Research and Statistics

April 2020

Karen E. Schetzina, MD, MPH, FAAP

Research and Statistics

Overview

- Levels of evidence and study design
- Therapy articles
- Bias
- Statistics
- Articles about diagnosis

Clinical Questions that arise

- Clinical findings/manifestations
- Etiology
- Differential diagnosis
- Diagnostic tests
- Prognosis
- Therapy
- Prevention
- Experience and meaning
- Improvement

Quality of Evidence

Levels of Evidence Ratings

- Centre for Evidence-Based Medicine, Oxford
  - Expert Opinion: LOE=5
  - Case Series: LOE=4
  - Case Control: LOE=3b
  - RCT: LOE=1b
  - SR with homogeneity: LOE=1a

Case report or series

- Describe patients or a series of patients with an interesting presentation of disease

Ecological Studies

- Studies correlating two or more characteristics of populations – unit of analysis is the group, not the individual
  - *Per capita consumption of cigarettes correlates with national mortality from lung cancer in countries throughout the world.*

Ecological Studies

- Advantages
  - Quick, easy, and cost-effective
- Disadvantages
  - Cannot be extrapolated to individuals
  - Correlations can be misleading

Cross-Sectional Study

- Measures exposure to suspected risk factors and the presence of a disease or condition in people in a population at the same time
  - *In a school survey, children who had disciplinary problems were found to spend more hours in front of the television at home.*

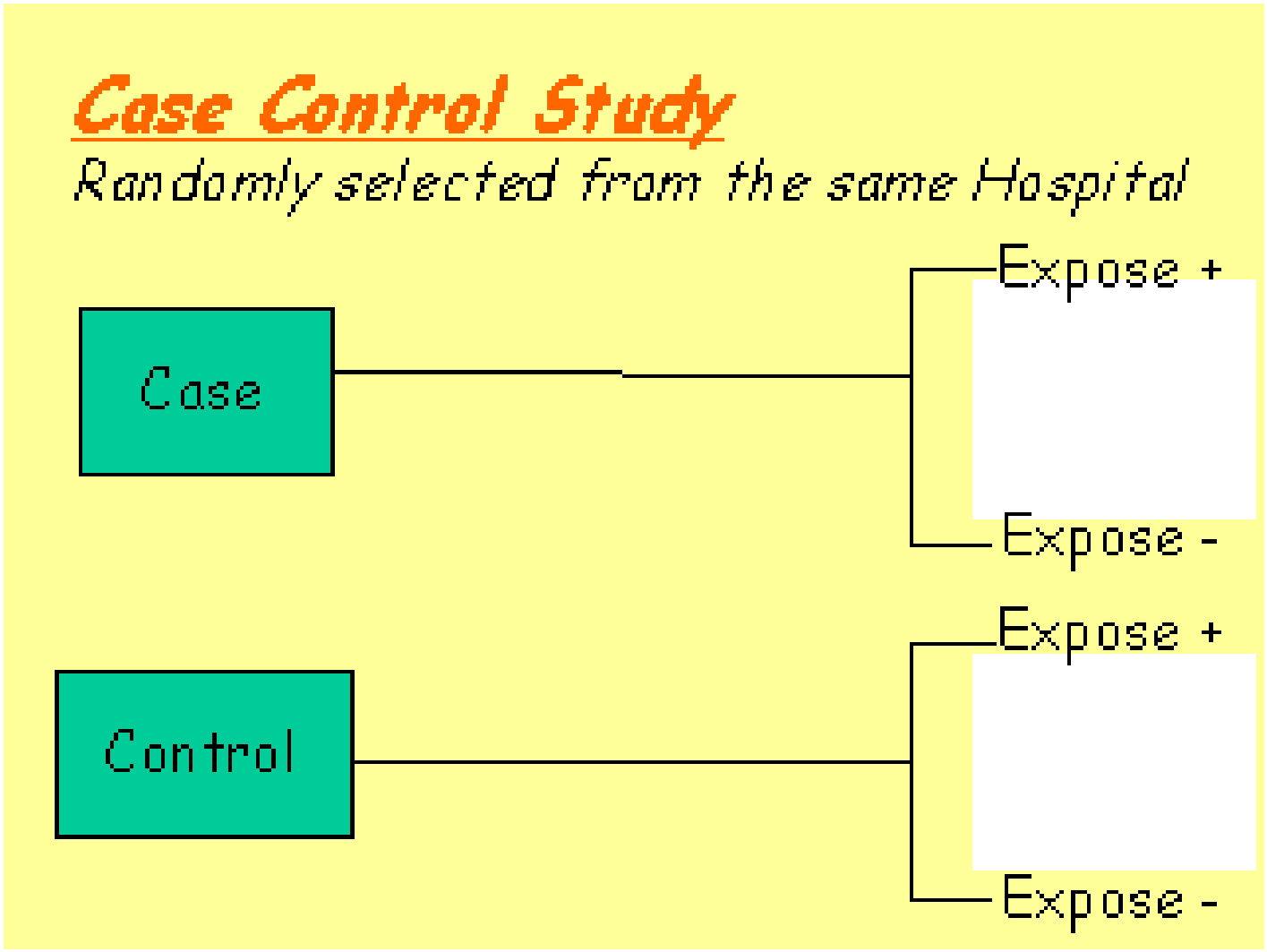
Cross-Sectional Study

- Advantages
  - Relatively quick and inexpensive
- Disadvantages
  - Difficult to assess temporal associations between risk factor and disease (tough to differentiate cause and effect)
  - See over-representation of diseases of long duration (because measuring prevalence)

Case Control Study

- People with and without a disease or condition of interest are identified and their prior exposure to suspected risk factors is measured.
  - *The proportion of smokers among lung cancer patients admitted to a hospital was found to be much higher than the proportion of smokers among patients admitted for other reasons*

Source:
Bambang Sutrisna, MD, MHS, DrPH, University of Indonedia,
http://www.pitt.edu/~super1/index.htm



