Introduction to Biostatistics and Epidemiology

David Dayya, D.O., M.P.H.
Saint Barnabas Hospital
Department Of Family Medicine

The Plain Truth Asserted!

- All studies are flawed!
- Most Studies are seriously flawed!
- What is the definition of a good study?

"Physicians can do more to admit the existence of uncertainty, both to themselves and to their patients. Although this will undoubtedly be unsettling, it is honest, and it opens the way for a more intensive search for ways to reduce uncertainty."

DAVID M. EDDY, MD, PhD

- ***Eddy DM. Clinical Decision Making. Chicago; American Medical Association, 1996***

Research Methods Lecture series

- 1) Intro to Biostatatistics and Epidemiology 1
- 2) Intro to Biostatatistics and Epidemiology 2
- 3) Screening
- 4) Medical Informatics
- 5) Advanced Topics in Statistics

Primary Objectives

- To introduce the physician to the fundamental principles that entail the process of research.
- To develop a foundation for Research Skills Development and the knowledge base necessary for interpreting the medical literature.

Anatomy of a Research Article

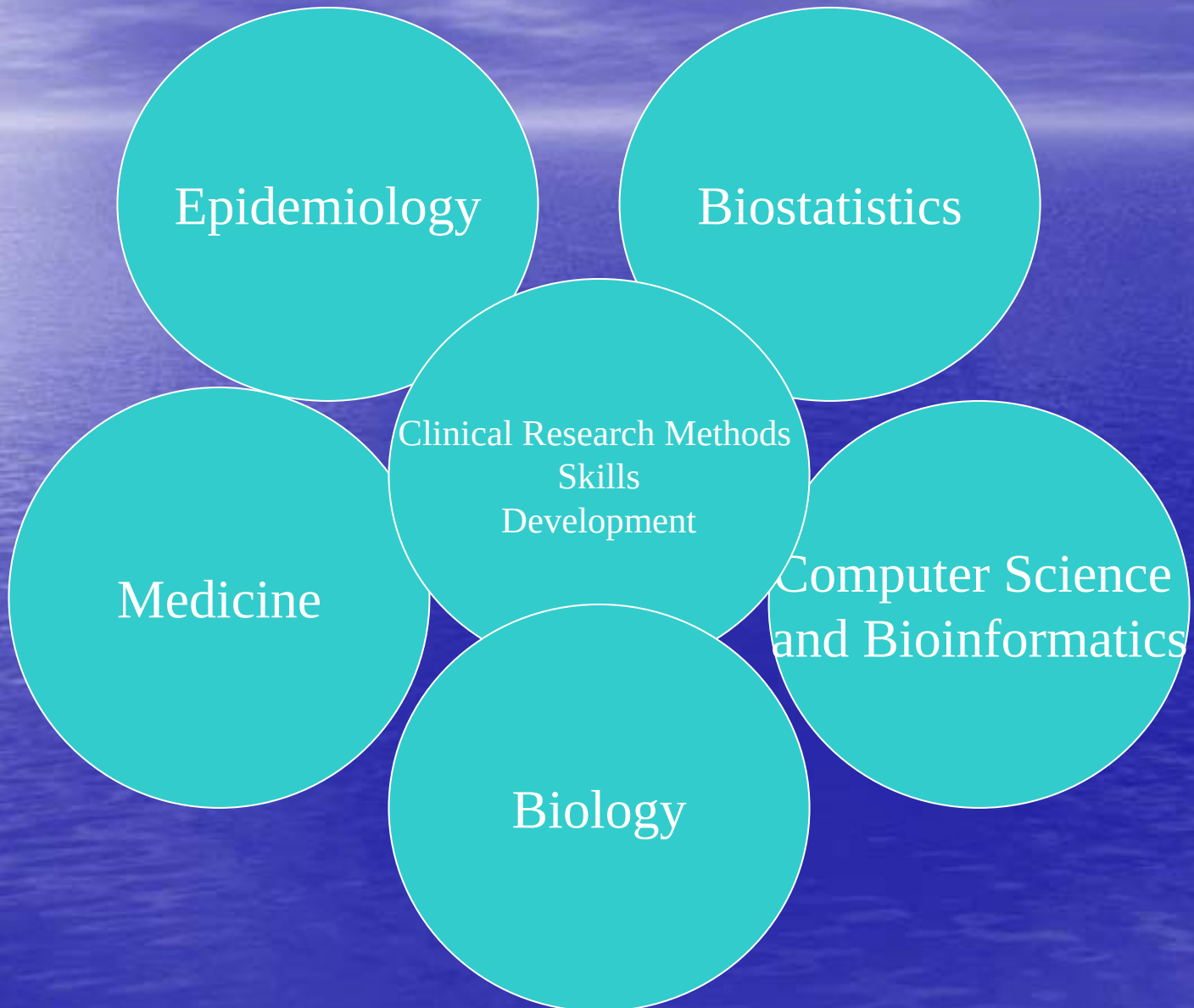
- Abstract
- Background/Introduction
- Methods
- Results
- Discussion/Conclusion

- A brief summary of the Study and its findings.
- Historic overview information related to the research question and its relevance.
- A detailed discussion of the research design including limitations.
- Descriptive statistics, graphical representation of data, and results of hypothesis testing and findings.
- An interpretation of the findings.

Overview

- Study Goals
- The Nature of Evidence and Causation
- Terminology
- Study Designs
- Strength of the Association
- Bias and Confounding
- References

The Interdisciplinary Nature of Clinical Research



Biostatistics (BY-oh-stuh-TIS-tix)

- The science of collecting and analyzing biologic or health data using statistical methods. Biostatistics may be used to help learn the possible causes of a disease or how often a disease occurs in a certain group of people. Also called biometry and biometrics. (NCI)

Epidemiology

Practical definition

The scientific study of the causes, influencing factors, and distribution of disease and health among people and applying that knowledge to the prevention and control of health problems.

