An Introduction to Statistical Methods to Support Evidence-Based Public Health

2012 Kansas Public Health Association Conference
Pre-Session: Analytic Tools for Public Health

Jo A. Wick, PhD
Assistant Professor of Biostatistics
University of Kansas Medical Center
Agenda

• Evidence-based practice in public health involves
  – Gathering evidence in the form of scientific data
  – Applying the scientific method to inform policy development, establishment or development of new programs aimed at improving public health
Agenda

• Evidence-based practice results in
  – Implementation of programs or policies with a high likelihood of success
  – More efficient use of public and private resources
Agenda

• Today, we will look at
  – An introduction to the concept of evidence-based public health
  – An overview of common analytic tools used by public health researchers to report the effects of an intervention, program or policy
  – Motivating examples from the public health literature to demonstrate the proper evaluation of competing evidence
Objectives

• Understand the history and role of evidence-based practice in public health
• Calculate and interpret epidemiologic measures of disease occurrence
• Calculate and interpret measures of effect used to compare the risk of disease between populations and subgroups
Objectives

• Recognize features, strengths and limitations of various types of study designs
• Differentiate between different levels of scientific evidence
• Understand the roles of chance, bias and confounding in the evaluation of literature
Evidence-Based Public Health
Today's Random Medical News

According to a report released today...

Jim Borgman
The Cincinnati Enquirer
King Features Syndicate
What is Public Health?

- “the science and art of preventing disease, prolonging life and promoting health through the organized efforts and informed choices of society, organizations, public and private, communities and individuals”\(^1\)

- Multidisciplinary
  - Epidemiology
  - Biostatistics
  - Health services
Early Public Health

- Religious restrictions on certain behaviors
  - Food
  - Indulgent behavior

- John Snow: Birth of Epidemiology
  - 1854 cholera outbreak in London
Modern Public Health

• Population based: 30-year gain in life expectancy in the US during the 20th Century
  – Safe water and food
  – Sewage treatment and disposal
  – Tobacco use prevention and cessation
  – Injury prevention
  – Control of infectious disease through immunization, etc.
Modern Public Health

• Prevent or resolve chronic disease
  – HIV/AIDS (communicable)
  – Diabetes
  – Cardiovascular disease
  – Cancer
  – Depression/mental illness

• Eradicate or prevent transmission of communicable disease
  – Waterborne diseases (e.g., malaria)
  – Influenza
  – STDs
  – Measles
Evidence-Based Policy

• In general, evidence-based policy is defined as the “incorporation of scientific evidence in selecting and implementing programs, developing policies, and evaluating progress”
Evidence-Based Public Health

• . . . the ‘process of integrating science-based interventions with community preferences to improve the health of populations.”^2
Evidence-Based Public Health

Decision-making

- Population characteristics needs, values, and preferences
- Best available research evidence
- Resources, including practitioner expertise
Evidence-Based Public Health

1. Community Assessment
2. Develop an initial statement of the issue
3. Quantify the issue
4. Determine what is known through the scientific literature
5. Develop and prioritize program and policy options
6. Develop an action plan and implement interventions
7. Evaluate the program or policy
The Scientific Method

Observe

Experiment

Revise Hypothesis
EBPH: Air Pollution

Exposure (e.g. air pollution)

Policy (e.g. air quality standards)

Health Effects

Public Health Impact Assessment
EBPH: Bullying

Exposure (e.g. bullying in schools)

Incident reports

Intervention implemented in school district

Intervention designed to reduce exposure
EBPH: Alcohol Consumption

Exposure (e.g., alcohol consumption)

Intervention designed (e.g., alcohol tax proposed to reduce consumption)

Intervention implemented (e.g., tax passed by city)

DUI reports Alcohol-related arrests
Evidence-Based Practice:
Hypothesis Testing

Evidence (Data)

Run experiment

Revise Hypothesis
Experiment

• An **experiment** is a process whose results are not known until after it has been performed.
  – The range of possible outcomes, \( e_1, \ldots, e_K \) are known in advance
  – We do not know the exact outcome, but would like to know the chances of its occurrence

• The **probability** of an outcome \( E \), denoted \( P(E) \), is a numerical measure of the chances of \( E \) occurring.

\[
0 \leq P(E = e_j) \leq 1 \quad \sum_{j=1}^{K} P(E = e_j) = P(e_1) + L + P(e_K) = 1
\]
Probability

- Relative frequency view:
  
  \[ P(E = e_j) = \frac{\text{# times } E = e_j}{\text{total # observations of } E} \]

- Probabilities for the outcomes of a random variable \( x \) are represented through a probability distribution:
Population Parameters

• Most often our research questions involve unknown population parameters:
  What is the average BMI among 5th graders in Wyandotte County, Kansas?
  What proportion of Kansas high-schoolers report being sexually active?