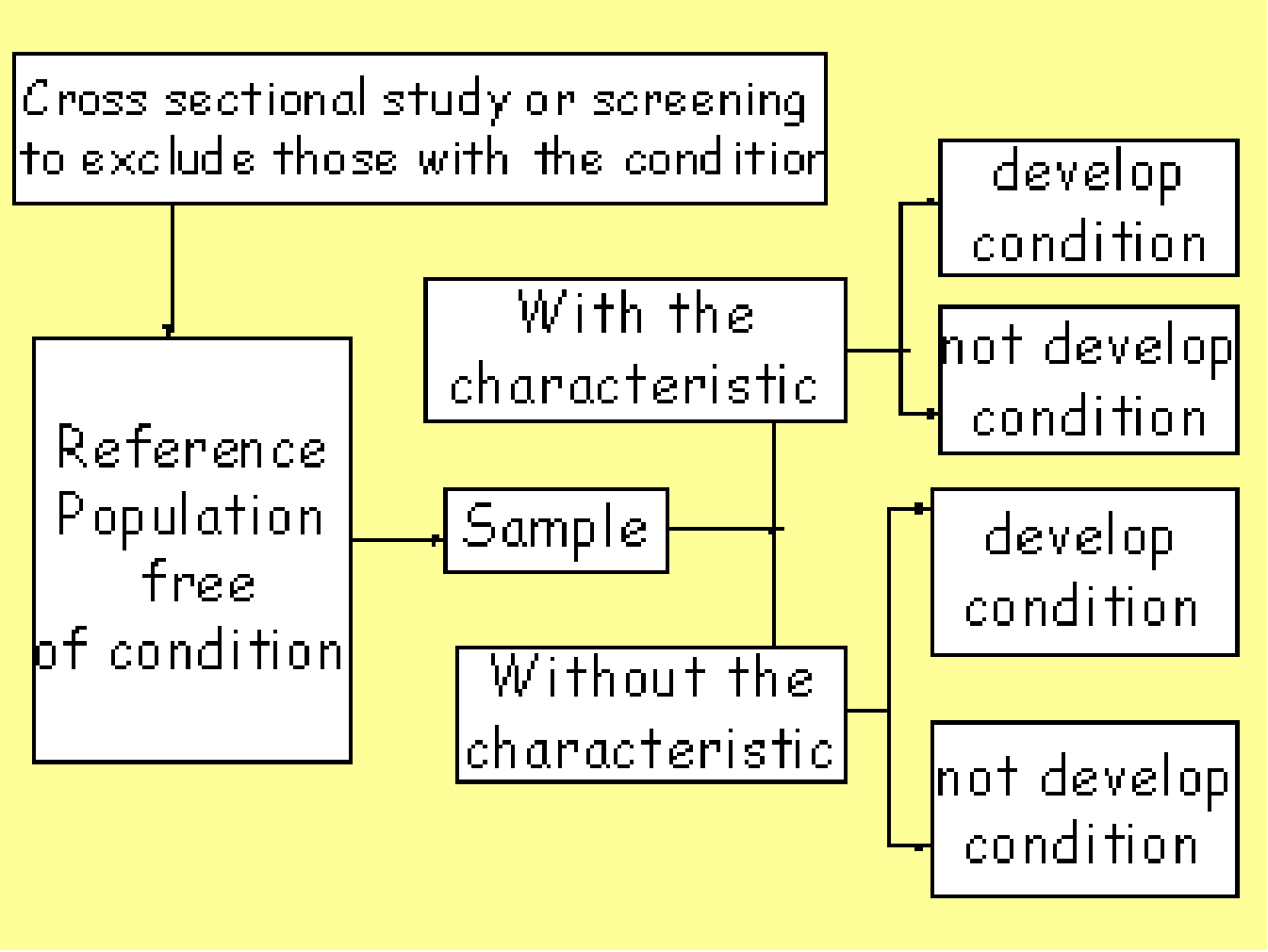
Case Control Study

- Advantages
  - Relatively quick and inexpensive
  - Good for rare diseases
- Disadvantages
  - Information on suspected risk factors may be incomplete
  - Potential problem of recall bias
  - Selection of an appropriate control group can be difficult

Cohort Study

- People with and without exposure to a suspected risk factor are identified and followed to determine whether they develop the disease(s) or condition(s) of interest.
  - *In the Framingham Heart Study, a cohort of disease-free individuals were evaluated for blood pressure levels, cholesterol levels, and other characteristics, and followed for over 40 years for the development of coronary heart disease.*

Cohort Study



Cohort Study

- Advantages
  - Can better establish temporality (because measure the risk factor before the disease is diagnosed)
  - Can get incidence rates
  - Can study many diseases
- Disadvantages
  - Take a long time to complete, require large sample size, and are expensive
  - Loss-to-follow-up bias may occur
  - Inefficient for studying rare diseases

