



UNIVERSITY OF HAWAII
 Office of Biostatistics & Quantitative Health Sciences
JOHN A. BURNS SCHOOL OF MEDICINE

Research Design & Biostatistics

Lecture 1

John J. Chen, Ph.D.
Professor
Office of Biostatistics & Quantitative Health Sciences
UH JABSOM

Surgery Resident Academic Session
October 17, 2018

Lecture Note: <http://biostat.jabsom.hawaii.edu/Education/training.html>

Outline

Lecture 1 (10/17/2018)


- The goal of statistics
- Introduction to descriptive biostatistics
- Some research design and data presentation issues

Lecture 2 (December, 2018)

- Large databases


Lecture 3 (January, 2019)

- Introduction to inferential statistics
- Some commonly used statistical approaches

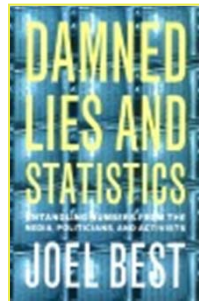
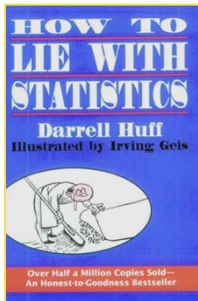


Lecture Objectives

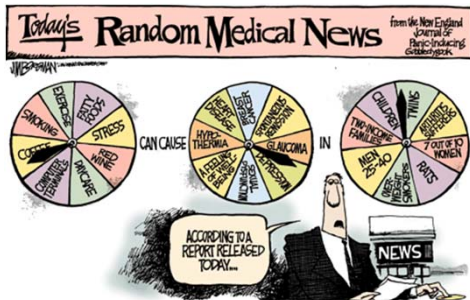
- Clarify the goal of statistics
- Grasp descriptive statistics
- Be familiar with various data presentation approaches
- Introduce key concepts of inferential statistics
- Survey some commonly used statistical approaches
- Understand basic research design principles
- Understand the pros and cons of large databases
- Build a foundation which will facilitate the active participation in clinical and translational research



Lies, Damned Lies, And Statistics



Medical Research, Media, and Public Health



Definition of Statistics

The theory and methodology for research (study) design, and for describing, analyzing, and interpreting information (data) generated from such studies, in which the data is subject to chance variation.

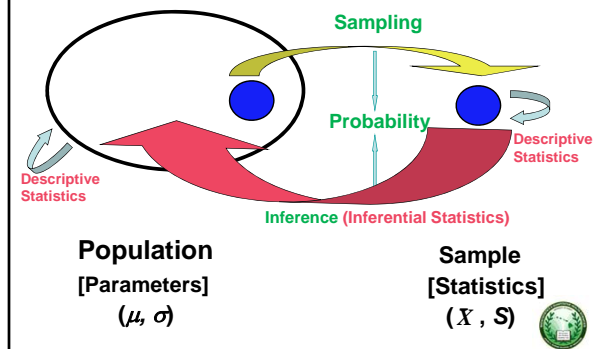


Population & Sample

- **Population:** the set of all subjects of interest having a common observable characteristic. For example, all patients who had heart surgery in US.
- **Sample:** a subset of a population, e.g., all heart surgery patients at QMC in 2017.
- **Parameter:** a summary measure of the population, e.g., the average LOS of the above population.
- **Statistic:** a summary measure of the sample, e.g., the average LOS of the above sample.



The Goal of Statistics



Properties of A “Good” Sample

- Adequate sample size (statistical power)
- Random selection (representative)

Commonly used sampling techniques

1. Simple random sample
2. Stratified sample
3. Systematic sample
4. Cluster sample
5. Convenience sample



Types of Data & Scales of Measurement

1. Qualitative variables - categorical

- Nominal: Categories, names (e.g., gender, eye color)
- Ordinal: Ordered data, intervals are not equal (e.g., satisfaction scores, grades of tumor)

2. Quantitative variables - numerical

- Discrete - no intermediate values (e.g., number of children per family)
- Continuous – intermediate values (e.g., temperature, birth weight)



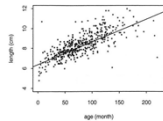
Types of Variables

Notes:

**Dependent (response) versus
Independent (explanatory) variables**

In linear regression analysis:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$



Sources of Data (Types of Studies)

Two major types of investigations:

Surveys versus experiments

Major difference: whether the investigator has control over which subjects enter each study group.