• To determine these values exactly would require a census.

• However, due to a prohibitively large population (or other considerations) a sample is taken instead.
Sample Statistics

- **Statistics** describe or summarize sample observations.
- They vary from sample to sample, making them **random variables**.
- We use statistics generated from samples to make **inferences** about the parameters that describe populations.
Sampling Variability

Population

μ
σ

Sampling
Distribution of \( \bar{x} \)

Samples

\[ \bar{x} = 0.1 \quad s = 0.92 \]

\[ \bar{x} = 0.5 \quad s = 1.02 \]

\[ \bar{x} = -0.1 \quad s = 0.98 \]
Types of Samples

• **Random sample**: each person has equal chance of being selected.

• **Convenience sample**: persons are selected because they are convenient or readily available.

• **Systematic sample**: persons selected based on a pattern.

• **Stratified sample**: persons selected from within subgroup.
Random Sampling

• For studies, it is optimal (but not always possible) for the sample providing the data to be representative of the population under study.

• Simple random sampling provides a representative sample (theoretically).
  – A sampling scheme in which every possible sub-sample of size $n$ from a population is equally likely to be selected
  – Assuming the sample is representative, the summary statistics (e.g., mean) should be ‘good’ estimates of the true quantities in the population.
    • The larger $n$ is, the better estimates will be.
Types of Data

• All data contains information.
• It is important to recognize that the hierarchy implied in the level of measurement of a variable has an impact on
  (1) how we describe the variable data and
  (2) what statistical methods we use to analyze it.
Levels of Measurement

- **Nominal**: discrete qualitative
- **Ordinal**: difference, order
- **Interval**: order, equivalence of intervals
- **Ratio**: difference, order, equivalence of intervals, absolute zero
Types of Data

- **Nominal**: Gender, Political Affiliation
- **Ordinal**: Rankings, Classification in college, Income (categorical), Distance
- **Interval**: Temperature in Celsius, Calendar Time
- **Ratio**: Time to complete a task, Production, Temperature in Kelvin

Information increases as the level of measurement increases.
Levels of Measurement

• The levels are in increasing order of mathematical structure—meaning that more mathematical operations and relations are defined—and the higher levels are required in order to define some statistics.

• At the lower levels, assumptions tend to be less restrictive and the appropriate data analysis techniques tend to be less sensitive.

• In general, it is desirable to have a higher level of measurement.
Hypothesis Testing

• **Null hypothesis “H₀”**: statement of no differences or association between variables
  – This is the hypothesis we test—the first step in the ‘recipe’ for hypothesis testing is to assume $H_0$ is true

• **Alternative hypothesis “H₁”**: statement of differences or association between variables
  – This is what we are trying to prove
Hypothesis Testing

• One-tailed hypothesis: outcome is expected in a single direction (e.g., administration of experimental drug will result in a decrease in systolic BP)
  – \( H_1 \) includes ‘<‘ or ‘>’

• Two-tailed hypothesis: the direction of the effect is unknown (e.g., experimental therapy will result in a different response rate than that of current standard of care)
  – \( H_1 \) includes ‘\( \neq \)’
Hypothesis Testing

• The statistical hypotheses are statements concerning characteristics of the population(s) of interest:
  – Population mean: $\mu$
  – Population variability: $\sigma$
  – Population rate (or proportion): $\pi$
  – Population correlation: $\rho$

• Example: It is hypothesized that the response rate for the intervention is greater than that in the control.
  – $\pi_{\text{Exp}} > \pi_{\text{SOC}} \leftarrow$ This is $H_1$. 
Decisions and Errors

• **Type I Error (α):** a true $H_0$ is incorrectly rejected
  - “An innocent man is proven GUILTY in a court of law”
  - Commonly accepted rate is $α = 0.05$

• **Type II Error (β):** failing to reject a false $H_0$
  - “A guilty man is proven NOT GUILTY in a court of law”
  - Commonly accepted rate is $β = 0.2$

• **Power (1 – β):** correctly rejecting a false $H_0$
  - “Justice has been served”
  - Commonly accepted rate is $1 – β = 0.8$
# Decisions and Errors

