STUDY DESIGN
CASE SERIES AND CROSS-SECTIONAL

Daniel E. Ford, MD, MPH
Vice Dean for Clinical Investigation
Johns Hopkins School of Medicine
Introduction to Clinical Research

STUDY DESIGN

- **Provides "differential diagnosis" of a study's strengths and weakness**
- **Determines confidence in results of study**
- **Facilitates critical appraisal of the medical literature**
- **Linked to research question**

STUDY DESIGNS AND CORRESPONDING QUESTIONS

• Ecologic	• What explains differences between groups?
• Case Series	• How common is this finding in a disease?
• Cross-sectional	• How common is this disease or condition?
• Case-control	• What factors are associated with having a disease?
• Prospective	• How many people will get the disease? What factors predict development?
• Randomized trial	• Does the outcome change if we change something?

2 x 2 TABLE

		Disease Status (Outcome) Yes	No	Total
Risk Factor (Exposure)	Yes	A	B	A + B
	No	C	D	C + D
Total		A + C	B + D	

Relative Risk = $\frac{A/A + B}{C/C + D}$ Odds Ratio = $\frac{AD}{BC}$

STUDY DESIGN DEFINITIONS

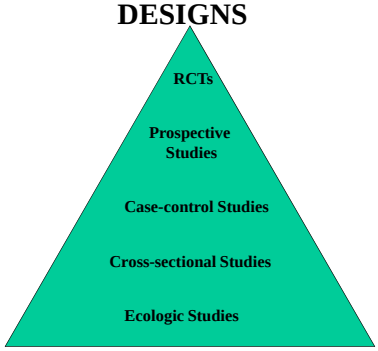
- **Based on sampling strategy, i.e., how we choose who gets into the study**
- **Sampling with regard to disease: cross-sectional and case-control studies**
- **Sampling with regard to exposure or treatment: prospective studies**

CRITERIA FOR CAUSAL INFERENCE

- **Experimental evidence**
- **Temporality**
- **Strength of the association**
- **Dose-response relationship**
- **Consistency in different populations**
- **Specificity: exposure leads to only 1 disease**
- **Biologic plausibility**
- **Coherence**
- **Analogy**

Not all study designs are created equal!

HIERARCHY OF STUDY DESIGNS



STUDY DESIGN EXAMPLE

- **Does higher dose of dialysis (Kt/v) result in lower mortality in hemodialysis patients?**

ECOLOGIC STUDIES

- **Sometimes called correlational studies**
- **Compares outcomes between groups, <u>not</u> individuals**
- **Useful to examine trends over time or to explain differences between groups**

2 x 2 Table

		Outcome: Yes	Outcome: No	Total
Exposure	Yes	a	b	a + b
	No	c	d	c+ d
Total		a + c	b + d	N

- **Letters represent group rates or means, not individuals**
- **Exposure can be assessed before or after outcome**

Kt/V AND MORTALITY IN 100 DIALYSIS UNITS

Mean Clinic Kt/V	Mortality: High	Mortality: Low
Acceptable	20	40
Low	20	20
	40	60

VITAL STATISTICS

- **Common data source for ecologic studies**
- **Describes disease patterns in entire geographic or political populations**
- **Routinely collect information from birth and death certificates; allow comparisons between countries over time**
- **Comparison by age, race, sex, geographic areas and time period**

VITAL STATISTICS ADVANTAGES

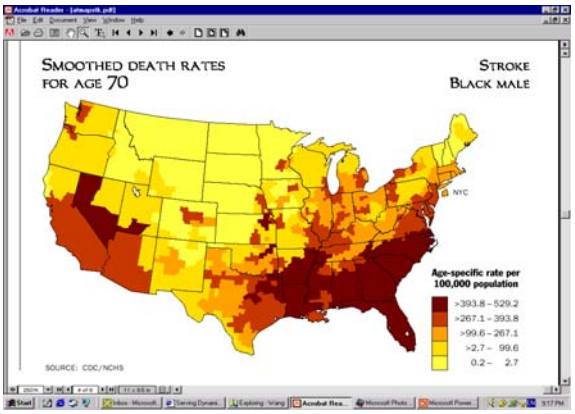
- **Inexpensive**
- **Representative of large groups and large geographic areas**
- **Available over long periods of time**
- **Uniform coding rules**

VITAL STATISTICS DISADVANTAGES

- **Group “ecologic” data -- not individual**
- **Uncertain accuracy of diagnoses**
- **Changes in ICD codes**
- **Variability in coding practices**
- **Limited to available data**
- **Mortality may not reflect incidence**

ECOLOGIC STUDIES DISADVANTAGES

- **Subject to ecologic fallacy**
  - logical fallacy in the interpretation of statistical data in an ecological study, whereby inferences about the nature of individuals are based solely upon aggregate statistics collected for the group to which those individuals belong
- **Lead to unusual conclusions if not testing biologically plausible hypotheses**
- **Usually done early in the investigation of a research question when cohort studies or clinical trials not available**