EPIDEMIOLOGY

epi – upon

demos – people

logos – study

The scientific study of the distribution and determinants of health-related states or events in specified populations, and the application of resulting knowledge to the prevention and control of health problems

John Last's *Dictionary of Epidemiology* (4th edition) published by Oxford University Press in 2001

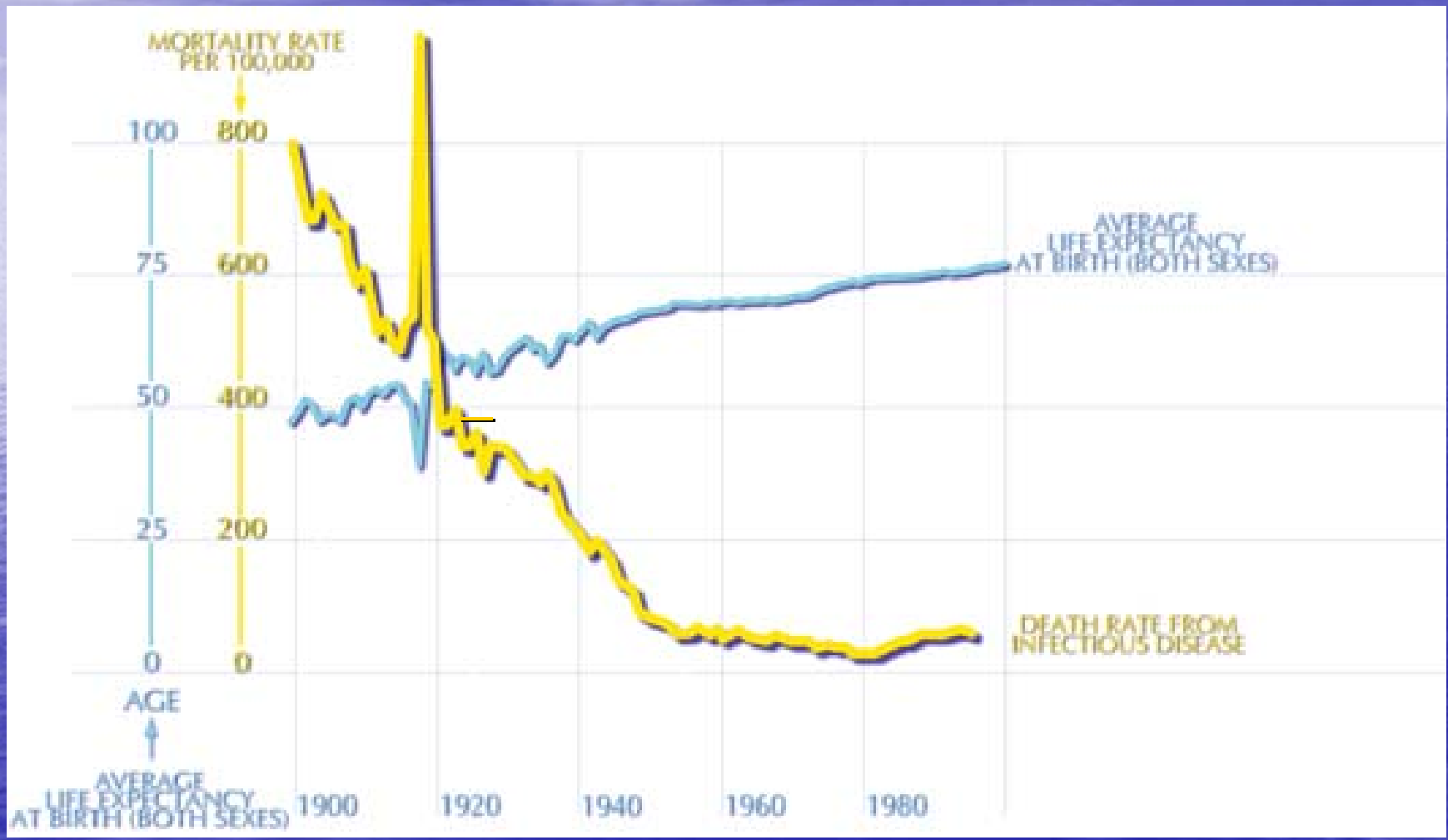
John Snow

"Greatest Doctor of all time!"

"A National Hero!"

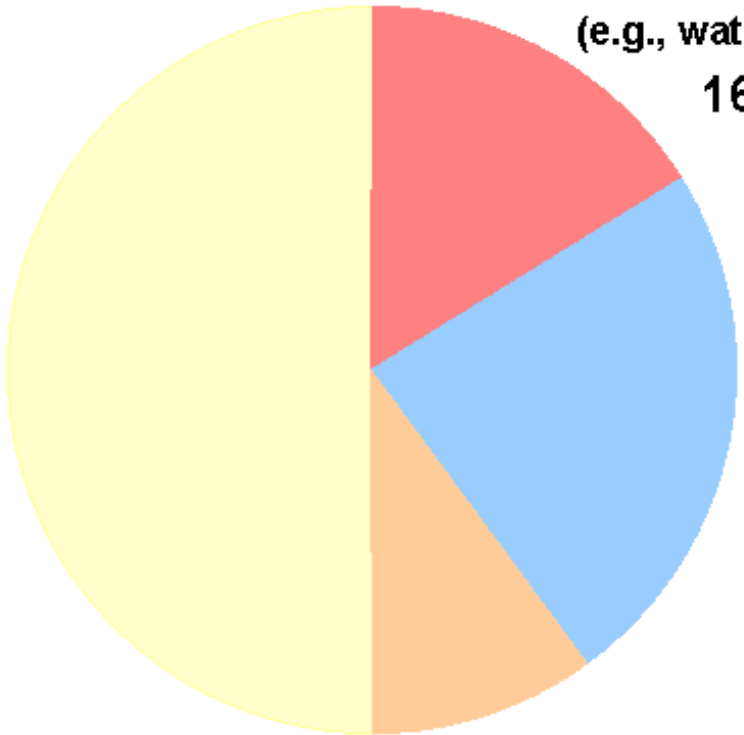
"Father of Epidemiology!"