<table>
<thead>
<tr>
<th>Conclusion</th>
<th>Truth</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_1$</td>
<td>$H_1$: Correct: Power</td>
</tr>
<tr>
<td>$H_0$</td>
<td>$H_0$: Type II Error</td>
</tr>
</tbody>
</table>
Power and Sample Size

- As effect of interest $\downarrow$, required $n \uparrow$.
- As required power $\uparrow$, required $n \uparrow$.
- As type I error rate $\downarrow$, required $n \uparrow$.
- **Power Calculations** ← an interesting interactive web-based tool to show the relationship between power and the sample size, variability, and difference to detect.
Basic Recipe for Hypothesis Testing

1. State $H_0$ and $H_1$
2. Assume $H_0$ is true
3. Collect the evidence—from the sample data, compute the appropriate sample statistic and the test statistic
4. Determine if the test statistic is large enough to meet the a priori determined level of evidence necessary to reject $H_0$ ( . . . or, is $p < \alpha$?)
Example: Carbon Monoxide

• An experiment is undertaken to determine the concentration of carbon monoxide in air.
• It is hypothesized that the actual concentration is significantly greater than 10 mg/m$^3$.
• Eighteen air samples are obtained and the concentration for each sample is measured.
  – The random variable (outcome) $x$ is carbon monoxide concentration in the sample.
  – The characteristic (parameter) of interest is $\mu$—the true average concentration of carbon monoxide in air.
Step 1: State $H_0$ & $H_1$

- $H_1: \mu > 10 \text{ mg/m}^3 \hookleftarrow \text{We think!}$
- $H_0: \mu \leq 10 \text{ mg/m}^3 \hookleftarrow \text{We assume in order to test!}$

Step 2: Assume $\mu = 10$
Step 3: Evidence

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>10.25</td>
<td>10.37</td>
<td>10.66</td>
</tr>
<tr>
<td>10.47</td>
<td>10.56</td>
<td>10.22</td>
</tr>
<tr>
<td>10.44</td>
<td>10.38</td>
<td>10.63</td>
</tr>
<tr>
<td>10.40</td>
<td>10.39</td>
<td>10.26</td>
</tr>
<tr>
<td>10.32</td>
<td>10.35</td>
<td>10.54</td>
</tr>
<tr>
<td>10.33</td>
<td>10.48</td>
<td>10.68</td>
</tr>
</tbody>
</table>

Sample statistic: $\bar{x} = 10.43$

Test statistic: $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{10.43 - 10}{1.02/\sqrt{18}} = 1.79$

What does 1.79 mean? How do we use it?
Student’s $t$ Distribution

- Remember when we assumed $H_0$ was true?

Step 2: Assume $\mu = 10$
Student’s $t$ Distribution

- That assumption set up this theoretical Student’s $t$ distribution from which the $p$-value can be calculated:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{10 - 10}{1.02/\sqrt{18}} = 0$$
Student’s $t$ Distribution

- Assuming the true air concentration of carbon monoxide is actually 10 mg/mm$^3$, how likely is it that we should get evidence in the form of 18 samples of air with mean concentration 10.43?

$$P(x)$$

---

Step 2: Assume $\mu = 10$

$$P(\bar{x} \geq 10.43)$$
Student’s t Distribution

- We can say how likely by framing the statement in terms of the probability of an outcome:

\[
t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{10 - 10}{1.02/\sqrt{18}} = 0
\]

\[p\text{-value} = P(t \geq 1.79) = 0.0456\]
Step 4: Make a Decision

- Decision rule: if $p \leq \alpha$, the chances of getting the actual collected evidence from our sample given the null hypothesis is true are very small.
  - The observed data conflicts with the null ‘theory.’
  - The observed data supports the alternative ‘theory.’
  - Since the evidence (data) was actually observed and our theory ($H_0$) is unobservable, we choose to believe that our evidence is the more accurate portrayal of reality and reject $H_0$ in favor of $H_1$. 
Step 4: Make a Decision

- What if our evidence had not been in as great of degree of conflict with our theory?
  - $p > \alpha$: the chances of getting the actual collected evidence from our sample given the null hypothesis is true are pretty high
  - We **fail to reject** $H_0$. 
Decision

• How do we know if the decision we made was the correct one?
  – We don’t!
  – If $\alpha = 0.05$, the chances of our decision being an incorrect rejection of a true $H_0$ are no greater than 5%.
  – We have no way of knowing whether we made this kind of error—we only know that our chances of making it in this setting are relatively small.
What does this look like?

