JOURNAL SERIES:

Statistics Review Part 1: Types of Variables, Incidence, and Prevalence

by Megan Heim, PharmD; Kevin Look, PharmD, MS; Amanda Margolis, PharmD, MS, BCACP

Statistics, or the collection, measurement, analysis, interpretation, and presentation of data, can be daunting for many practitioners. However, it is vitally important for practicing clinicians to be acquainted with basic statistical concepts in order to be able to properly interpret and evaluate the findings from research articles. This includes being able to interpret statistics, knowing how statistics are calculated, and recognizing whether their use is appropriate. With this knowledge, the reader will be able to better recognize the quality of evidence generated by studies, identify whether a study is designed well (or poorly), and appreciate the importance of a study's findings.

This is the first article in a series of articles designed to review the basic, fundamental concepts of biostatistics. Topics that will be covered in the series will include standard deviations and confidence intervals, power and statistical significance, and more. This article describes the different types of variables found in research, and discusses the concepts of incidence and prevalence. Familiarity with these concepts is required for understanding concepts in later articles.

Objectives

1. Define dependent and independent variables.
2. Categorize types of variables as either nominal, ordinal, interval, or ratio variables.
3. Describe the difference between the concepts of incidence and prevalence.

Types of Variables

It is important to understand statistical terminology in order to understand and/or perform statistical analyses utilized in medical research. Understanding the different types of variables that exist is an important first step in understanding research. A variable is a characteristic of a unit that may assume more than one set of values (e.g., age, weight, income, etc.). Variables are classified as either independent or dependent variables. In research studies, the independent variable is the variable that is varied or manipulated by the researcher. For example, patients in a clinical trial may be assigned to take a placebo or an anti-hypertensive medication. The type of study medication that the patient is taking is an independent variable. The dependent variable is the outcome or response that is measured, such as blood pressure. In research, we are typically interested in understanding the effect of the independent variable (or treatment) on the dependent variable (or outcome); for example, what is the effect of an anti-hypertensive medication on blood pressure?

Variables can be further broken down into categorical and continuous variables. A categorical variable has values that function as labels rather than as numbers. An example of a categorical variable is blood type (A, B, AB, or O). A continuous variable has numeric values, and the relative magnitude of the values is significant. This means that data can be ranked low to high or high to low. An example of a continuous variable is body weight.

Continuous and categorical variables can be further classified into sub-categories. Subtypes of categorical variables include nominal and ordinal variables. A nominal variable is a non-continuous and non-ordered variable. For nominal variables, a number or code is arbitrarily assigned to characteristics for data categorization. As an example, gender may be defined as 1 for "female" and 2 for "male." The number or code selected is purely arbitrary and designated at the discretion of the researcher without regard to any rank order.

For ordinal variables, numbers are used to denote rank-order, without defining a magnitude of difference between numbers. In other words, categories are ranked in a specific order. For example, data from a questionnaire can be categorized as 5 for "strongly agree," 4 for "agree," 3 for "neutral," 2 for "disagree," and 1 for "strongly disagree." For ordinal variables, the difference between categories is not linear; for example, the difference between "strongly agree" and "agree" may not be the same as the difference between "agree" and "neutral." Numbers assigned to these categories are not arbitrary, as the order of the numbers is meaningful. Disease classification systems are often ordinal, such as the New York Heart Association (NYHA) Heart Failure classification system.

For continuous variables, sub-categories include interval and ratio variables. As discussed previously, a continuous variable has numeric values, and the relative magnitude of the values is significant. This means that data can be ranked low to high

or high to low and, additionally, data units must follow a linear order. For interval variables, numbers represent units of equal magnitude and rank order on a scale without an absolute zero. In other words, there is a consistent level of magnitude of difference between the data units; however, the zero point on the scale is arbitrary. Additionally, interval variables have a defined unit of measure. The Fahrenheit scale for temperature is an example of an interval variable. The zero value is arbitrary, the number of the scale is consistent, and there is a defined unit of measure.