Life Expectancy and Infectious Disease



MARIAN KOSHLAND
SCIENCE MUSEUM
OF THE NATIONAL ACADEMY OF SCIENCES

Factors Influencing Life Expectancy



Environmental Factors
(e.g., water quality)
16%

Lifestyle Risk Factors
(e.g., smoking, poor diet)
50%

Human Biolgy
(e.g., genetic predisposition to disease)
24%

Medical Care
10%

Source: Hinkle LE, ed. The Effect of the Man Made Environment on Health and Behavior: A Report of the Inter-University Board of Collaborators. Atlanta, Georgia: Centers for Disease Control, 1977.

What are the goals or purpose of any study?

- To determine if a Cause-Effect association or relationship exists.
- To rule out a Cause-Effect association or relationship.
- To establish or explain observed trends.

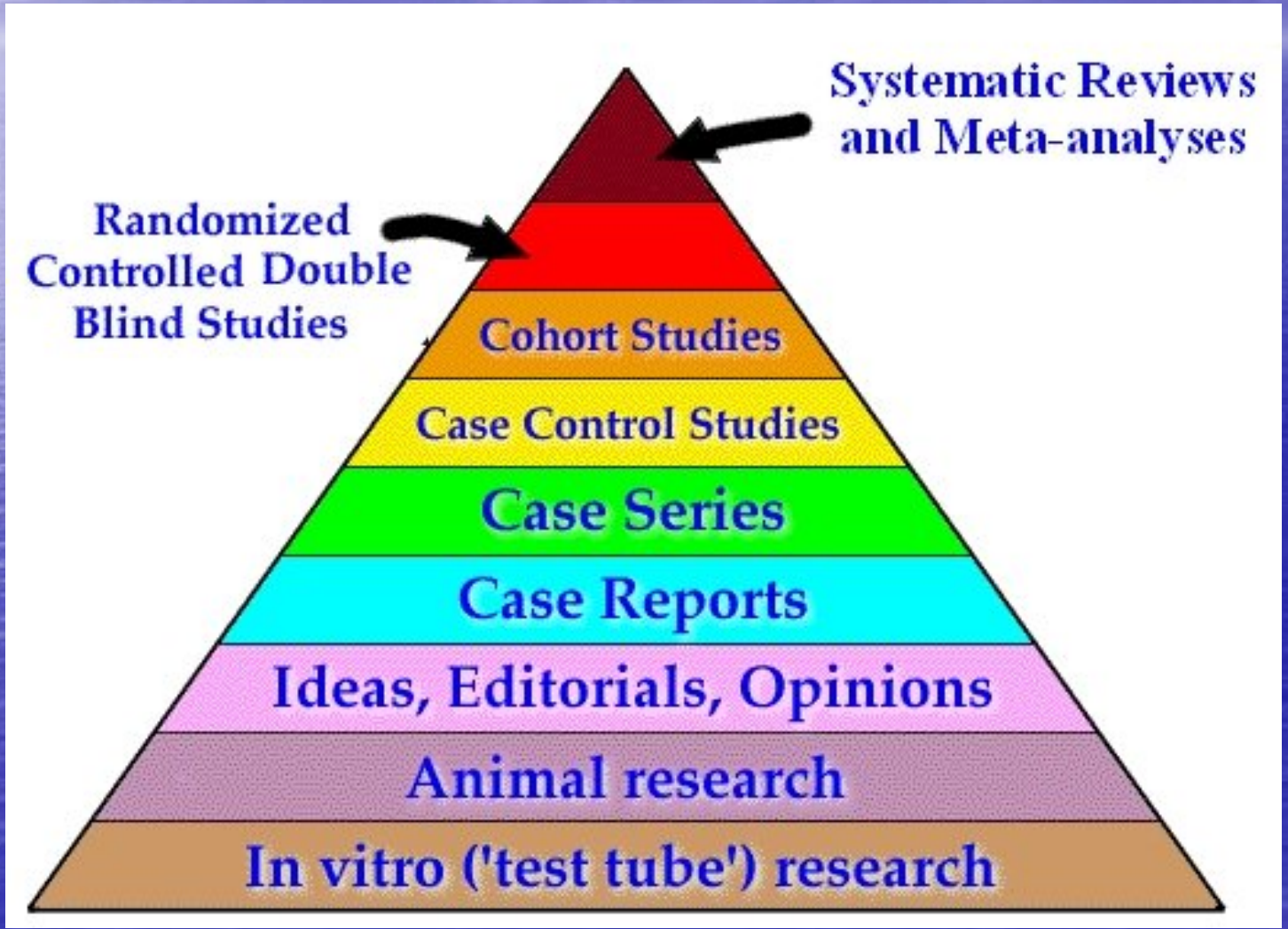
Evidence and Causation

- What is evidence?
- Good evidence, poor evidence, and no evidence?
- Levels of Evidence?
- How do you determine cause and effect relationship?
- How do you conclude that an association constitutes causation.