Some examples of survey researches

Prospective (cohort) studies

Retrospective (case-control) studies

Cross-sectional studies

Some examples of experimental studies:

Lab experiments

Clinical trials



Descriptive Statistics

Qualitative data:

- Frequencies
- Percentages

Quantitative data:

- **Measures of central tendency**
Mean, Median, Mode
- **Measures of variability (dispersion)**
Standard deviation, Variance, Range, Interquartile range



Measures of Variability

1. Variance:

$$\text{Sample variance} = s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}$$

2. Standard deviation (SD):

$$\text{Sample SD} = s = \sqrt{s^2}$$

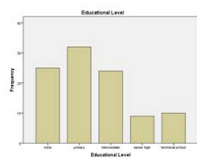
3. Range:

$$\text{Range} = \text{max} - \text{min}$$

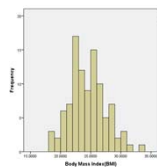


Ways of Displaying Data

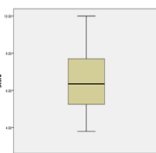
Bar chart



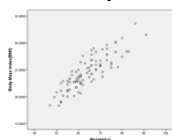
Histogram



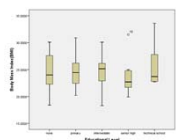
Box Plot



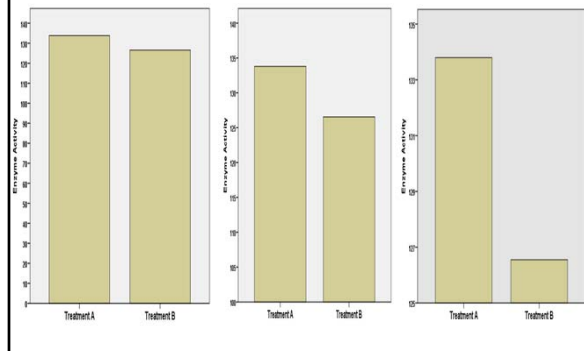
Scatterplot



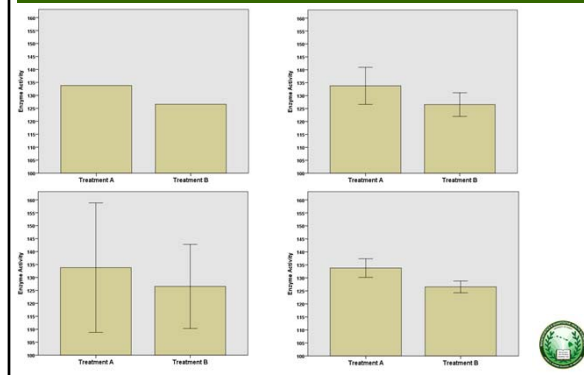
"Box and Bar" Plot



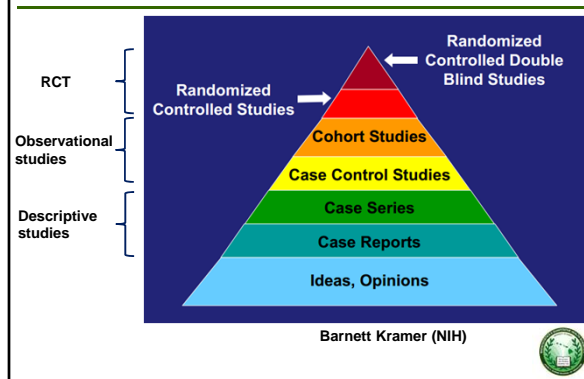
Different Scales



Different Error Bars



Clinical Research & Scientific Evidence



Tan et al. Long-term Survival Following Partial vs Radical Nephrectomy Among Older Patients With Early-Stage Kidney Cancer. JAMA 2012; 307:1629-1635.

ABSTRACT

OBJECTIVE: To compare long-term survival between partial and radical nephrectomy in older patients with early-stage kidney cancer.

DESIGN: Retrospective cohort study.

SETTING: National Cancer Institute Surveillance, Epidemiology, and End Results (SEER) database.

PARTICIPANTS: 1,000 patients with early-stage kidney cancer who underwent partial or radical nephrectomy between 1975 and 1999.

MEASUREMENTS AND MAIN RESULTS: The median age at diagnosis was 68 years. The median follow-up was 10 years. The 5-year survival rates were 50% for partial nephrectomy and 45% for radical nephrectomy. The 10-year survival rates were 35% for partial nephrectomy and 30% for radical nephrectomy. The 15-year survival rates were 25% for partial nephrectomy and 20% for radical nephrectomy. The 20-year survival rates were 15% for partial nephrectomy and 10% for radical nephrectomy. The 25-year survival rates were 10% for partial nephrectomy and 5% for radical nephrectomy. The 30-year survival rates were 5% for partial nephrectomy and 0% for radical nephrectomy. The 35-year survival rates were 0% for partial nephrectomy and 0% for radical nephrectomy. The 40-year survival rates were 0% for partial nephrectomy and 0% for radical nephrectomy. The 45-year survival rates were 0% for partial nephrectomy and 0% for radical nephrectomy. The 50-year survival rates were 0% for partial nephrectomy and 0% for radical nephrectomy. The 55-year survival rates were 0% for partial nephrectomy and 0% for radical nephrectomy. The 60-year survival rates were 0% for partial nephrectomy and 0% for radical nephrectomy. The 65-year survival rates were 0% for partial nephrectomy and 0% for radical nephrectomy. The 70-year survival rates were 0% for partial nephrectomy and 0% for radical nephrectomy. The 75-year survival rates were 0% for partial nephrectomy and 0% for radical nephrectomy. The 80-year survival rates were 0% for partial nephrectomy and 0% for radical nephrectomy. The 85-year survival rates were 0% for partial nephrectomy and 0% for radical nephrectomy. The 90-year survival rates were 0% for partial nephrectomy and 0% for radical nephrectomy. The 95-year survival rates were 0% for partial nephrectomy and 0% for radical nephrectomy. The 100-year survival rates were 0% for partial nephrectomy and 0% for radical nephrectomy.

