

## 1 Introduction

Reading: SW Chapter 1

### What is statistics/biostatistics/biometry?

Examples of medical and research problems:

1. A couple is deciding whether or not to have a child, because of the existence of certain diseases within the family. With present understanding of genetics, they are told that the probability that a child of theirs having this defect is 0.01. What might they want to do? How would the type of disease affect this? What if the probability is 0.50? What other factors besides probabilities and the type of disease would be pertinent?
2. Research question: Is HPV (human papilloma virus) a risk factor for cervical dysplasia? How does one approach answering this question? One possibility: Becker et al. (1994) conducted a case-control study. The women in the study were patients at UNM clinics. The 175 cases were women, aged 18-40, who had cervical dysplasia. The 308 controls were women aged 18-40 who did not have cervical dysplasia. Each women was classified as positive or negative, depending on the presence of HPV. The data collected from the study are summarized below.

HPV Outcome	Cases	Controls
Positive	164	130
Negative	11	178
Sample size	175	308

The results can be summarized in a number of ways. The proportion positive among cases is  $164/175 = 0.94$ . the proportion positive among controls is  $130/308 = 0.42$ . This gives an odds ratio of  $164 * 178 / (11 * 130) = 20.4$ . Do these results indicate that HPV is a risk factor for cervical dysplasia?

3. Research question: Is a new drug more effective in treating an illness than a previously used drug? How to approach this question? One possibility: conduct a clinical trial (Phase II) with one treatment group where all patients receive the new drug. The old drug has an assumed cure rate obtained from repeated use of this treatment.

**Outcomes and conclusions:** Assume old drug cures 70%. If 9 people out of 10 with the illness were cured with new treatment, then what would you conclude? If 6 were cured? If 90 out of a sample of 100?

**Alternative possibility:** conduct a clinical trial (Phase III) with two groups (new treatment, old treatment), and randomize patients to the two groups.

**Other possible outcomes of interest:** reduction in fever, pain, itching of skin rash in 24-hour period (quantify reduction), reduction in tumor size.

## SO WHAT DOES STATISTICS LEND TO THESE PROBLEMS?

### 1. What is statistics?

- Statistics is concerned with the STUDY, DESCRIPTION, and MANAGEMENT of variability.
- There are many ways to define statistics, but common components in the definitions are: variation; uncertainty; inference.
- Biostatistics is the subset of statistics that is concerned with applications in biological/medical areas.

### 2. What should you get out of an introductory course in biostatistics?

- Understand basic statistical concepts
- Be able to read papers in your field and understand the statistical results, and, hopefully, the statistical methods that were used.
- Be able to determine appropriate statistical methods to use and implement them – in simple analyses.
- Be able to determine when you can't do something and seek out help from a statistician.

## ASPECTS OF STATISTICS THAT WE WILL BE CONCERNED WITH

- Descriptive statistics and exploratory data analysis: ways to describe data using graphical displays and numerical summaries
- Basic ideas of probability as a means of quantifying uncertainty
- Statistical inference: wish to draw some conclusions from data, based on hypothesis testing and estimation methods.

### Types of data/situations we will examine:

- data on one continuous variable (one, two and multiple samples)
- discrete data (single sample and two-way tables, including logistic regression)
- data on two or more continuous variables (linear regression and correlation, and survival analysis)