Evidence Based Medicine Hierarchy

I	Strong evidence from at least one published systematic review of multiple well-designed randomised controlled trials.
II	Strong evidence from at least one published properly designed randomised controlled trial of appropriate size and in an appropriate clinical setting.
III	Evidence from published well designed trials without randomisation, single group pre-post, cohort, time series or matched case-controlled studies.
IV	Evidence from well-designed nonexperimental studies from more than one centre or research group.
V	Opinions of respected authorities, based on clinical evidence, descriptive studies or reports of expert consensus committees.

Evidence Pyramid



Hill's Postulates for Causation (1965)

- **Strength of Association:** The larger the relative effect, the more likely the causal role of the factor.
- **Dose-response:** If the risk increases with increasing dose of the risk factor, the more likely the causal role of the factor. (Pancreatic cancer and coffee)
- **Consistency:** If similar associations are found in different studies in different populations, the more likely the causal role of the factor. (Literature review)
- **Temporality:** Risk factor exposure must precede the outcome. (Effect cannot precede cause)
- **Intervention:** Reduction or removal of the risk factor must reduce the risk of the outcome.
- **Biological Plausibility** a plausible mechanism exists that may explain the risk. (thimersol and autism)
- **Coherence**: Associations between the risk factor and the outcome must be consistent with existing knowledge. (exercise and obesity)

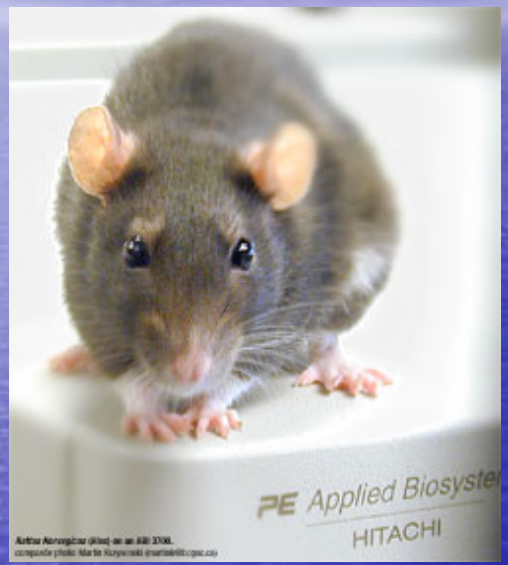
Definitions

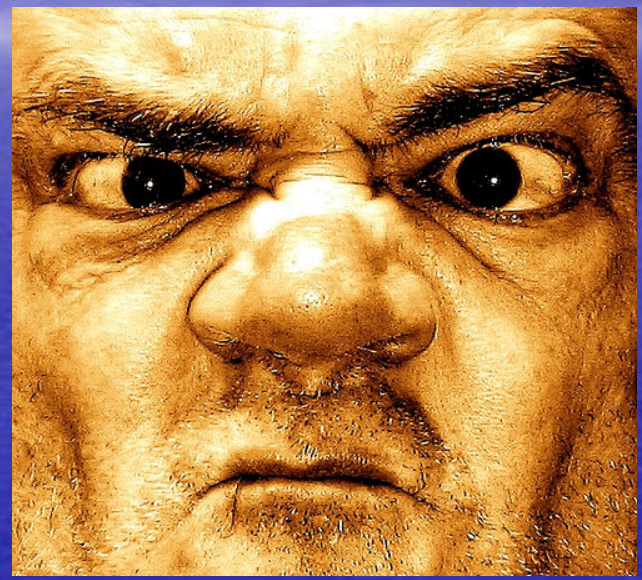
- Incidence = New cases of disease in a population, rate.
- Prevalence = Existing cases of disease in a population, proportion.
- Exposed = A segment of a population that is brought into contact with something.
- Unexposed = Comparison group, i.e. Control group, placebo group.

Definitions

- Population “relative”
- Sample - Segment of the population under study.
- Subjects or Participants
- Inclusion/Exclusion criteria
- Internal/External Validity

Species to Species Extrapolation





Reasons for Observed Effects or Apparent Associations (Three possibilities)

- True effect or association
- Random Error or Chance
- Systematic Error

Factors that must be considered in making conclusions about our observations include:

- Correlation or association does not necessarily imply causation nor does it necessarily imply coincidence. The converse is also true that lack of correlation or association doesn't necessarily imply a lack of causation.
- Efforts must be made to control or exclude any potential confounding variables and sources of bias.

Categories of Studies

- Retrospective
- Studies individuals already with disease and attempts to establish if there was past exposure. Can use prevalent or incident cases.
- Observes diseased individuals for evidence of past exposure.
- i.e. Case controlled study.

- Prospective
- Studies individuals with exposure and follows them forward in time from exposure to development of disease. “Incidence Type Study”.
- Observes exposed individuals for development of disease.
- i.e. Cohort, Incidence, Follow up, or Longitudinal Study.

Retrospective vs. Prospective

Incidence of disease cannot be calculated from this case control data.

Common sources of control general population, hospital patients, relatives of cases, associates or friends of cases. (related sample)

Sampling is random, systematic or paired.

Cases and controls selected from a medical facility/facilities, community or general population.

Subjects are diseased at onset of the study. Can select Incident cases or Prevalent cases.

Bias more common.

Rare Diseases

Less Cost

Less difficult logistics

Ethical considerations

Incidence of disease can be obtained from the Prospective data.

Common sources for control are a placebo group that is comparable to the exposed subjects.

Randomized pairing is preferred in the selection process.

Sample is drawn from the popoulation and randomized into exposure or control group.

Subjects are free of disease at onset of study.

Bias is less common.