STUDY DESIGNS AND CORRESPONDING QUESTIONS

• **Ecologic**	• **What explains differences between groups?**
• **Case Series**	• **How common is this finding in a disease?**
• **Cross-sectional**	• **How common is this disease or condition?**
• **Case-control**	• **What factors are associated with having a disease?**
• **Prospective**	• **How many people will get the disease? What factors predict development?**

CASE REPORTS

- **Make observations about medical phenomena in an individual patient**
- **Simple description of clinical data without comparison group**
- **Observations should be comprehensive and adequately detailed**

Kt/V AND MORTALITY CASE REPORT

- **55 year old man has been on dialysis for 35 years**
- **On home dialysis daily during that time**
- **No evidence of hypertension, cardiovascular disease, LVH**

CASE REPORTS ADVANTAGES

- **Easy and inexpensive to do in hospital**
- **Provides information on new disease or new therapy**
- **Useful in conveying “clinical experience”**
- **Helpful in hypothesis formation**

CASE REPORTS DISADVANTAGES

- **Biased selection of subjects so that conclusions are difficult to generalize**
- **Were the findings a chance happening or characteristic of the disease?**
- **Is the "exposure" really higher than a comparison group?**

CASE REPORTS EXAMPLES

- **Asbestos and mesothelioma**
- **Pneumocystis pneumonia**
- **Legionnaire's Disease**

DECIDING TO PUBLISH

- **What observations have been made prior to this report?**
- **What new phenomenon is illustrated?**
- **What further studies should be done?**

CASE SERIES

- **Group of patients with a disease or outcome**
- **Usually consecutive series**
- **Detailed observations**
- **No comparison group – difficult to address etiologic questions**

2 x 2 TABLE
Case Series

		Outcome Yes	No	Total
Exposure	Yes	a		a
	No	c		c
Total		a +c		N

DISTRIBUTION OF Kt/V IN 100 PATIENTS WHO DIED DURING THE FIRST YEAR OF DIALYSIS

N = 100

Low Kt/V	70%
Acceptable Kt/V	30%

- review records on 100 patients who died

CASE SERIES OBSERVATIONS

- **Should have clear definitions of the phenomena being studied**
- **Same definitions should be applied equally to all individuals in the series**
- **All observations should be reliable and reproducible (consider blinding assessors)**

CASE SERIES PRESENTATION OF FINDINGS

- **Proportions of the study populations with the outcome with confidence intervals**
- **Means, standard errors for continuous variables**
- **Consider reporting data for important subgroups separately?**

CASE SERIES ADVANTAGES

- **Informs patients and physicians about natural history and prognostic factors**
- **Easy and inexpensive to do in hospital settings**
- **Helpful in hypothesis formation**
- **Can help answer the question of why this outcome occurred**

CASE SERIES LIMITATIONS

- **Cases may not be representative**
- **Outcome may be a chance finding, not characteristic of disease**
- **Begs the question "Compared to what?"**

STUDY DESIGNS AND CORRESPONDING QUESTIONS

• **Ecologic**	• **What explains differences between groups?**
• **Case Series**	• **How common is this finding in a disease?**
• **Cross-sectional**	• **How common is this disease or condition?**
• **Case-control**	• **What factors are associated with having a disease?**
• **Prospective**	• **How many people will get the disease? What factors predict development?**

CROSS-SECTIONAL STUDIES

- **Make observations concerning the prevalence and characteristics of a disease in a well-defined population during a defined period of time (period prevalence)**
- **Estimate prevalence**
- **Examine characteristics associated with condition or disease by comparing cases to noncases**

2 x 2 TABLE
Cross-sectional Study

		Disease: Yes	No	Total
Exposure	Yes	a	b	a + b
	No	c	d	c + d
Total		a + c	b+ d	N

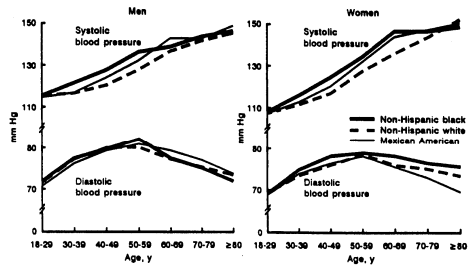
- Draw a 1% random sample of all hemodialysis patients dialyzed in 1996
- Assess Kt/V in all patients and their vital status at the end of 1996

PREVALENCE OF LOW Kt/V AND MORTALITY

	Dead	Alive	Total
Low Kt/V	400	1,000	1,400
High Kt/V	350	1,250	1,600
	750	2,250	3,000

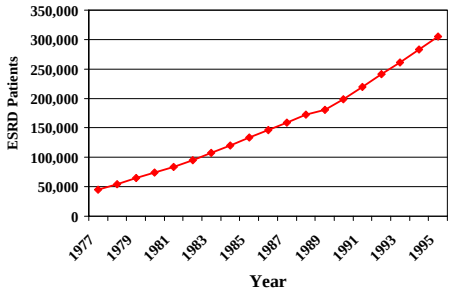
Odds Ratio = $\frac{(400)\,(1{,}250)}{(350)\,(1{,}000)} = 1.4$