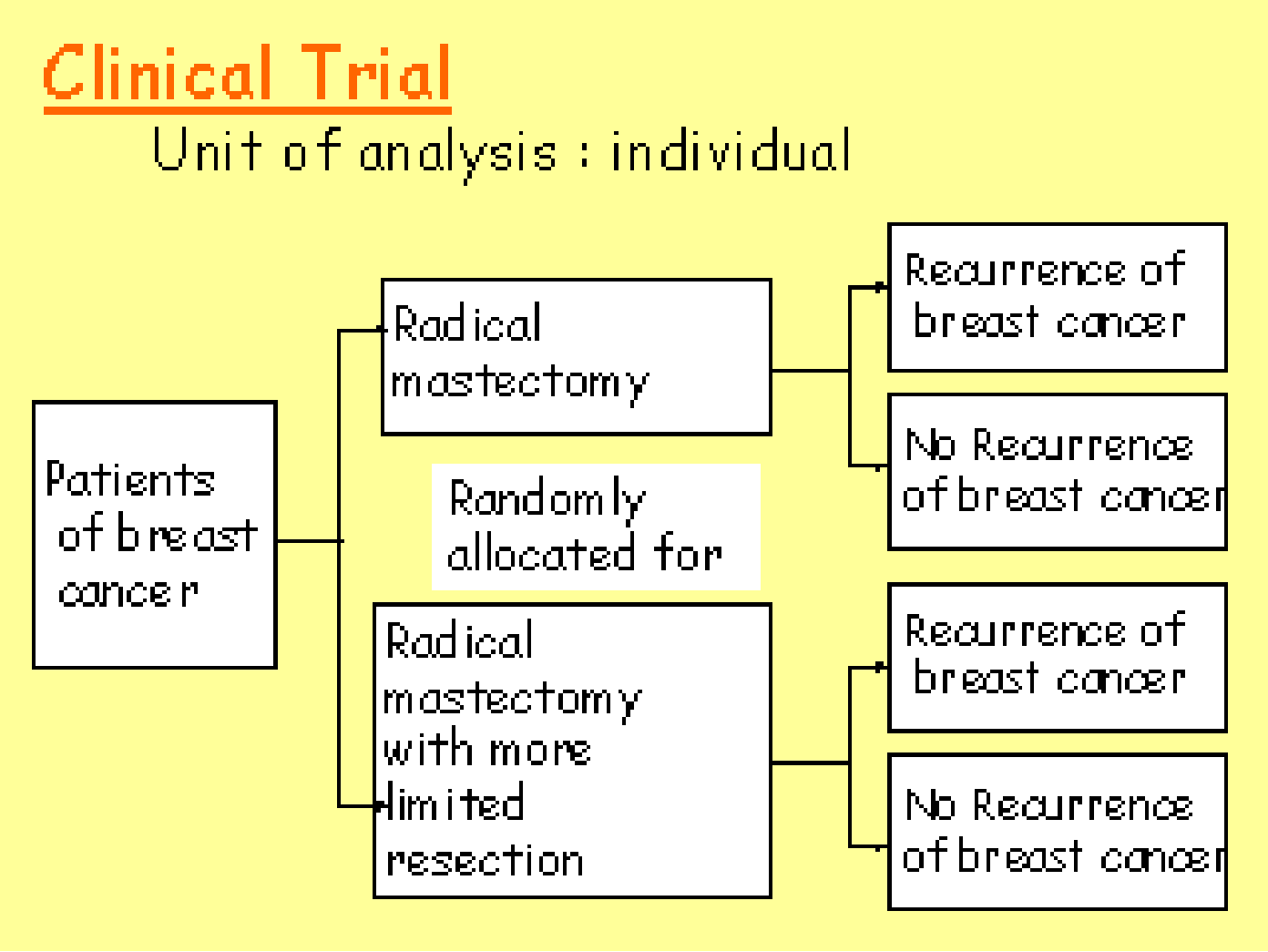
Analysis of Results

Expose	Disease		Total
	+	-	
+	a	b	a+b
-	c	d	c+d
Total	a+c	b+d	a+b+c+d

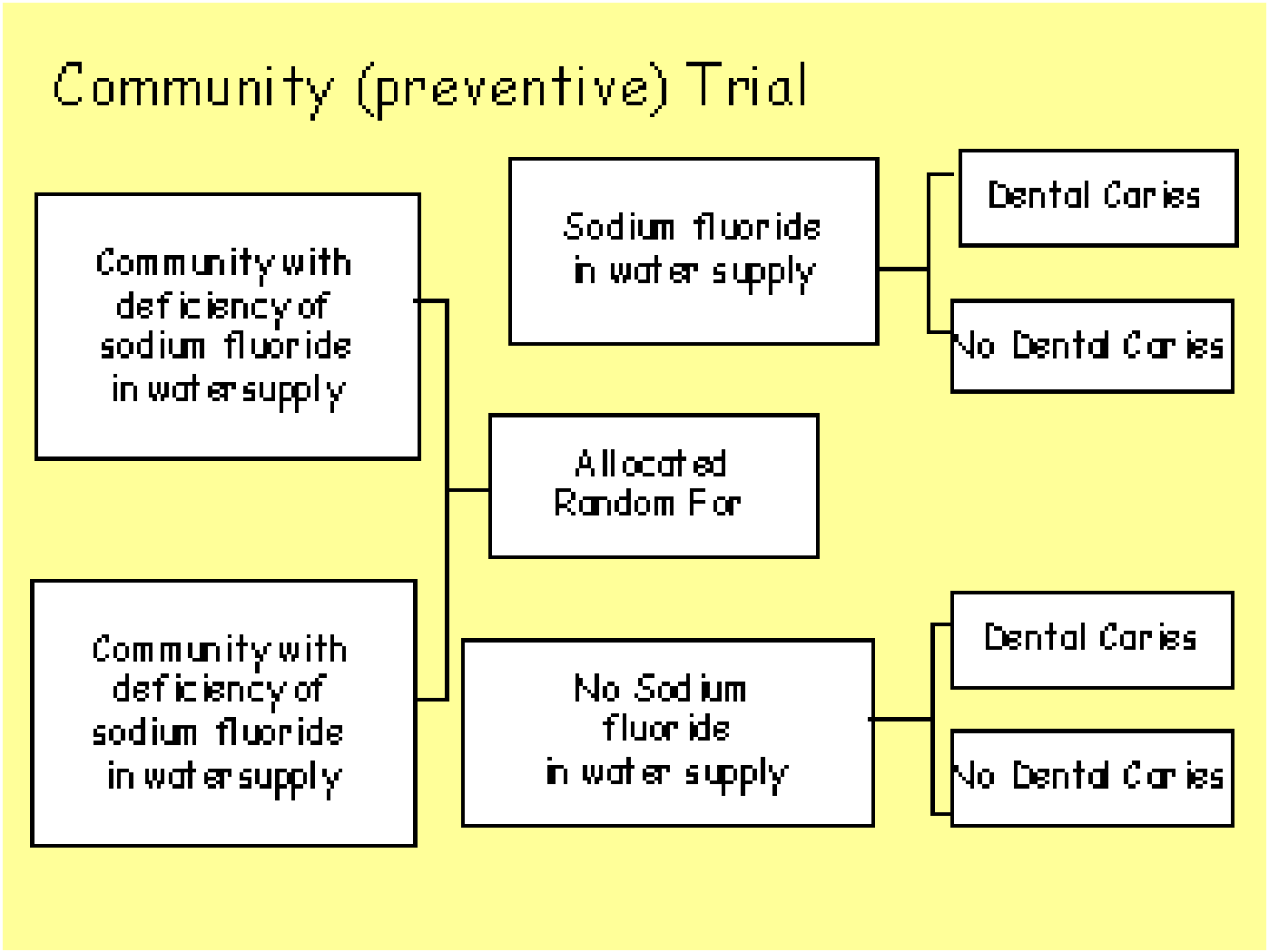
Experimental Study

- Basically a prospective cohort study in which the exposure and the persons to be exposed are determined by the investigator. The randomized controlled trial is the strongest method to determine cause and effect in human subjects.

