

# A Basic Analysis Plan for Examining Bivariate and Multivariate Associations

Four Easy Steps to Greatness

Rod Hayward

One of the most common types of data analysis and statistical comparisons is the examination of bivariate (raw or unadjusted) and multivariate (independent or adjusted) associations (e.g. "Does surgical mortality increase with increasing age, and if so, is this association independent of sex, health status and comorbidities?"). A basic understanding of how to do such analyses is important for a researcher to know, and are easy to learn and do. However, for every 20 studies that use such analyses, perhaps 1 or 2 produce meaningful results. A high quality study comes from asking good questions and making sure that the variables put into the analytic model are accurate and complete, not by doing fancy statistics (by the way, the title is meant to be tongue in cheek). The analysis must be done correctly, but first you should ask yourself whether it needs to be done at all.

Below, you will find a check list of questions to ask as you plan and do an analysis of bivariate and independent associations.

## ***Step 1. Carefully Define the Study Objective/Question, Hypothesis & Conceptual Model.***

- a) What is the precise association that you wish to address and how would you state this as a study objective/question?
- b) Why have you selected this study question?
- c) What is your idealized model/design (i.e., what will affect or influence my outcome measure [DepVar] and how)?
- d) What is your study model/design (i.e., which components of my theoretical model are not adequately measured and how important are these omissions)?
- e) Based upon your theoretical model, what do you expect to find?

## ***Step 2. Develop a thorough understanding of your data's structure and the distribution of key variables?***

- a) What does one observation in your dataset represent (one encounter, one person, one person-year, one clinic, etc)? Is this the unit of analysis that makes sense for your data analysis?
- b) Explore the variables in the dataset using *-list- -tab- -sum,d-* (are they complete, sensible, skewed, etc)
- c) Make a codebook and/or some tables of the descriptive (univariate) statistics.

## ***Step 3. What is the magnitude and statistical significance of the bivariate association(s) pertaining to the study objective/question?***

- a) What specific DepVar(s) and IndVar(s) are needed to address my study objective/question?
- b) What statistical comparisons and presentations of effect size should I use to examine the bivariate association (also known as the "raw" or "unadjusted" associations)? Refer to handouts of "Measures of Effect Size & Associations" and "What Statistical Test Should I Use?". Make sample tables before you start the data analysis.
- c) Are the data in the correct format needed for the statistical comparison(s)? If you need to make new variables, use *-gen- -recode- -label-* to make new variables for your analysis, but do not overwrite variables in the original dataset.
- d) Run the analyses and fill in the table summarizing bivariate effect sizes and statistical significance.
- e) Always look at the raw data (scatter plots, cross tabs, etc) and do not just look at the p values and summary statistics.

**Step 4. What Are the Effect Sizes & Statistical Significance of the Independent (Multivariate) Association(s)?**

- a) What statistic comparisons and presentations of effect size should be used? Refer to handout on "What Statistical Test Should I Use?". Make sample tables before you start the analyses.
- b) Are the data in the correct format needed for the statistical comparisons? For dichotomous variables, make them "0/1 variables.
- c) Have I included all available confounders in the model? Have I included things that may be pathways and not confounders at all?
- d) Make a table summarizing the multivariate (independent or adjusted) effect sizes and statistical significance.