Common Diseases

Greater Costs

More difficult logistics

Ethical considerations

Studies Continued

- Prospective
- Observational study
- Retrospective cohort
- Retrospective
- Nested case Control

Cross Sectional Study

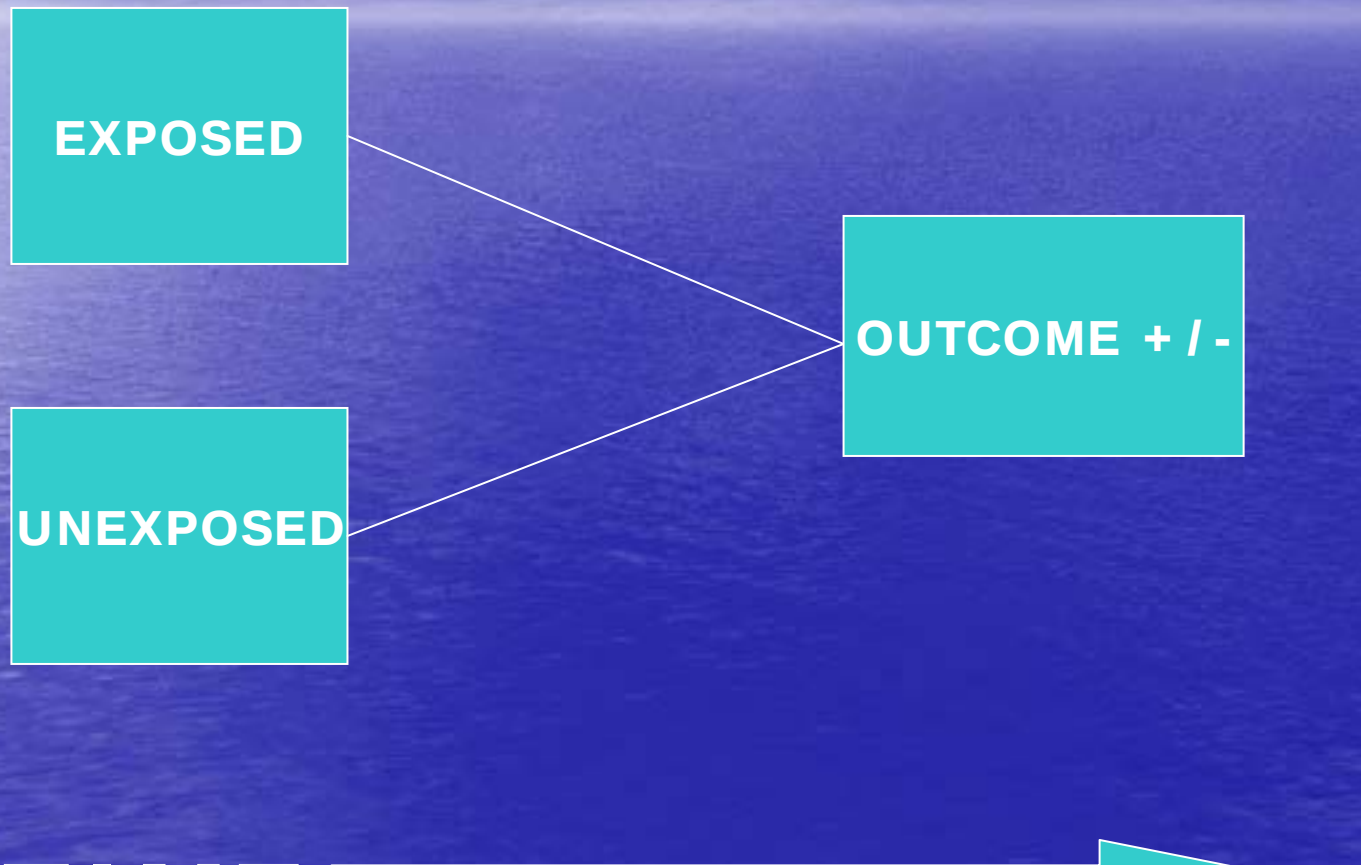
Case Report

- Anecdotal data but important
- Potentially hypothesis generating
- May lead to Observational, Case-Control, or Pilot Cohort study
- The “root” or route to establishing cause effect relationships.

Randomized Clinical Intervention Trial

- The Experimental Study (Gold Standard)
- A prospective study with randomization of the subjects to either the exposed or unexposed group. i.e. Coin toss, random numbered selection, computer generated random selection etc.
- Can include “Blinding” i.e. Single Blinded, Double Blinded, or Triple Blinded.