Source:
Bambang Sutrisna, MD, MHS, DrPH, University of Indonedia,
http://www.pitt.edu/~super1/index.htm



Source:
Bambang Sutrisna, MD, MHS, DrPH, University of Indonedia,
http://www.pitt.edu/~super1/index.htm



Experimental Study

- Advantages
  - Limits bias/confounding
- Disadvantages
  - Expensive
  - May take a long time
  - External generalizability may be limited

Systematic Reviews / Meta-Analysis

Strengths

- Summarize data and findings from multiple trials
- Can be more generalizable if different patient populations were included

Limitations

- Variable data reporting
- Methodology of individual trials
- Heterogeneity of outcomes
- May place too much weight on findings

Reporting Guidelines

Table. Publication reporting systems by study type and website access link		
Type of Study	**Standard**	**Link**
Randomized Clinical Trials	CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials (6).	http://www.equator-network.org/reporting-guidelines/consort/
Observational studies	The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies (7).	http://www.equator-network.org/reporting-guidelines/strobe/
Systematic reviews and Meta-analyses	Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement (8).	http://www.equator-network.org/reporting-guidelines/prisma/
Diagnostic validity studies	Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Standards for Reporting of Diagnostic Accuracy (9).	http://www.equator-network.org/reporting-guidelines/stard/
Case reports	The CARE guidelines: consensus-based clinical case reporting guideline development (10).	http://www.equator-network.org/reporting-guidelines/care/
Statistical analysis	Basic Statistical Reporting for Articles Published in Biomedical Journals: The "Statistical Analyses and Methods in the Published Literature" or The SAMPL Guidelines (11).	http://www.equator-network.org/reporting-guidelines/sampl/

http://www.scielo.org.co/img/revistas/rcog/v65n1/en_v65n1a01t.jpg

Quality Improvement

http://squire-statement.org/index.cfm?fuseaction=Page.ViewPage&PageID=471

Implementation Science:
REAIM Model Domains Adapted from Glasgow, Vogt, & Boles (1999)

Domain	Definitions
Reach	Proportion of the target intervention that participated in the intervention
Effectiveness	Success rate if implemented as in guidelines; defined as positive outcomes minus negative outcomes
Adoption	Proportion of settings, practices, and plans that will adopt this intervention
Implementation	Extent to which the intervention is implemented as intended in the real world
Maintenance	Extent to which a program is sustained over time

2

Definitions

- Hypothesis or research question
- Population – Collection of units from which a sample may be drawn
- Sample – A selected subset of a population
- Variable - “Any quantity that varies. Any attribute, phenomenon, or event that can have different values.”
- Predictor Variable – Risk factor
- Outcome Variable – Disease or condition of interest
- Association or Correlation – Degree to which variables change together

Definitions

- Estimate – Incorporates some degree of error
- Prevalence - “The number of events in a given population at a designated time”
- Incidence – “The number of new events in a defined population within a specified period of time”
- Mean – average
- Variance – A measure of dispersion or variation
- Standard Deviation/Error – Square root of the variance


3

Validity=Truth

- **External Validity** - Are the results of the study generalizable to other populations of interest? Are the results valid for this other population?

- **Internal Validity** - Do the study results represent the truth for the population studied? All studies are flawed to some degree. To reduce the effect of bias and confounding on a study's results, the study must be correctly designed, executed, and analyzed.

4. An advertisement in a medical journal stated that "2000 subjects with sore throats were treated with our new medicine. Within four days, 94% were asymptomatic." the advertisement claims that the medicine was effective. Based on the evidence given, the claim:

A. Is correct
B. May be incorrect because the conclusion is not based on a rate.
C. May be incorrect because of failure to recognize a long-term cohort phenomenon.
D. May be incorrect because no test of statistical significance was used.
E. May be incorrect because no control or comparison group was involved.

Answer: An advertisement in a medical journal stated that "2000 subjects with sore throats were treated with our new medicine. Within four days, 94% were asymptomatic." the advertisement claims that the medicine was effective. Based on the evidence given, the claim:

A. Is correct
B. May be incorrect because the conclusion is not based on a rate.
C. May be incorrect because of failure to recognize a long-term cohort phenomenon.
D. May be incorrect because no test of statistical significance was used.
E. **May be incorrect because no control or comparison group was involved.**

All studies are flawed!

5. The major purpose of random assignment in a clinical trial is to:

A. Help ensure that study subjects are representative of the general population
B. Facilitate double-blinding
C. Facilitate measurement of outcome variables
D. Try to have the study groups comparable on baseline characteristics
E. Reduce selection bias in allocation of treatment