CONCLUSIONS: Long-term survival was significantly better for partial nephrectomy than for radical nephrectomy in older patients with early-stage kidney cancer.

KEY WORDS: kidney cancer, partial nephrectomy, radical nephrectomy, survival.

ABBREVIATIONS: SEER, Surveillance, Epidemiology, and End Results; OS, overall survival; RFS, recurrence-free survival; CSS, cause-specific survival; QoL, quality of life.

INTRODUCTION: Kidney cancer is the 14th most common cancer in the United States. The incidence of kidney cancer has increased steadily over the past several decades. The most common type of kidney cancer is renal cell carcinoma (RCC). RCC is a heterogeneous disease with several subtypes. The most common subtype is clear cell RCC. The other subtypes are papillary RCC, chromophobe RCC, and collecting duct RCC. The prognosis for RCC depends on the stage of the disease at diagnosis. The most common stage at diagnosis is localized disease. The 5-year survival rate for localized RCC is 50%. The 10-year survival rate for localized RCC is 35%. The 15-year survival rate for localized RCC is 25%. The 20-year survival rate for localized RCC is 15%. The 25-year survival rate for localized RCC is 10%. The 30-year survival rate for localized RCC is 5%. The 35-year survival rate for localized RCC is 0%. The 40-year survival rate for localized RCC is 0%. The 45-year survival rate for localized RCC is 0%. The 50-year survival rate for localized RCC is 0%. The 55-year survival rate for localized RCC is 0%. The 60-year survival rate for localized RCC is 0%. The 65-year survival rate for localized RCC is 0%. The 70-year survival rate for localized RCC is 0%. The 75-year survival rate for localized RCC is 0%. The 80-year survival rate for localized RCC is 0%. The 85-year survival rate for localized RCC is 0%. The 90-year survival rate for localized RCC is 0%. The 95-year survival rate for localized RCC is 0%. The 100-year survival rate for localized RCC is 0%.

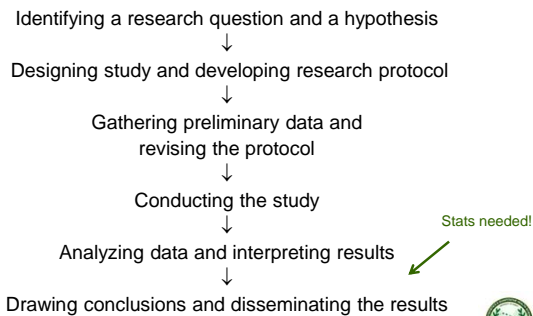
CONCLUSIONS: Long-term survival was significantly better for partial nephrectomy than for radical nephrectomy in older patients with early-stage kidney cancer.

KEY WORDS: kidney cancer, partial nephrectomy, radical nephrectomy, survival.

ABBREVIATIONS: SEER, Surveillance, Epidemiology, and End Results; OS, overall survival; RFS, recurrence-free survival; CSS, cause-specific survival; QoL, quality of life.

INTRODUCTION: Kidney cancer is the 14th most common cancer in the United States. The incidence of kidney cancer has increased steadily over the past several decades. The most common type of kidney cancer is renal cell carcinoma (RCC). RCC is a heterogeneous disease with several subtypes. The most common subtype is clear cell RCC. The other subtypes are papillary RCC, chromophobe RCC, and collecting duct RCC. The prognosis for RCC depends on the stage of the disease at diagnosis. The most common stage at diagnosis is localized disease. The 5-year survival rate for localized RCC is 50%. The 10-year survival rate for localized RCC is 35%. The 15-year survival rate for localized RCC is 25%. The 20-year survival rate for localized RCC is 15%. The 25-year survival rate for localized RCC is 10%. The 30-year survival rate for localized RCC is 5%. The 35-year survival rate for localized RCC is 0%. The 40-year survival rate for localized RCC is 0%. The 45-year survival rate for localized RCC is 0%. The 50-year survival rate for localized RCC is 0%. The 55-year survival rate for localized RCC is 0%. The 60-year survival rate for localized RCC is 0%. The 65-year survival rate for localized RCC is 0%. The 70-year survival rate for localized RCC is 0%. The 75-year survival rate for localized RCC is 0%. The 80-year survival rate for localized RCC is 0%. The 85-year survival rate for localized RCC is 0%. The 90-year survival rate for localized RCC is 0%. The 95-year survival rate for localized RCC is 0%. The 100-year survival rate for localized RCC is 0%.