For ratio variables, numbers signify units of equal magnitude and rank order on a scale with an absolute zero. Like interval variables, there is a consistent level of magnitude of difference between the data units. As opposed to interval variables, the absolute zero for ratio variables is not arbitrary. Examples of ratio variables include blood pressure and heart rate.

TABLE 1. Definitions and examples of types of variables[9]

Variable	*Sub-Category*	*Definition*	*Example*
Categorical	Nominal	The number or code is arbitrarily assigned to characteristics for data categorization	Gender classification: 1 for "female" and 2 for "male" Disease classification: 1 for "hypertension," 2 for "hyperlipidemia," and 3 for "diabetes"
	Ordinal	A number is used to denote rank-order, without defining a magnitude of difference between numbers	Survey questionnaire categorization: 5 for "strongly agree," 4 for "agree," 3 for "neutral," 2 for "disagree," and 1 for "strongly disagree" NYHA heart failure classifications levels: class I, II, III, or IV
Continuous	Interval	Numbers represent units of equal magnitude and rank order on a scale *without* an absolute zero	Fahrenheit scale for temperature
	Ratio	Numbers signify units of equal magnitude and rank order on a scale *with* an absolute zero	Blood pressure, heart rate, weight, height

Incidence and Prevalence

Two other concepts that are important to understand when assessing medical literature and statistical methods and that are commonly confused are incidence and prevalence. Incidence is a measure of the rate of occurrence of a condition. In other words, incidence is the number of new cases of a condition that develop during a specific period of time. For example, the incidence of diabetes may be documented as the number of new cases of diabetes per year within a given sample or population. Incidence can be measured using methods such as randomized controlled trials (e.g., to determine the incidence of adverse effects of a medication) or cohort studies, which follow groups of patients over time.

Prevalence, on the other hand, is the proportion of a population found to have a condition at a single point in time. For example, the prevalence of diabetes may be documented as the percentage of the United States population with diabetes. Prevalence is measured using cross-sectional study designs, such as a survey or a census, or by using administrative data such as medical, hospital, or prescription drug claims.

Summary

This article reviewed some fundamental concepts and terminology of statistical methods, including types of variables, incidence, and prevalence. The next article in this series will cover relative risk, relative risk reduction, absolute risk reduction, and number needed to treat. ●

Practice questions

1. The amount of time in minutes it takes to walk one mile is an example of what kind of variable?
   a. Nominal
   b. Ordinal
   c. Interval
   d. Ratio
2. Classifying educational experience into categories including "high school," "college," and "graduate school," is an example of which type of variable?
   a. Nominal
   b. Ordinal
   c. Interval
   d. Ratio
3. "One out of every three adults in the United States has hypertension." This statement is an example of which type of the following concepts?
   a. Incidence
   b. Prevalence

Answers:

1. **d** The time in minutes it takes to walk one mile is a continuous variable as it is expressed in a numeric value and the magnitude of the value is significant. The type of continuous variable is a ratio variable because there is an absolute zero.
2. **b** Classifying educational experience into categories is a categorical variable as the value functions as a label rather than a numeric value. The type of categorical variable is ordinal since categories can be ranked in a specific order.
3. **b** This statement is an example of prevalence since the statement is an expression of a proportion of a population found to have a condition at a single point in time.

Megan Heim is a PGY2 Ambulatory Care Pharmacy Resident at the William S. Middleton Memorial Veterans Hospital in Madison, WI. Amanda Margolis is a Lecturer at the UW-Madison School of Pharmacy and a Clinical Pharmacist at the William S. Middleton Memorial Veterans Hospital in Madison, WI. Kevin Look is a PhD Candidate in the Social and Administrative Sciences Division at the UW-Madison School of Pharmacy in Madison, WI.

References and suggestions for further review:

1. Gaddis GM, Gaddis ML. Introduction to Biostatistics: Part 1, Basic Concepts. Annals of Emergency Medicine. 1990;19:86-89.
2. Gaddis GM, Gaddis ML. Introduction to Biostatistics: Part 2, Descriptive Statistics. Annals of Emergency Medicine. 1990;19:309-315.