- For a single experiment, Assume air concentration is 10
  - $H_0: \mu = 10$
  - $H_1: \mu > 10$

Samples from air have mean concentration close to 10
  - For $n = 18$ samples, $\bar{x} \approx 10$

Evidence supports assumption
  - $p > 0.05$
  - Fail to reject $H_0: \mu = 10$
What does this look like?

• For a single experiment,

  Assume air concentration is 10
  • $H_0: \mu = 10$
  • $H_1: \mu > 10$

  Samples from air have mean concentration much greater than 10
  • For $n = 18$ samples, $\bar{x} \approx 11$

  Evidence conflicts with assumption
  • $p < 0.05$
  • Reject $H_0: \mu = 10$
Which test do I use?

• What kind of outcome do you have?
• How many samples do you have?
  – Are they related or independent?
# Types of Tests

<table>
<thead>
<tr>
<th>Measurement Level</th>
<th>Population Parameter</th>
<th>Hypotheses</th>
<th>Sample Statistic</th>
<th>Inferential Method(s)</th>
</tr>
</thead>
</table>
| Nominal           | Proportion $\pi$      | $H_0: \pi = \pi_0$  
                  |           | $H_1: \pi \neq \pi_0$  
                  |           | $p = \frac{x}{n}$  
                  |           | Binomial test or  
                  |           | $z$ test (if $np > 10 \& nq > 10$) |
| Ordinal           | Median $M$            | $H_0: M = M_0$  
                  |           | $H_1: M \neq M_0$  
                  |           | $m = p_{50}$  
                  |           | Wilcoxon signed-rank test |
| Interval          | Mean $\mu$           | $H_0: \mu = \mu_0$  
                  |           | $H_1: \mu \neq \mu_0$  
                  |           | $\bar{x}$  
                  |           | Student’s $t$ or Wilcoxon (if non-normal or  
                  |           | small $n$) |
| Ratio             | Mean $\mu$           | $H_0: \mu = \mu_0$  
                  |           | $H_1: \mu \neq \mu_0$  
                  |           | $\bar{x}$  
                  |           | Student’s $t$ or Wilcoxon (if non-normal or  
                  |           | small $n$) |
Types of Tests

• **Parametric methods**: make assumptions about the distribution of the data (e.g., normally distributed) and are suited for sample sizes large enough to assess whether the distributional assumption is met.

• **Nonparametric methods**: make no assumptions about the distribution of the data and are suitable for small sample sizes or large samples where parametric assumptions are violated.
  - Use ranks of the data values rather than actual data values themselves.
  - Loss of power when parametric test is appropriate.
# Types of Tests

<table>
<thead>
<tr>
<th>Measurement Level</th>
<th>Population Parameters</th>
<th>Hypotheses</th>
<th>Sample Statistics</th>
<th>Inferential Method(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal</td>
<td>π₁, π₂</td>
<td>H₀: π₁ = π₂ &lt;br&gt;H₁: π₁ ≠ π₂</td>
<td>$p_i = \frac{x_i}{n_i}$ &lt;br&gt;$p_j = \frac{x_j}{n_j}$</td>
<td>Fisher’s exact or Chi-square (if cell counts &gt; 5)</td>
</tr>
<tr>
<td>Ordinal</td>
<td>M₁, M₂</td>
<td>H₀: M₁ = M₂ &lt;br&gt;H₁: M₁ ≠ M₂</td>
<td>$m_1, m_2$</td>
<td>Median test</td>
</tr>
<tr>
<td>Interval</td>
<td>μ₁, μ₂</td>
<td>H₀: μ₁ = μ₂ &lt;br&gt;H₁: μ₁ ≠ μ₂</td>
<td>$\bar{x}_1, \bar{x}_2$</td>
<td>Student’s $t$ or Mann-Whitney (if non-normal, unequal variances or small $n$)</td>
</tr>
<tr>
<td>Ratio</td>
<td>μ₁, μ₂</td>
<td>H₀: μ₁ = μ₂ &lt;br&gt;H₁: μ₁ ≠ μ₂</td>
<td>$\bar{x}_1, \bar{x}_2$</td>
<td>Student’s $t$ or Mann-Whitney (if non-normal, unequal variances or small $n$)</td>
</tr>
</tbody>
</table>
Smoking Cessation

• Two types of therapy: $x = \{\text{behavioral therapy, literature}\}$

• Outcome: $y = \text{number of cigarettes smoked per day after six months of therapy}$
Smoking Cessation

- Research question: Is behavioral therapy in addition to education (1) better than education alone (2) in getting smokers to quit?
- $H_0: \mu_1 = \mu_2$ versus $H_0: \mu_1 \neq \mu_2$
Smoking Cessation

Education Only

Education + Behavioral Therapy

% Reduction
Smoking Cessation

Conclusion: Adding behavioral therapy to cessation education resulted in a significantly greater reduction in smoking at six months post-therapy when compared to education alone ($t_{30.9} = -2.87, p < 0.01$).

