the survival functions (e.g. comparing the Kaplan-Meier survival graphs with the Log-rank test) between the main study groups is recommended prior to multivariable modeling.
    - Considerations for survival regression (Cox) modeling:
        - Important assumption #1: proportionality of hazard is one of the main assumptions of the Cox regression models, hence the name proportional hazard model. This assumption should be tested whenever a Cox modeling method is used. Statistical (Schoenfeld's residual test) or graphical (log-minus-log plot) method or both can be used for this evaluation. Hazard proportionality for each covariate needs to be assessed in a univariate setting, and if statistically / visually significant time interaction is found, the variable needs to be considered as the "time-varying variable" in all subsequent models.
        - Important assumption #2: Like regression analysis, independence of the covariates needs to be checked in separate regression or correlation analysis models, and if significant interaction is found, addition of the interaction term to the multivariable model needs to be considered.
        - Think about any events that may 'compete' with your event of interest (e.g. post-ablation liver transplantation competing with death). These events are called "competing risk events" and require consideration in the analysis using "Competing risk Cox models". Please consult with your statistician if there is possibility of such events in the study.
- **Selection bias:**
    - Confounding effect from the selection bias cannot be resolved simply with multivariable adjustment. Therefore, it is strongly recommended to use an appropriate matching method (such as propensity score matching, propensity score weighting, etc.) if such a confounding effect is detected.
    - When using propensity score, post-matching / weighting balance of the applied covariates should be reported using an appropriate method such as Standardized Mean Difference (SMD). Any unbalanced variables after matching or weighting, needs to be added to the final analysis for a multivariable model.
- **Multiple comparisons:**
    - Exploratory studies (such as –Omics data analysis) are exposed to the incidental significant association finding depending on the number of the variables that they evaluate. P value significance level adjustment using an appropriate method such as False Discovery Rate (FDR) is needed for such studies. The results can be summarized in a Smile plot.
- **Large sample size:**
    - Although large sample size (usually over 1000 cases) are valuable source of research, there are issues with them as well. One of the most important issues with such studies is shrinkage of the type one error. In practical sense, it means that small associations can become statistically significant merely due to the large size of the study population. This issue is commonly seen in studies using

insurance or other national registry data. In such studies, the *'effect sizes'* of the studied associations should be calculated and reported beside the P values.

- **Meta-analysis:**
    - There is a debate of using random-effect versus fixed-effect models in meta-analysis studies. Random-effect models assume that the effect size varies among studies and meta-analysis tries to calculate the average effect size. In contrast, fixed-effect models assume an identical effect size and sampling error being the only reason for variation between studies hence largely ignoring effect size from the smaller studies. Since it is very unlikely that such condition will be true in a real-world scenario, it is highly recommended that random-effect models be applied in all meta-analyses unless the authors have a robust evidence of a fixed effect.
- **Novel measurement:**
    - When presenting a novel measurement (e.g. a radiographic index, etc.) in a study that demonstrates a significant association with an outcome (e.g. post-treatment tumor progression), the validity of the measurement should also be demonstrated. It is recommended that the measurement is repeated blindly by two or more observers and the reliability indices such as Inter-rater reliability, intraclass correlation coefficient (ICC), inter-observer agreement or standard error of measurement (SEM) be calculated and reported.


Comparative study
Numbers (mean, median)
Frequencies, ratios
Two groups
More than 2 groups
Small sample size (Expected numbers < 5 in >25% of the table cells)
Small sample size (< 10 in one arm)
Small sample size (< 10 in one arm)
Yes
No
Yes
No
Yes
No
Fisher's exact test
Chi-square test
Non-parametric Mann-Whitney U test
Unequal/non-normal distributions (Kolmogorov-Smirnov test)
Non-parametric Kruskal-Wallis test
ANOVA test
Yes
No
Unequal variances (Bartlett's test of sphericity)
Student's t-test
Yes
No
Compare standard deviations, report appropriate P value if SDs are significantly different.
Welch's ANOVA test
Consider appropriate adjustment method (e.g. Bonferroni's or Tukey's) for ad hoc two-by-two comparisons

Further readings:

1. Lantz B. The large sample size fallacy. Scand J Caring Sci. 2013 Jun;27(2):487-92.
2. A B Haidich. Meta-analysis in medical research. Hippokratia. 2010 Dec;14(Suppl 1):29-37.
3. Margarita Stolarova, Corinna Wolf, Tanja Rinker, Aenne Brielmann. How to assess and compare inter-rater reliability, agreement and correlation of ratings: an exemplary analysis of mother-father and parent-teacher expressive vocabulary rating pairs. Front Psychol. 2014 Jun 4;5:509.
4. Fixed-Effect Versus Random-Effects Models. In: Introduction to Meta-Analysis. Michael Borenstein M, Hedges LV, Higgins JPT and Rothstein HR. 2009 John Wiley & Sons, Ltd.
5. Newson R. Multiple-test procedures and smile plots. The Stata Journal. 2003; 3(2):109-132.
6. Imbens/Wooldridge. Estimation of Average Treatment Effects Under Unconfoundedness. Lecture Notes 1, Summer '07. https://www.nber.org/WNE/lect_1_match_fig.pdf Accessed: 9/11/2020.