MEAN BLOOD PRESSURE BY AGE AND GENDER, U.S., 1991



Burt, *Hypertension*, 1995

Number of Medicare ESRD Patients on Dialysis in the United States



Physician Visits Per Person in the Last Six Months of Life: 2002-2003

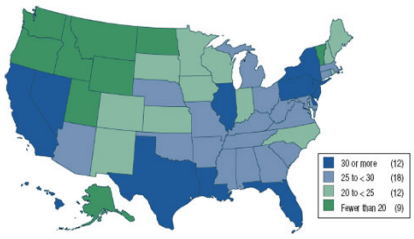


Table 2. Odds Ratios of Elevated C-Reactive Protein Level by Recent History of Major Depression in Men and Women*

	Odds Ratio (95% Confidence Interval)	*P* Value
Men		
No depression	1.0 (Reference)	
Severe depression <1 year	3.00 (1.39-6.48)	.006
Severe depression >1 year	1.95 (0.81-4.69)	.14
Women		
No depression	1.0 (Reference)	
Severe depression <1 year	0.76 (0.44-1.33)	.33
Severe depression >1 year	1.39 (0.80-2.42)	.24

*Adjusted for age, African American race, body mass index, total cholesterol, log triglycerides, diabetes, systolic blood pressure, smoking status, alcohol use, estrogen use in women, aspirin use, ibuprofen use, and self-reported health status. *P* value for trend = .01 (men) and .70 (women); *P* value for interaction (<1 year) = .01.

Table 3. Odds Ratios for Elevated C-Reactive Protein Level by Recurrent* Major Depression in Men and Women†

	Odds Ratio (95% Confidence Interval)	*P* Value
Men		
No depression	1.0 (Reference)	
Single episode	1.13 (0.25-5.18)	.87
Recurrent	3.55 (1.55-8.14)	.003
Women		
No depression	1.00 (Reference)	
Single episode	0.73 (0.36-1.47)	.37
Recurrent	0.72 (0.45-1.14)	.16

*Recurrent depression was defined as having 2 or more episodes of severe major depression.
†Adjusted for age, African American race, body mass index, total cholesterol, log triglycerides, diabetes, systolic blood pressure, smoking status, alcohol use, estrogen use in women, aspirin use, ibuprofen use, and self-reported health status. *P* value for interaction (recurring) = .002.

SAMPLING

- **Process of obtaining a sample of a population for study**
- **Goal should be sample representative of all individuals you are stating the results apply**
- **Variety of methods available**

CROSS-SECTIONAL STUDIES SAMPLING THE POPULATION

- **Derive a sampling “frame”**
- **Choose a sampling strategy**
- **Maximize response rate**

CROSS-SECTIONAL STUDIES TYPES OF SAMPLING

- ***Simple random*--each individual has the same probability of being chosen**
- ***Stratified random*– first create strata and then select randomly within strata. If most variance is between strata, gives lower sampling variance**
- ***Systematic*— ex., select every 4th person, used commonly in clinical research, akin to stratified random sample if list is ordered**
- ***Cluster***

RESPONSE RATES AND SAMPLING

- **Sample size of 500**
  - **5% of 10,000=500**
  - **75% of 666=500**
- **Which study provides the most valid causal inference?**
- **Are persons who do not respond (can't be found or say no) likely to be different than those who do respond?**

NONRESPONSE IN SAMPLING CROSS SECTIONAL STUDIES

- **Minimize non-response**
  - **smaller sample size allows more intensive recruitment**
  - **collect data on non-responders, if possible**
  - **intensively recruit a sub-sample of non-responders**

CROSS-SECTIONAL STUDIES ADVANTAGES

- **Inexpensive for common diseases**
- **Should be able to get a better response rate than other study designs**
- **Relatively short study duration**
- **Can assess gradation of association (higher exposure higher outcome)**

CROSS-SECTIONAL STUDIES DISADVANTAGES

- **Unsuitable for rare or short duration diseases (prevalence = incidence x duration)**
- **Disease process may alter exposure – reverse causality**
- **No data on temporal relationship between risk factors and disease development**
- **If data collected for other reasons (secondary analysis of existing data) may not have high quality data, particularly to assess confounding**

Defining Cross-Sectional Studies

- **How short is the assessment period?**
  - **Symptom questionnaire and then physical exam**
- **Cases accumulated over long period of time**
- **Time trends of multiple cross-sectional studies (smoking rates in population over time)**

STUDY DESIGNS AND CORRESPONDING QUESTIONS

• **Ecologic**	• **What explains differences between groups?**
• **Case Series**	• **How common is this finding in a disease?**
• **Cross-sectional**	• **How common is this disease or condition?**
• **Case-control**	• **What factors are associated with having a disease?**
• **Prospective**	• **How many people will get the disease? What factors predict development?**