USING BIOSTATISTICS AND ANALYZING QUANTITATIVE DATA

5

Yardlee S. Kauffman and Daniel M. Witt

"The most confusing part to me about conducting my research project was determining what statistical test is most appropriate."

—Former PGY2 Ambulatory Care Resident

LEARNING OBJECTIVES

- Differentiate between continuous and discrete variables.
- Differentiate between descriptive and inferential statistics.
- Given data from a research project, select an appropriate statistical test for comparing study outcomes.
- Identify software that can be used to assist with statistical analysis.
- Identify when to consult someone with biostatistics expertise.

INTRODUCTION

Frequently, the most challenging component of a research project is determining the appropriate statistical approach to analyze the data. Establishing a reasoned analytical plan prior to collecting study data will help to minimize frustrations as your project progresses. As discussed in Chapter 1, the study objectives outlined in the study protocol should serve as a guide for your analytical plan.

Evidence-based medical interventions are important to achieve optimal patient outcomes and utilize healthcare resources effectively. Assessing the value of interventions based on clinical experience alone can lead to erroneous conclusions due to biologic variability, the influences of the placebo effect, and various sources of confounding and/or bias (see Chapter 2). Most resident

DOI 10.37573/9781585285617.005

projects will involve choosing a sample of patients to study from a larger population. *Biostatistics* provide a framework for determining the extent to which chance or bad luck in drawing your study sample from a larger population influences any observed differences between the study groups. When biostatistics are correctly applied to research data, you are able to quantify the likelihood that any differences observed between the groups being studied are due to chance alone. When biostatistics helps you to rule out chance as the explanation for any observed differences between groups, your results are said to be *statistically significant*.

VARIABLE TYPES

Your project will likely involve collecting independent and dependent data variables. Examples of *independent variables* include interventions or processes that you are interested in studying such as a new pharmacy diabetes service, a new process for administering vaccines, a quality improvement program, a target drug program, an intravenous-to-oral antibiotic conversion program, or a new drug that requires less monitoring. *Dependent variables* are what you expect to change as a result of your independent variable; for example, glucose control, vaccination rates, quality metrics, pharmacy costs, length of stay, or quality of life.

The first step in defining an analytical plan using biostatistics is to determine what types of dependent and independent data variables you will be collecting during your study. The term *variable* refers to the fact that you cannot predict the outcomes of your project beforehand. There are two broad types of variables: discrete and continuous.

Discrete Variables

Discrete variables can often can be placed into distinct categories; they are derived by counting outcomes fulfilling various criteria.[1] Discrete variables can be subcategorized as dichotomous, nominal, and ordinal:

- *Dichotomous variables* have only two possible outcomes that are mutually exclusive and exhaustive. Examples include mortality, sex category, pregnancy status, history of various comorbid conditions, and assignment to experimental versus control group.
- *Nominal variables (sometimes referred to as categorical variables)* have three or more possible outcomes that are mutually exclusive and exhaustive, but have no particular order ranking of severity. Examples include eye color, gender identity, and race category.
- *Ordinal variables* have two or more categories that are ranked in a specific order or severity, but the scale between ranks is arbitrary. For example, New York Heart Association class IV heart failure is not twice as severe as class II heart failure. Other examples of ordinal variables include Likert scale responses and severity scales, such as CHA_2DS_2-VASc and HAS-BLED.

Dichotomous and nominal variables are often summarized using percentages or proportions. Ordinal variables are also summarized using percentages or proportions when there are only two categories, but they should be summarized using median and interquartile range (IQR) (i.e., the 25th and 75th percentiles) when there are more than two categories.

Continuous Variables

Continuous variables measure quantity and are generally derived from measurements taken during a study.[1] Continuous variables can take on any value within a given interval and are arranged in a specific order with a consistent change in magnitude between units of measure. Continuous data are sometimes subcategorized into ratio and interval data. *Ratio data* have an absolute minimum or zero point. In essence, when the variable equals zero, there is a complete absence of the effect being measured. For example, a plasma drug concentration of zero indicates there is no measurable