<table>
<thead>
<tr>
<th>Number of Cigarettes Smoked Per Day</th>
<th>Levene's Test for Equality of Variances</th>
<th>t-test for Equality of Means</th>
<th>95% Confidence Interval of the Difference</th>
</tr>
</thead>
</table>

Reject $H_0: \mu_1 = \mu_2$
Confidence Intervals

• . . . have a tricky interpretation.
• The meaning of a 95% confidence interval is not:
  – “The probability (or chances) that the true difference in smoking reduction is between LB (5%) and UB (17%) is 95%.”
Confidence Intervals

• Suppose we actually took sample after sample . . .
  – 100 of them, to be exact

  – 95% confident means: “In 95 of the 100 samples, our interval will contain the true unknown value of the parameter. However, in 5 of the 100 it will not.”
Confidence Intervals

• Suppose we actually took sample after sample . . .
  – 100 of them, to be exact
  – Our “confidence” is in the procedure that produces the interval—i.e., it performs well most of the time.
  – Our “confidence” is not directly related to our particular interval.
Quantitative Measures of Disease Occurrence in Populations
Measures of Disease Occurrence

• Absolute counts of disease incidence are important but make comparisons between groups difficult.
  – As $\uparrow N$, $\uparrow$ number of cases likely

• We need to consider the number of cases relative to the size of the population at risk.
Measures of Disease Occurrence

- Ratios: allows us to compare the number of people with disease in one population with the number with disease in another population.
- Proportions: fraction of people within a population with a certain characteristic.
- Rates: involves both a time frame of interest and a unit of the population.
## Ratio

- Female-to-male death ratio in US population for 2005:

<table>
<thead>
<tr>
<th>2005 All-cause mortality</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Females</td>
<td>269,368</td>
</tr>
<tr>
<td>Males</td>
<td>243,324</td>
</tr>
</tbody>
</table>

\[
\frac{269,368}{243,324} = 1.107
\]

- Interpretation: The male-to-female death ratio is 1:1.107; for every one male death in 2005, 1.107 female deaths occurred.
Proportion

• Proportion of female deaths in US population for 2005:

<table>
<thead>
<tr>
<th>2005 All-cause mortality</th>
<th>Females</th>
<th>269,368</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td>243,324</td>
<td></td>
</tr>
</tbody>
</table>

\[
\frac{269,368}{269,368 + 243,324} = 0.525
\]

• Interpretation: 52.5% of all-cause deaths in the US for 2005 occurred in females.
Rate

• Rate of consultation for knee surgery
  – A group of 742 people with knee pain
  – Followed up for 3 years after completing a survey
  – During the follow-up period, 202 consultations for knee surgery were recorded among the 742 subjects

\[
\frac{202}{742 \times 3} = 0.091 \text{ consultations per person per year}
\]
Rates: Incidence & Prevalence

• An **incidence rate** of a disease is the number of new cases of a disease in a population during a given time period.

• For a given time period, incidence is defined as:

\[
\frac{\text{# of newly-diagnosed cases of disease during period}}{\text{# of individuals at risk during period}}
\]

• Only those free of the disease at time \( t = 0 \) can be included in numerator or denominator.
Incidence Rate

• 13361 members of an at-risk population fill out a survey
  – All are followed for the 12 calendar months of 2007
  – New cases of CVD diagnosed during 2007 are counted to provide a measure of the incidence of CVD in the at-risk population
    \[
    \frac{264}{13097 + 264} = 0.02 \text{ new diagnoses per year}
    \]

• Incidence rate of 2% per year; 20 per 1000 responders per year
Time as Differential Factor

Disease with rapid resolution or death

Disease with prolonged time to resolution or death
Time as Differential Factor

- Person-time at risk: time during which the event was a possibility for an individual member of the population, and for which it would have been counted as an event had it occurred.