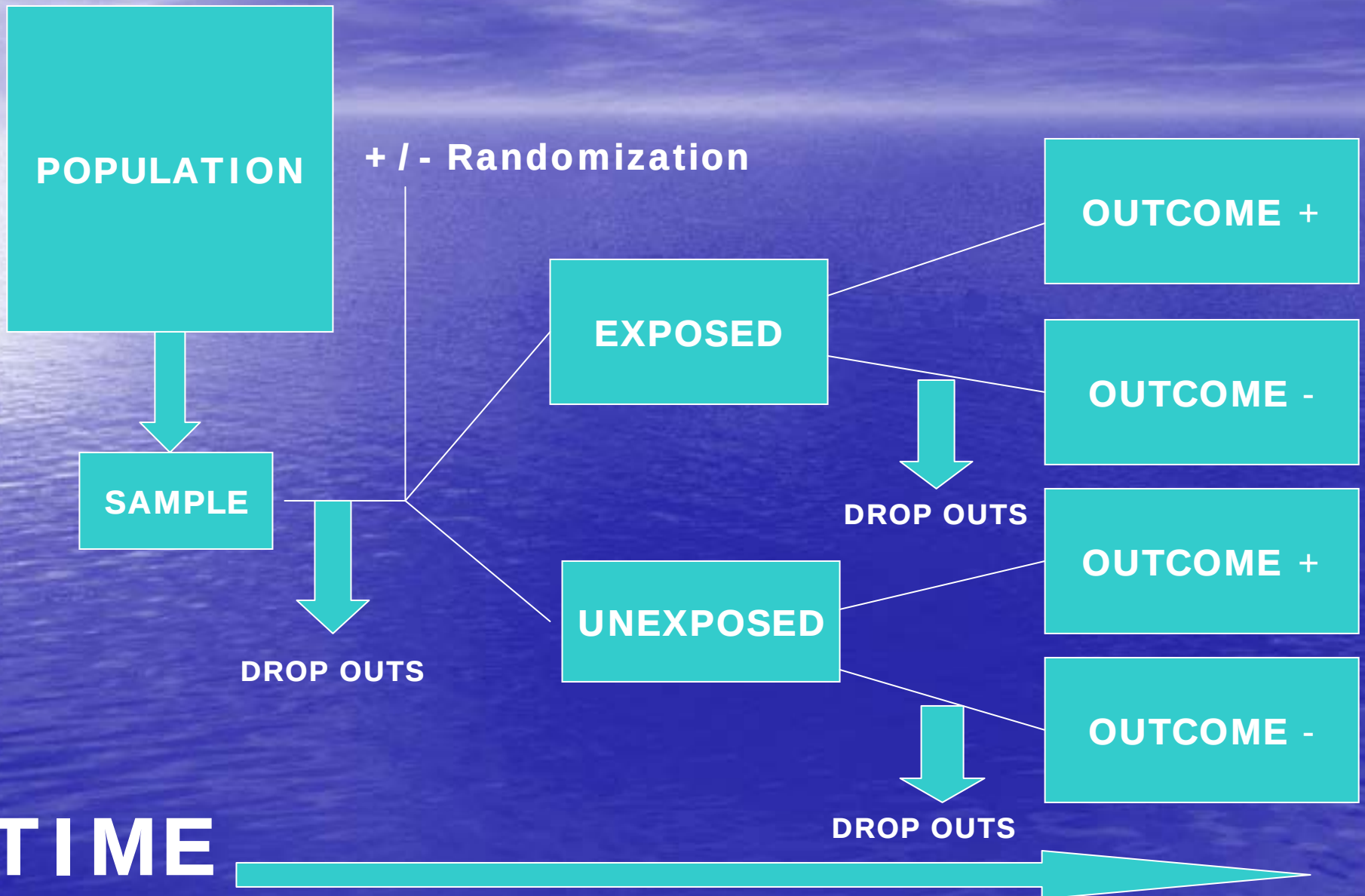
Prospective Design



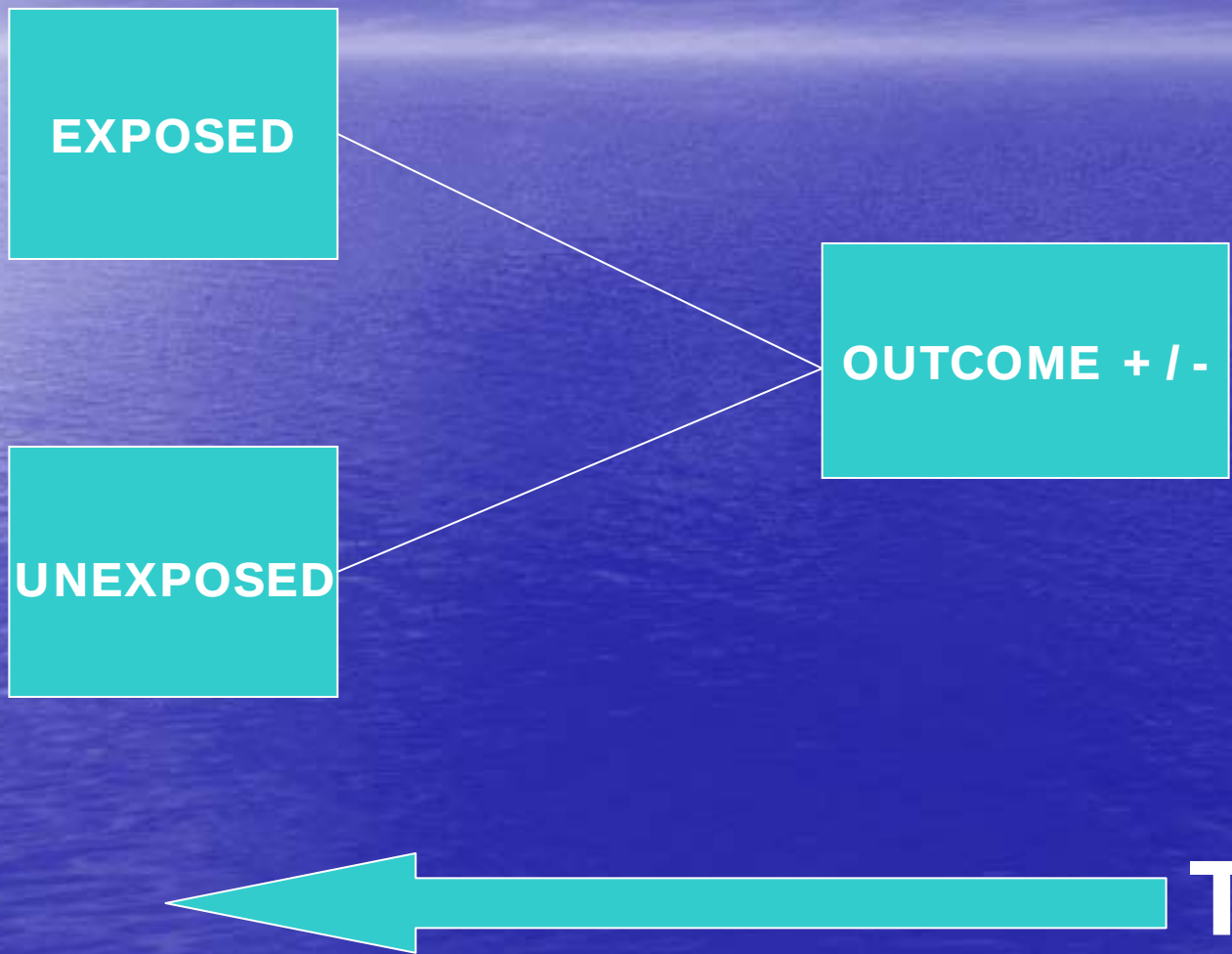
Blinding

- **Single Blinded** = Subject or patient unaware if they are in the exposed or unexposed group.
- **Double Blinded** = Subject or patient + principal investigator unaware if the subject is in the exposed or unexposed group.
- **Triple Blinded** = Subject or patient + principal investigator + Data examiner