Biomedical Research Process




Basic Principles of Experimental Design

- Replications
- Randomization
- Blocking (stratification)
- Blinding
- Factorial experiments


Handling A Confounding Variable (Z)

- If you can, fix a variable.
- If you can't, stratify it.
- If can't fix or stratify a variable, randomize it.


$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \epsilon$$

Data & Data Dictionary				
SPSS: Honolulu Heart Study (partial data)				
				
ID	Educational Level	Variable	Education	
1	1	2	Description/Label	
2	2	1	Education Level	
3	3	1	Data Type	
4	4	2	Num – Categorical variable	
5	5	2	Length	
6	6	4	Allowable Values	
7	7	1	1=none	
8	8	3	2=primary	
9	9	5	3=intermediate	
10	10	2	4=senior high	
11	11	4	5=technical school	
12	12	1	6=university or above	
13	13	1	Notes	
14	14	2	Required field. No missing allowed.	
15	15	3		
16	16	4		
17	17	3		
18	18	5		
19	19	1		
20	20	4		
21	21	1		
22	22	3		

Warning Signs



Data generation



Data consumption

Master of Science in Clinical and Translational Research



The Clinical and Translational Research (CTR) graduate program will prepare graduates with skills for successful careers in clinical and translational research and research support.

Clinical Research (CR) Track

Develop knowledge and skills to investigate clinical research topics through coursework and research projects focused on research design, methodologies, quantitative methods, scientific writing, ethical issues, and the capacity in obtaining research funding.

Career

Research, research support, data analyst positions at:

- Academia
- Hospitals
- Government agencies
- Healthcare organizations
- Pharmaceutical companies

How to Apply

Visit <http://manoa.hawaii.edu/graduate/content/clinical-research> to either fill out an application or download a PDF form

Application Deadline: May 30

Quantitative Health Sciences (QHS) Track

Courses and research projects focus on biostatistical and bioinformatic methods development and application to improve population and individual health. Students will acquire big data skills and master the scientific principles and methodologies that underlie basic science, clinical, and translation research.

Program Curricula

~2-year 34 total credit hours graduate program

- Plan A (Thesis): 24 credits of didactic courses
- Plan B (Capstone Project): 28 credits of didactic courses

For more information

Phone: (808) 692-1840
Email: GradCTR@hawaii.edu
Web: <http://msctr.jabsom.hawaii.edu>

MSCTR Curriculum

- BIOM 640 Introduction to Clinical Research (3 credits)
- BIOM 641 Legal & Regulatory Issues and Bioethics (2 credits; cross-listed with CMB626)
- BIOM 644 Translational Research Methods (2 credits)
- BIOM 645 Clinical Protocol Development (3 credits)
- BIOM 654 Medical Genetics (2 credits)
- QHS 601 Biomedical Statistics I (3 credits; cross-listed with TRMD 655)
- QHS 602 Biomedical Statistics II (3 credits)
- QHS 610 Bioinformatics I (3 credits; cross-listed with TRMD 653)
- QHS 611 Bioinformatics II (3 credits)
- QHS 620 Introduction to Clinical Trials (2 credits)
- QHS 621 Design and Analysis of Clinical Trials (2 credits)
- QHS 650 Secondary Data Analysis (2 credits)
- QHS 651 Secondary Data Analysis Practicum (2 credits)
- QHS 675 Biostatistical Consulting (2 credits)
- QHS 676 Biostatistical Consulting Practicum (1 - 2 credits)

MSCTR Graduate Program Website: msctr.jabsom.hawaii.edu



Collaboration with A Biostatistician

1. Early and often
2. Start the discussion when you have the initial idea
3. It is an iterative process
4. A collaborative effort: equal and fair
5. Ask questions so you can discuss about the general statistical approach without the statistician

JABSOM Office of Biostatistics & QHS

<http://biostat.jabsom.hawaii.edu>



NIH U54MD007601



NIH U54MD007584



Lecture Objectives

- Clarify the goal of statistics
- Grasp descriptive statistics
- Be familiar with various data presentation approaches
- Introduce key concepts of inferential statistics
- Survey some commonly used statistical approaches
- Understand basic research design principles
- Understand the pros and cons of large databases
- Build a foundation which will facilitate the active participation in clinical and translational research