- Population-time at risk: sum of person-times at risk for all population members
  - Different from clock time, as the times are occurring simultaneously for many people.
Incidence Rate

- Limitation: an incidence rate of 1 event per 100 person-years (.01 events per person per year) could be obtained in many ways!
  - Follow 100 people for 1 year, observe 1 event
  - Follow 50 people for 2 years, observe 1 event
- Often only accounts for the first event
Incidence and Prevalence

• A **prevalence rate** is a rate that is taken at a snapshot in time (cross-sectional).

• At any given point, the prevalence is defined as

\[
\frac{\text{# with the illness}}{\text{# of individuals at risk}}
\]

• The **prevalence** of a disease includes both new incident cases and survivors with the illness.
Incidence and Prevalence

- Prevalence is equivalent to incidence multiplied by the average duration of the disease.
- The two are equal in diseases with long durations and low incidence.
- Prevalence is greater than incidence if the disease is long-lasting.
- Diseases with high incidence rates may have low prevalence if they are rapidly fatal or quickly cured.
What is the period prevalence during February? 6/20 = .3
What is the point prevalence on February 28? 1/20 = .05
What is the incidence in February? 4/17 = .235
Care with Interpretation

• Chronic disease: low incidence, long duration
• Acute, common disease: high incidence, short duration
• Preventive measures may lower incidence (e.g., vaccination, public health campaigns)
• Clinical interventions may shorten duration or they may decrease mortality which would result in an increase in disease duration
Incidence and Prevalence

- Prevalence rates are generally used to describe the extent of disease in a population (disease burden)
  - Descriptive, demonstrates public health need
- Incidence rates look at the rate at which new cases of disease develop
  - Good for study cause of disease or to look at the order in which events occur
Measurement Error

• To this point, we have assumed that the presence of disease can be measured perfectly.

• However, mismeasurement of outcomes is common in the medical field due to fallible tests and imprecise measurement tools.
## Diagnostic Testing

<table>
<thead>
<tr>
<th>Diagnostic Test Result</th>
<th>True Disease State</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Present (D+)</td>
<td>Absent (D-)</td>
<td></td>
</tr>
<tr>
<td>Positive (T+)</td>
<td>True Positive (TP)</td>
<td>False Positive (FP)</td>
<td></td>
</tr>
<tr>
<td>Negative (T-)</td>
<td>False Negative (FN)</td>
<td>True Negative (TN)</td>
<td></td>
</tr>
</tbody>
</table>
Sensitivity and Specificity

• **Sensitivity** of a diagnostic test is the probability that the test will be positive among people that have the disease.

\[ P(T+ | D+) = TP/(TP + FN) \]

• Sensitivity provides no information about people that do not have the disease.

• **Specificity** is the probability that the test will be negative among people that are free of the disease.

\[ \Pr(T− | D−) = TN/(TN + FP) \]

• Specificity provides no information about people that have the disease.
Positive Diagnosis

Negative Diagnosis

Diseased

Non-Diseased

Diagnosed positive

SN = 24/30 = 0.80

SP = 56/70 = 0.80

Prevalence = 30/100 = 0.30
Diseased
Healthy
Diseased

A perfect diagnostic test has SN = SP = 1

Positive Diagnosis
Negative Diagnosis
A 100% inaccurate diagnostic test has $SN = SP = 0$

- **Positive Diagnosis:** Red dots
- **Negative Diagnosis:** Green dots

- **Healthy:** Red dots
- **Diseased:** Green dots
Sensitivity and Specificity

- **Example**: 100 HIV+ patients are given a new diagnostic test for rapid diagnosis of HIV, and 80 of these patients are correctly identified as HIV+

  What is the sensitivity of this new diagnostic test?

- **Example**: 500 HIV− patients are given a new diagnostic test for rapid diagnosis of HIV, and 50 of these patients are incorrectly specified as HIV+

  What is the specificity of this new diagnostic test? (Hint: How many of these 500 patients are correctly specified as HIV−?)
Positive and Negative Predictive Value

- **Positive predictive value**: probability that a person with a positive diagnosis actually has the disease.
  \[ Pr(D+ | T+) = \frac{TP}{TP + FP} \]
  - If a patient tests positive for the disease, what are the chances they actually have it?