unaware if the subject is in the exposed or unexposed group.

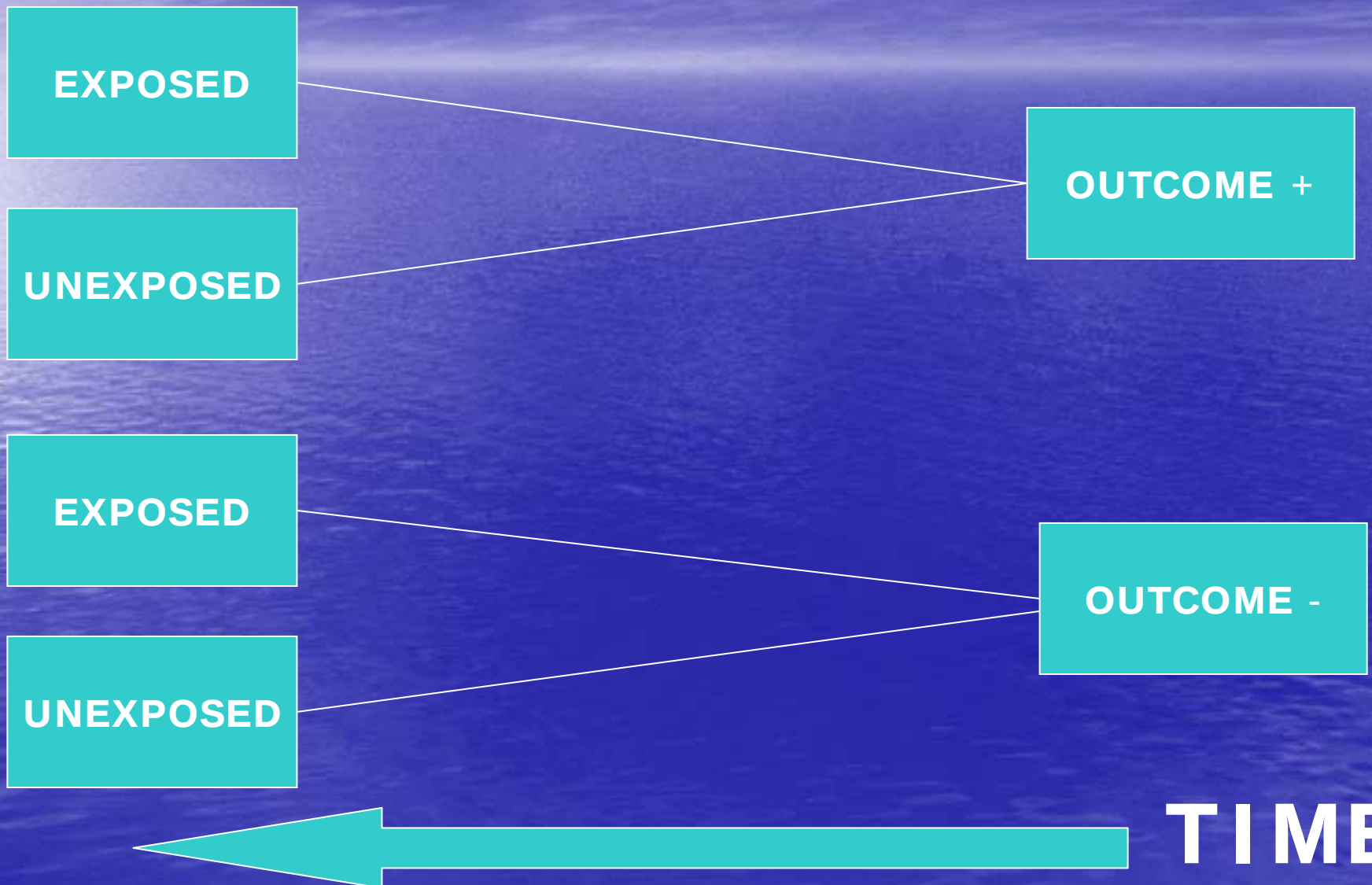
Prospective Design



Retrospective Design



Retrospective Design



2X2 TABLE

	OUTCOME Present	OUTCOME Absent	
Exposure Present	a	b	a+b
Exposure Absent	c	d	c+d
	a+c	b+d	a+b+c+d Population

Prevalence

Prevalence = (a+c)/(a+b+c+d)

Defined as the proportion of those people in the population with the disease.