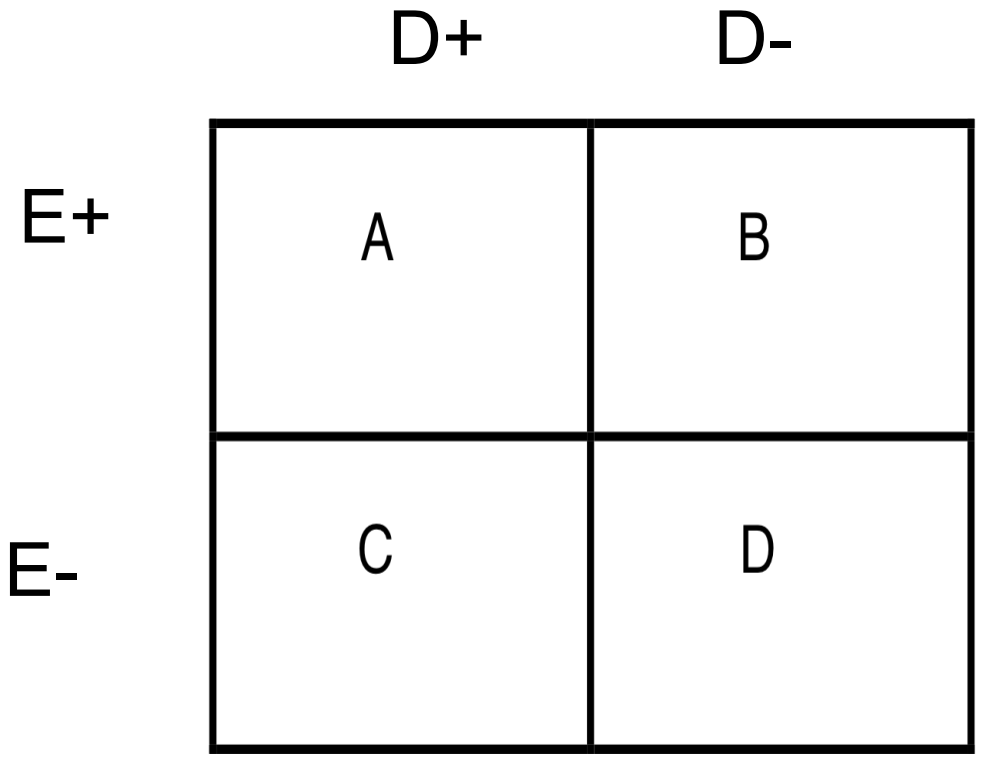
Answer: The major purpose of random assignment in a clinical trial is to:

A. Help ensure that study subjects are representative of the general population
B. Facilitate double-blinding
C. Facilitate measurement of outcome variables
D. Try to have the study groups comparable on baseline characteristics
E. Reduce selection bias in allocation of treatment

Critical Reading – Review

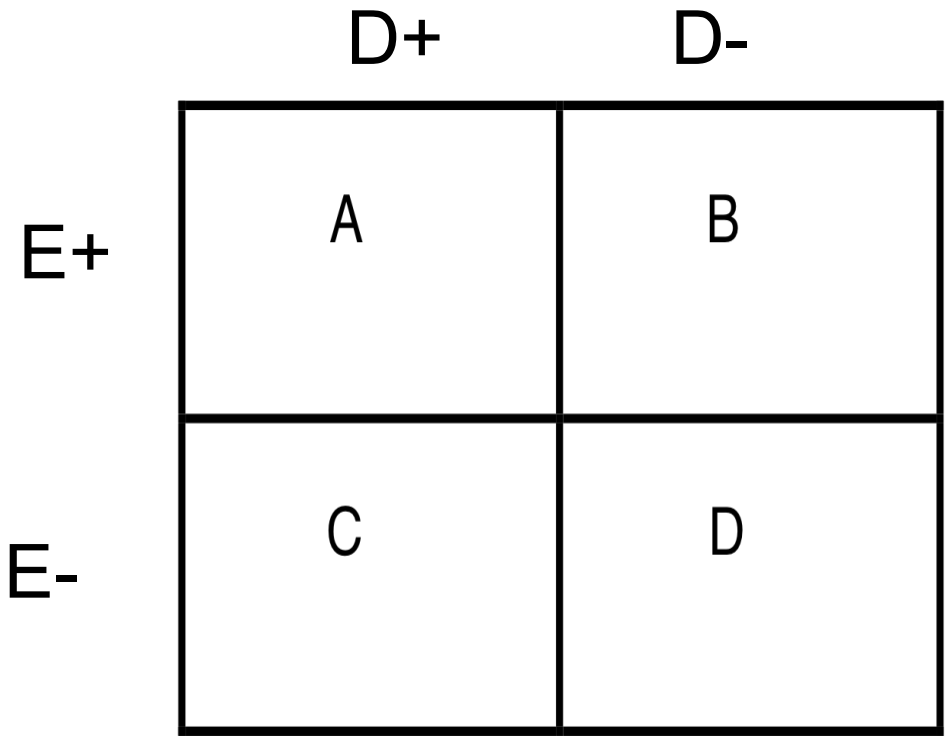
- Odds Ratio – The ratio of two odds. For **rare diseases**, this approximates relative risk. Commonly calculated in cross-sectional studies and case control studies, and from logistic regression.
  - Interpretation:
    - >1 suggests positive association
    - <1 suggests negative association
    - =1 suggests no difference between groups

Odds Ratio – General Definition



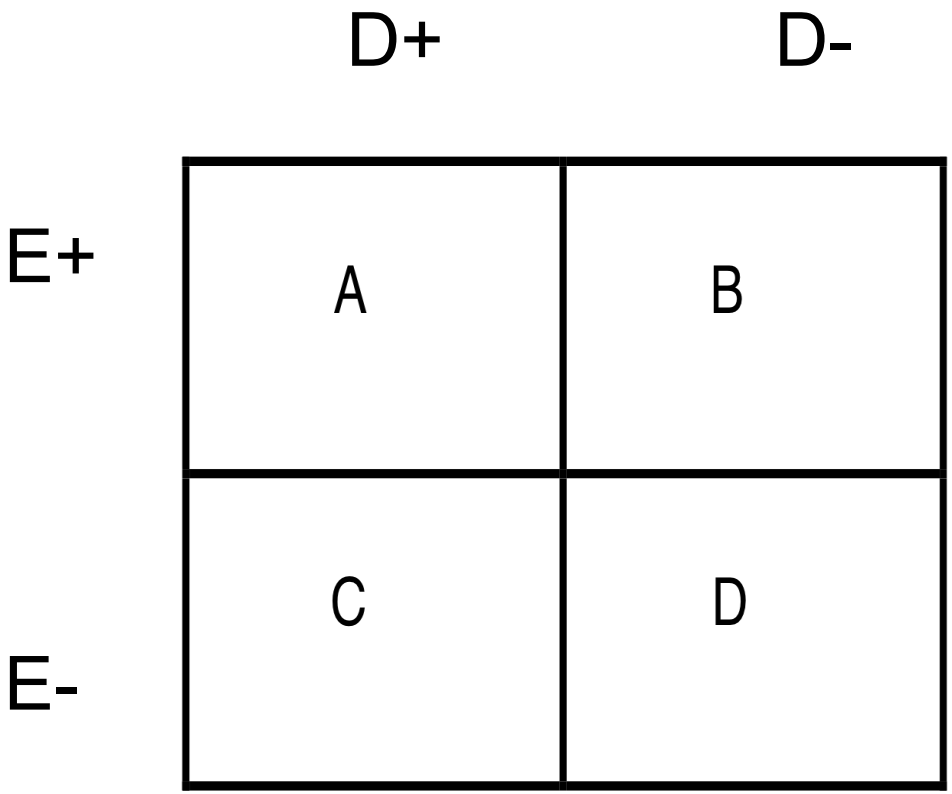
OR = odds of disease for E+ = A/B = AD
odds of disease for E- = C/D BC