- **Negative predictive value**: probability that a person with a negative test does not have the disease.
  \[ Pr(D− | T−) = \frac{TN}{TN + FN} \]
  - Similarly, if a patient tests negative for the disease, what are the chances they are truly disease free?
Positive Diagnosis
Negative Diagnosis

Diseased
Healthy

Diagnosed

PPV = 0.63
NPV = 0.90

Diseased
Non-Diseased
Positive Diagnosis
Negative Diagnosis
A perfect diagnostic test has $\text{PPV} = \text{NPV} = 1$
A 100% inaccurate diagnostic test has PPV = NPV = 0
PPV and NPV

- **Example**: 50 patients given a new diagnostic test for rapid diagnosis of HIV test positive, and 25 of these patients are actually HIV+.

  What is the PPV of this new diagnostic test?

- **Example**: 200 patients given a new diagnostic test for rapid diagnosis of HIV test negative, but 2 of these patients are actually HIV+.

  What is the NPV of this new diagnostic test? (Hint: How many of these 200 patients testing negative for HIV are truly HIV−?)
Types of Study Designs and the Quality of Evidence
Types of Evidence

• Scientific evidence: “empirical evidence, gathered in accordance to the scientific method, which serves to support or counter a scientific theory or hypothesis”
  – Type I: descriptive, epidemiological
  – Type II: intervention-based
  – Type III: intervention- and context-based
Types of Evidence

• Type I: descriptive, epidemiological
  – Clinic or controlled community setting
  – Example: Smoking causes lung cancer
  – “Something should be done.”

• Type II: intervention-based
  – Socially-intact groups or community wide
  – Example: Intervention reduces smoking rates
  – “This particular intervention should be implemented.”
Types of Evidence

- Type III: intervention- and context-based
  - Socially-intact groups or community wide
  - Example: understanding the political challenges of intervention in particular audience segments
  - “This is how an intervention should be implemented.”
Types of Evidence

Scientific literature in systematic reviews
Scientific literature in one or more journal articles
Public health surveillance data
Program evaluations
Qualitative data
  - Community members
  - Stakeholders
Media/marketing data
Word of mouth
Personal experience

Objective

Subjective
Types of Studies

Epidemiological Studies

Descriptive Studies
- Populations
  - Ecological
- Individuals
  - Case Reports
  - Case Series
  - Cross Sectional

Analytic Studies
- Observational
- Experimental
  - Case Control
  - Cohort
  - RCT

Complexity and Confidence
Cross-Sectional Studies

• Designed to assess the association between exposure and disease
• Selection of study subjects is based on both their exposure and outcome status
• No direction of inquiry
Cross-Sectional Studies

Defined Population

Gather data on Exposure & Disease

Exposed
Diseased

Exposed
No Disease

Not Exposed
Diseased

Not Exposed
No Disease
Cross-Sectional Studies

• Cannot determine \textit{causal} relationships between exposure and outcome
• Cannot determine \textit{temporal} relationship between exposure and outcome
## Analysis of Cross-Sectional Data

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Prevalence of **disease** compared in exposed versus non-exposed groups:

\[
p(D^+ | E^+) = \frac{a}{a + b}
\]

\[
p(D^+ | E^-) = \frac{c}{c + d}
\]
Analysis of Cross-Sectional Data

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Prevalence of exposure compared in diseased versus non-diseased groups:

\[
p(E^+ | D^+) = \frac{a}{a + c}
\]

\[
p(E^+ | D^-) = \frac{b}{b + d}
\]
Case-Control Studies

• Designed to assess the association between disease occurrence and past exposures

• Selection of study subjects is based on their disease status

• Direction of inquiry is backward
Case-Control Studies

- Exposed
- Unexposed

- Defined Population
  - Gather data on Disease

- Diseased
- No Disease

Time

Direction of Inquiry
Case-Control Studies

- Incident versus prevalent cases
- Selection of appropriate controls
- Temporal sequence of exposure and outcome
- Exposure of cases and controls (similar?)
### Analysis of Case-Control Data

<table>
<thead>
<tr>
<th>Disease</th>
<th>+</th>
<th>-</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposure</td>
<td>+</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>c</td>
<td>d</td>
</tr>
<tr>
<td>Total</td>
<td>a+c</td>
<td>b+d</td>
<td></td>
</tr>
</tbody>
</table>

Odds ratio: \( \frac{\text{odds of case exposure}}{\text{odds of control exposure}} \)

\[
OR = \frac{\frac{a}{c}}{\frac{b}{d}} = \frac{ad}{bc}
\]
Cohort Studies

• Designed to assess the association between exposures and disease occurrence
• Selection of study subjects is based on their exposure status
• Direction of inquiry is forward
Cohort Studies

Defined Population

Gather data on Exposure

Exposed

Disease

No Disease

Not Exposed

Disease

No Disease

Direction of Inquiry

Time
Cohort Studies

- Attrition or loss to follow-up
- Time and money!
- Inefficient for very rare outcomes
- Bias
  - Outcome ascertainment
  - Information bias
  - Non-response bias
## Analysis of Cohort Data

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Disease</th>
<th>Total</th>
<th>Person-time of Observation</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>+</td>
<td>a+b</td>
<td>$PTE$</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>c+d</td>
<td>$PTO$</td>
</tr>
<tr>
<td>Total</td>
<td>a+c</td>
<td>b+d</td>
<td></td>
</tr>
</tbody>
</table>

Relative Risk: $\frac{\text{risk of disease in exposed}}{\text{risk of disease in unexposed}}$

$$RR = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}$$
Randomized Controlled Trials

• Designed to test the association between exposures and disease
• Selection of study subjects is based on their assigned exposure status
• Direction of inquiry is forward
Randomized Controlled Trials

Defined Population

Randomize to Exposure

Exposed (Treated)

Disease
No Disease

Not Exposed (Control)

Disease
No Disease

Direction of Inquiry

Time
Randomized Controlled Trials

- Resource-heavy
- Ethical concerns
- Feasibility
- Blinding
Randomization

• Fixed allocation
  – Simple
  – Block
  – Stratified

• Adaptive allocation
  – Baseline adaptive
  – Response adaptive
Other Protections Against Bias

- Blinding
  - Single, double, triple
- Control
  - Placebo
Validity

External Validity  Internal Validity

Design  Implementation

Research Question  Study Plan  Actual Study

Truth in Reality  Truth in the Study  Findings in the Study

Inference #2  Inference #1
Internal Validity

• Degree to which conclusions correctly describe what actually happened in the study
External Validity

• Degree to which conclusions are applicable to target population
• Also referred to as generalizability
  – How well does your study population reflect your reference population?
Inferential Statistics

- **Nuisance variation** occurs when undesired variables affect the outcome.
- Nuisance variation can systematically distort results in a particular direction—referred to as **bias**.
  - Example: All heavier subjects assigned to one weight loss treatment.
- It can **increase the variability** of the outcome being measured.
  - Example: Failing to control for severity of illness or source of admission in a study of hospital quality.
Threats to Valid Inference

Statistical Conclusion Validity

- **Low statistical power** - failing to reject a false hypothesis because of inadequate sample size, irrelevant sources of variation that are not controlled, or the use of inefficient test statistics.

- **Violated assumptions** - test statistics have been derived conditioned on the truth of certain assumptions. If their tenability is questionable, incorrect inferences may result.

- Many methods are based on approximations to a normal distribution or another probability distribution that becomes more accurate as **sample size increases**. Using these methods for small sample sizes may produce unreliable results.
Threats to Valid Inference

• Statistical Conclusion Validity
  • **Reliability** of measures and treatment implementation.
  • Random variation in the **experimental setting** and/or subjects.
  • Inflation of variability may result in not rejecting a false hypothesis.
Threats to Valid Inference

- Internal Validity
  - **Uncontrolled events** - events other than the administration of treatment that occur between the time the treatment is assigned and the time the outcome is measured.
  - **The passing of time** - processes not related to treatment that occur simply as a function of the passage of time that may affect the outcome.
Threats to Valid Inference

• Internal Validity
  • **Instrumentation** - changes in the calibration of a measuring instrument, the use of more than one instrument, shifts in subjective criteria used by observers, etc.
  • **The “John Henry” effect** - compensatory rivalry by subjects receiving less desirable treatments.
  • **The “placebo” effect** - a subject behaves in a manner consistent with his or her expectations.
Threats to Valid Inference

- **External Validity—Generalizability**
  - **Reactive arrangements** - subjects who are aware that they are being observed may behave differently than subjects who are not aware.
  - **Interaction of testing and treatment** - pretests may sensitize subjects to a topic and enhance the effectiveness of a treatment.
Threats to Valid Inference

- **External Validity—Generalizability**
  - **Self-selection** - the results may only generalize to volunteer populations.
  - **Interaction of setting and treatment** - results obtained in a clinical setting may not generalize to the outside world.
Questions?

• Jo Wick
  – jwick@kumc.edu
  – (913) 588-4790
  – Biostatistics.kumc.edu
References

- For a complete list of references used to create this document, please email jwick@kumc.edu.