Relative Risk



$$RR = \frac{\text{Risk of disease for E+}}{\text{Risk of disease for E-}} = \frac{A/(A + B)}{C/(C + D)}$$

Absolute Risk



- AR = (Risk for E+) - (Risk for E-) =

  A/(A + B) - C/(C + D)

6. Calculate the following statistics

- Suppose researchers conducted a study with 2000 people: 1000 took a new drug to prevent stroke for five years, and 1000 were given standard therapy. At the end of the trial, 2% of the people in the standard therapy group had experienced a stroke, compared to only 1% in the group taking the new drug.
- Calculate: RRR, ARR, NNT

Answers

RRR = 1%/2% = 50%

•Relative calculations may be misleading

ARR=2%-1%=1%

•Sounds like more modest benefit. One man in 100 will receive benefit.

NNT=1/ARR=100.

•100 men had to receive the new drug for 5 years for one man to benefit (for one less stroke to occur).

7. A clinician-researcher wishes to answer the question "How many of my patients would need to receive this preventive intervention to prevent one of them from developing disease?"

A. Number needed to treat (NNT)
B. Attributable risk (AR)
C. Population Impact Number (PIN)
D. Attributable Risk Reduction (ARR)
E. Population Attributable Risk (PAR%)

ANSWER: A clinician-researcher wishes to answer the question "How many of my patients would need to receive this preventive intervention to prevent one of them from developing disease?"

A. Number needed to treat (NNT)

B. Attributable risk (AR)

C. Population Impact Number (PIN)

D. Attributable Risk Reduction (ARR)

E. Population Attributable Risk (PAR%)

8. What is the purpose of using double-blinding in an RCT?

A. Achieve greater comparability of cases and controls
B. Avoid placebo effects
C. Avoid objective and subjective bias
D. Reduce the effects of sampling variation
E. Reduce the effects of loss to follow-up

ANSWER: What is the purpose of using double-blinding in an RCT?

A. Achieve greater comparability of cases and controls
B. Avoid placebo effects
C. Avoid objective and subjective bias
D. Reduce the effects of sampling variation
E. Reduce the effects of loss to follow-up

Bias

- Systematic Error – Deviation of results from the truth or - any process or effect at any stage of a study from its design to its execution to the application of information from the study, that produces results or conclusions that differ systematically from the truth.
  - Initial selection of participants for a study
  - Continued participation in a study
  - Methods of measurement

Selection Bias

- **Selection Bias –** A bias in assignment that arises from study design rather than by chance. These can occur when the study and control groups are chosen so that they differ from each other by one or more factors that may affect the outcome of the study (a potential problem in case control studies).

Advantages of randomization:

- Achieve non-predictability of the assigned to treatment
- Less worry that any subjective biases of the investigators, whether inadvertent or purposeful, are not introduced into the selection for treatment.
- Also hope that it will make groups similar based on characteristics other than treatment assignment, but this is not guaranteed (due to chance - see Table 1 in paper)

Blinding can occur without concealed allocation (and vice versa)

- **Example: Use of surfactant in NICU**
- **Example: PT vs surgery for DJD**

- **Trials with unconcealed allocation consistently overestimate benefit by 40%.**

Schultz KF, Chalmers I, Hayes RJ, et al. JAMA 1995;273:408-12
Schultz KF, Grimes DA. Lancet 2002;359:614-18

Intention-to-Treat Analysis

- Patients are analyzed in the groups to which they are assigned
- Attempts to reflect "real world" clinical situation in which not all patients are compliant
- Compliant subjects always do better "overall"

Non-Response Bias

- **Non-response bias** - How do respondents and non-respondents differ in regard to the study question? In general, respondents tend to be more educated compared to non-respondents.

Loss-to-Follow-Up Bias

- **Loss-to-follow-up Bias** – Even if the study sample was representative of the population from which it was derived at the beginning of a study, it may not be by the end of the study. This is a potential problem in cohort studies and clinical trials.
  - It may be more difficult to maintain long-term follow-up of patients of lower SES.
  - Patients may drop out of a clinical trial because of symptoms they are having that may be due to the study drug.

Measurement Bias

- **Measurement bias – Were measurement methods consistently different between groups in a study?**
  - **Lead-Time Bias**: If study patients are not enrolled at similar, well-defined points in the course of their illness, differences in outcome over time may simply reflect differences in the duration of their illness. For example, persons diagnosed using screening tests will be observed to live longer than those diagnosed based on clinical symptoms.
  - **Recall Bias**: Systematic Error due to the differences in accuracy or completeness of recall to memory of past events or experiences. A potential problem in case-control studies, for example.

Confounding

- **Confounding** may be considered "a confusion of effects" - attributing a result or disease to a specific risk factor when it is in fact due to another factor It can lead to over- or under-estimation of an effect or can even change the direction of the effect.

Confounding

- Researchers may attempt to control confounding in several different ways:
  - **Matching**: Infants with intussusception were matched to controls of the same age and birth location (they attempted to match them to infants born in the same hospital on the same day). Age is related both to the probability of having been vaccinated with RRV-TV and to the risk of intussusception.
  - **Regression** (a statistical procedure): "Variables used to adjust the odds ratios were related to both the risk of intussusception and to vaccination with RRV-TV." The reported adjusted odds ratios were adjusted for sex, mother's level of education, type of health insurance, type of mild or formula used for feeding, and time of first intake of solids.

Effect Modification

- Does the relationship between the predictor variable (risk factor) and outcome variable (disease) vary among different subgroups of a population? (Statistical term is “interaction”).
  - **Example**: “The risk of intussusception three to seven days after the first dose of RRV-TV was lower among infants fed breast milk (adjusted odds ratio, 10.7; 95%CI, 1.4 to 78.7) than among other vaccinated infants (adjusted odds ratio, 43.3; 95%CI, 12.7 to 148.1). However, the difference between these two estimates was not statistically significant (p=0.22).”

Hill's Causal Criteria

- Strength
- Consistency
- Specificity
- Temporality
- Biologic gradient
- Plausibility
- Coherence
- Experimental evidence
- Analogy

9. A journal publishes results of an RCT that showed a statistically significant difference in outcomes between treatment and control groups. The editorial that accompanies the article argues however that the results were not clinically significant. Why may this be?

Answer

- The effect size was very small and the sample size was very large.
- Statistical significance means that the difference was not likely to have occurred by chance.
- A large sample size can show even a small effect/contrast between groups as significance.
- It is important to consider effect sizes in interpreting study results.

10. In many studies examining the association between estrogens and endometrial cancer of the uterus, a one-sided significance test was used. The underlying assumption justifying a one-sided rather than a two-sided test is:

A. The distribution of the proportion exposed followed a "normal" distribution.
B. The expectation prior to doing the study was that estrogens cause endometrial cancer of the uterus.
C. The pattern of association could be expressed by a straight line function.
D. The type II error was the. Most important potential error to avoid.
E. Only one control group was being used.

Answer: In many studies examining the association between estrogens and endometrial cancer of the uterus, a one-sided significance test was used. The underlying assumption justifying a one-sided rather than a two-sided test is:

A. The distribution of the proportion exposed followed a "normal" distribution.
B. The expectation prior to doing the study was that estrogens cause endometrial cancer of the uterus.
C. The pattern of association could be expressed by a straight line function.
D. The type II error was the. Most important potential error to avoid.
E. Only one control group was being used.

Hypothesis Testing

- Random sampling error exists in all epidemiological studies. Hypothesis testing allows us to account for this random error and to determine whether a result is "statistically significant."
- **Hypothesis Testing** – Statistically test the study hypothesis against the **null hypothesis** (the null hypothesis is the nothing hypothesis - says there is no association between two variables – i.e. between risk factor and disease).
- **Study Hypothesis** – i.e. - There <u>is an association</u> between sex & race and physicians' recommendations for cardiac catheterization.

Potential for Error

- **Type** I **error** = the incorrect rejection of a true null hypothesis (a "false positive")

- **Type II error** = incorrectly retaining a false null hypothesis (a "false negative").

p-Value

- **Test statistic** – A value quantifying the degree of association between two variables that is calculated from the statistical test procedure. For example, a chi-square statistic.
- **p-Value** - The probability of obtaining a value for the test statistic as extreme or more extreme as that observed if the null hypothesis were true (also calculated from the statistical test procedure). A p-Value quantifies the degree of random variability in the sampling process.

p-Value

- **Statistical Significance** – Most researchers are willing to declare that a relationship is statistically significant if the chances of observing the relationship in the sample when nothing is going on in the population are less than 5%. This is why the commonly accepted cut point for calling a result “statistically significant is $p<0.05$.

Confidence Intervals

- Another value that can be calculated from statistical test procedures that accounts for random sampling error.
- 95% Confidence Intervals (95% CI) are commonly reported.
- 95% CI – A range of values computed from the sample that should contain the true population parameter with 95% probability in repeated collections of the data (i.e. a range of values that is almost sure to contain the true population parameter).

Confidence Intervals

- The width of a confidence interval is inversely proportionate to the sample size of the study.
- For risk ratios and odds ratios, if the confidence interval includes the value “1,” the association is not “statistically significant.”
- If the confidence intervals for measures in two groups overlaps, the two groups do not differ “significantly” with respect to that measure.

Important!

- p-Values and Confidence Intervals assume that there is no **bias, or systematic error**, in the study - i.e., they do not account for bias in the study. They do not assure that the association is real. They do not quantify **clinical significance**. It is important not to completely discount values that are not statistically significant. One must also look at trends and how the results compare to previous studies.

11. A randomized trial comparing the efficacy of two dugs showed a difference between the two (with a p value of <0.05). Assume that in reality, however, the two drugs do not differ. This is therefore an example of:

A. Type I error (alpha error)
B. Type II error (beta error)
C. 1-alpha
D. 1-beta
E. None of the above

11. A randomized trial comparing the efficacy of two drugs showed a difference between the two (with a p value of <0.05). Assume that in reality, however, the two drugs do not differ. This is therefore an example of:

A. Type I error (alpha error)

B. Type II error (beta error)

C. 1-alpha

D. 1-beta

E. None of the above

12. What is the probability and odds of the following in Vegas?

- Heads in a fair coin toss
- Drawing a red card from a standard deck
- Drawing a club card from a standard deck

Answer: What is the probability and odds of the following in Vegas?

- Heads in a fair coin toss

1/2=50%, 1:1

- Drawing a red card from a standard deck

26/52=50%, 1:1

- Drawing a club card from a standard deck

13/52=25%, 1:3

Sample Size

- If no difference found, was study power adequate?
  - Power is the ability of the study to find a difference IF one truly exists
  - Studies with a larger sample size will have greater power than those with smaller sample size

Diagnosis


Copyright 2002 by Randy Glasbergen.
www.glasbergen.com
GLASBERGEN
"We can't find anything wrong with you, so we're going to treat you for Symptom Deficit Disorder."

Use of BNP in Diagnosis of Congestive Heart Failure

- Are the results of this diagnostic article valid?
- Are the valid results of this diagnostic study important?
- Can we apply this valid, important evidence about a diagnostic test in caring for our patient?

Are the results of this diagnostic study valid?

Was there an independent, blind comparison with a reference ("gold") standard of diagnosis?	
Was the diagnostic test evaluated in an appropriate spectrum of patients (like those in whom it would be used in practice)?	
Was the reference standard applied regardless of the diagnostic test result?	
Was the test (or cluster of tests) validated in a second, independent group of patients?	

***Examine key elements of study design.**

Are the valid results of this diagnostic study important?

	Disease + Has CHD	Disease - Does not have CHD
Test + BNP 80 or more	A 95	B 11
Test - BNP<80	C 2	D 142

Sensitivity

- SnNout

When a sign, test or symptom has a high sensitivity, a negative result tends to rules out the diagnosis.

Specificity

- SpPin

When a sign, test or symptom has an extremely high specificity (say, over 95%), a positive result tends to rule in the diagnosis.

Predictive Values

- Positive and negative predictive values are similar to specificity and sensitivity, but are dependent on prevalence of disease in the population (your pre-test probability).
- The positive predictive value is simply the post-test probability of disease after a positive test result.
- The negative predictive value is the post-test probability of NO disease after a negative test result.

13. Calculate the following statistics

- Sensitivity=a/(a+c)=
- Specificity=d/(b+d)=
- Likelihood ratio for a positive test result

LR+=sensitivity/(1 − specificity)=

- Likelihood ratio for a negative test result
- LR-=(1-sensitivity)/specificity=

Are the valid results of this diagnostic study important?

- **Pre-test odds=**

prevalence/(1 − prevalence)=

- **Post-test odds=pre-test odds×LR**

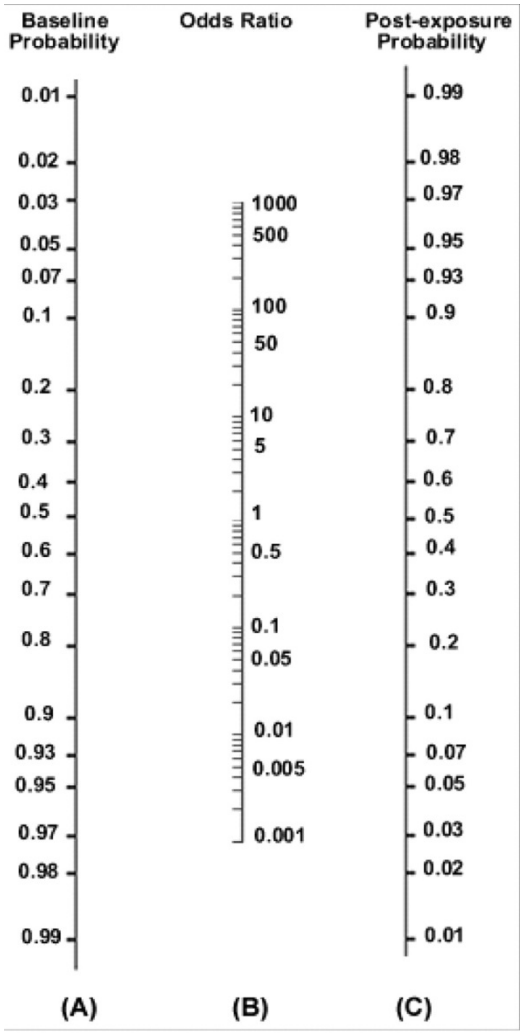
- **Post-test probability=**

post-test odds/(post-test odds+1)

Interactive Nomogram

- http://www.cebm.net/index.aspx?o=1161

Nomogram to calculate postexposure probability given estimates of the odds ratio and baseline probability. This nomogram is equivalent to the Bayes' nomogram, but with different labels.



Page, J. et al. Evid Based Med 2003;8:132-134

Can we apply this valid, important evidence about a diagnostic test in caring for our patient?

Is the diagnostic test available, affordable, accurate, and preceise in our setting?	
Can we generate a clinically sensible estimate of our patient's pre-test probability (from personal experience, prevalence statistics, practice databases, or primary studies)?	
Will the resulting post-test probabilities affect our management and help our patient? **Could it move across a test–treatment threshold?** **Would our patient be a willing partner in carrying it out?**	
Would the consequences of the test help our patient?	

14. Two pediatricians want to investigate a new laboratory test that identifies streptococcal infections. Dr. Kidd uses the standard test, which has a sensitivity of 90% and a specificity of 96%. Dr. Childs uses the new test, which is 96% sensitive and 96% specific.

If 200 patients undergo both test which of the following is correct?

A. Dr. Kidd will correctly identify more people with streptococcal infection than Dr. Childs
B. Dr. Kidd will correctly identify fewer people with streptococcal infection than Dr. Childs
C. Dr. Kidd will correctly identify more people without streptococcal infection than Dr. Childs
D. The prevalence of streptococcal infection is needed to determine which pediatrician will correctly identify the larger number of people with disease

Answer:Two pediatricians want to investigate a new laboratory test that identifies streptococcal infections. Dr. Kidd uses the standard test, which has a sensitivity of 90% and a specificity of 96%. Dr. Childs uses the new test, which is 96% sensitive and 96% specific.
If 200 patients undergo both test which of the following is correct?

A. Dr. Kidd will correctly identify more people with streptococcal infection than Dr. Childs
B. Dr. Kidd will correctly identify fewer people with streptococcal infection than Dr. Childs
C. Dr. Kidd will correctly identify more people without streptococcal infection than Dr. Childs
D. The prevalence of streptococcal infection is needed to determine which pediatrician will correctly identify the larger number of people with disease

References

- Last, John M. *A Dictionary of Epidemiology*, 4th edition, Oxford Press (2001).
- *Clinical Epidemiology & Evidence-Based Medicine Glossary,* http://www.vetmed.wsu.edu/courses-jmgay/GlossClinEpiEBM.htm#Introduction and Usage, June 22, 2002.
- Gordis, Leon. *Epidemiology*. WB Saunders, (1996).
- *Information Mastery: Evidence-Based Practice in the Real World*. Abingdon, VA. Dec 3-4, 2004