Prevalence Example

	Lung Cancer Positive	Lung Cancer Negative	
Smoker	63	99937	100000
Nonsmoker	7	99993	100000
	70	199930	200000

Prevalence Example

70/200000 = 0.00035

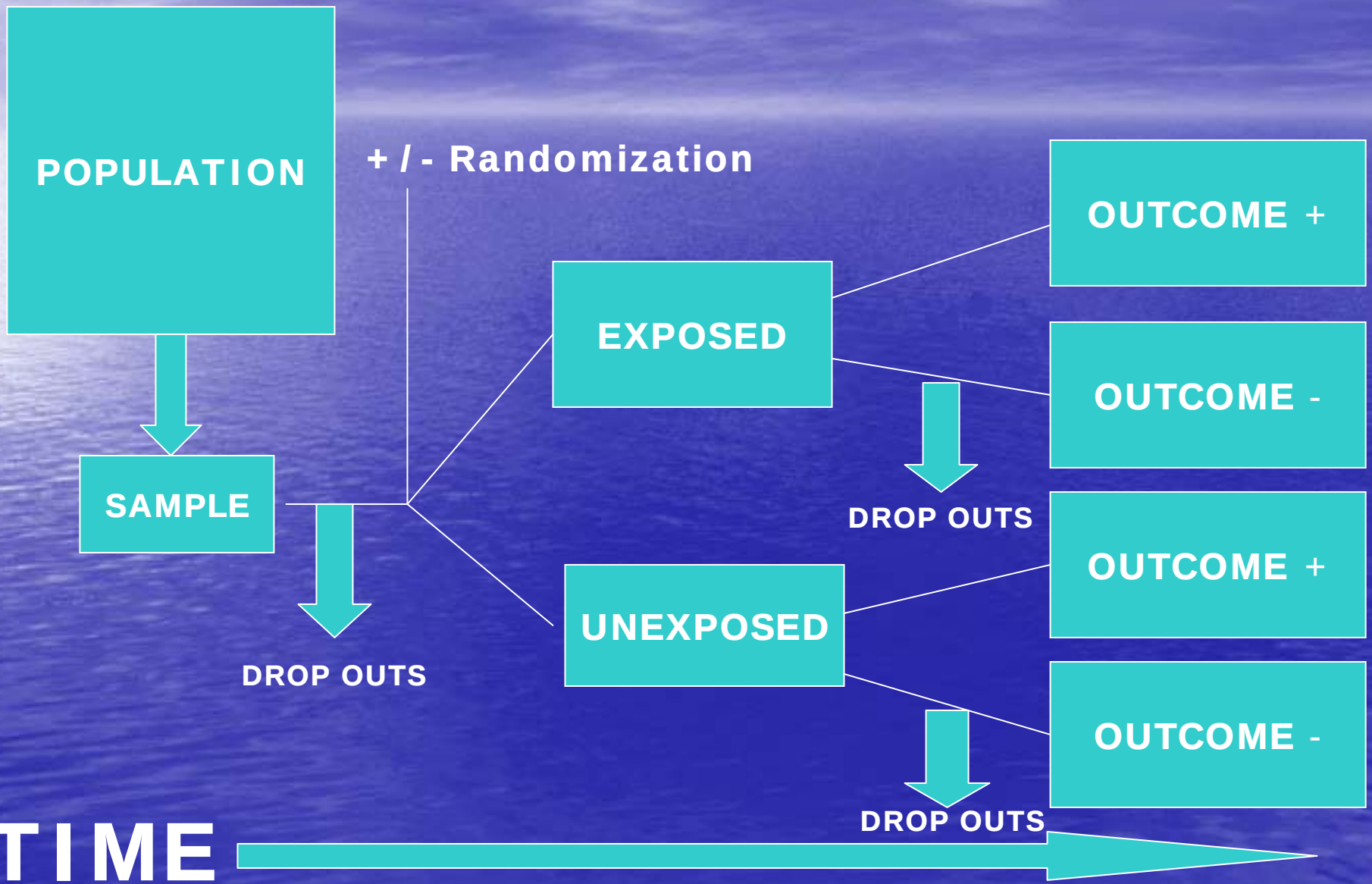
0.035%

OR 35/1000 OR 3500/100000

Strength of the Association

- Relative Risk Ratio
- Odds Ratio

Prospective Design



Relative Risk Ratio

Pertains to a prospective study.
A= Number of subjects who were exposed and developed the outcome.
A+B= Total number of subjects that were exposed in our population.
A/A+B = Proportion of subjects that were exposed and developed the outcome. RE (Risk of exposed group)

C = Number of subjects who were not exposed and developed the Outcome.
C+D = Total number of subjects that were not exposed in our population.
C/C+D = Proportion of subjects that were not exposed and developed the outcome. RU(Risk of unexposed group)

A/(A+B)/C/(C+D) = Relative Risk Ratio = Probability proportion Relative(Risk of one group compared to another.)
RE/RU>1, =1, <1

Relative Risk Ratio

A/A+B = 63/100000 = 0.00063

0.063 %

C/C+D = 7/100000 = 0.00007

0.007%

A/A+B C/C+D = 9

Risk Increase vs. Risk Reduction

- 100X risk =

9900% increased risk.

- 10X risk =

900% increased risk.

1.9X risk = 90% increased risk.

- 1.5X risk =

50 % increased risk.

- 1/2X risk = 50% decreased risk.

- 1/10X risk = 90% decreased risk.
- 1/100X risk = 99% decreased risk.
- 1/1000X risk = 99.9% decreased risk.

Attributable Risk

- (Rate in the exposed population) –(Rate in the unexposed population)
- [a/(a+b)]-[c/(c+d)]
- Sum of attributable risks may be greater than 100% because the disease may have multiple risk factors.
- Population Attributable Risk (PAR) is the proportion of a disease in a population attributed to a particular or given exposure)
- PAR = [PE(RR-1)]/[PE(RR-1)+1]

Attributable Risk Example

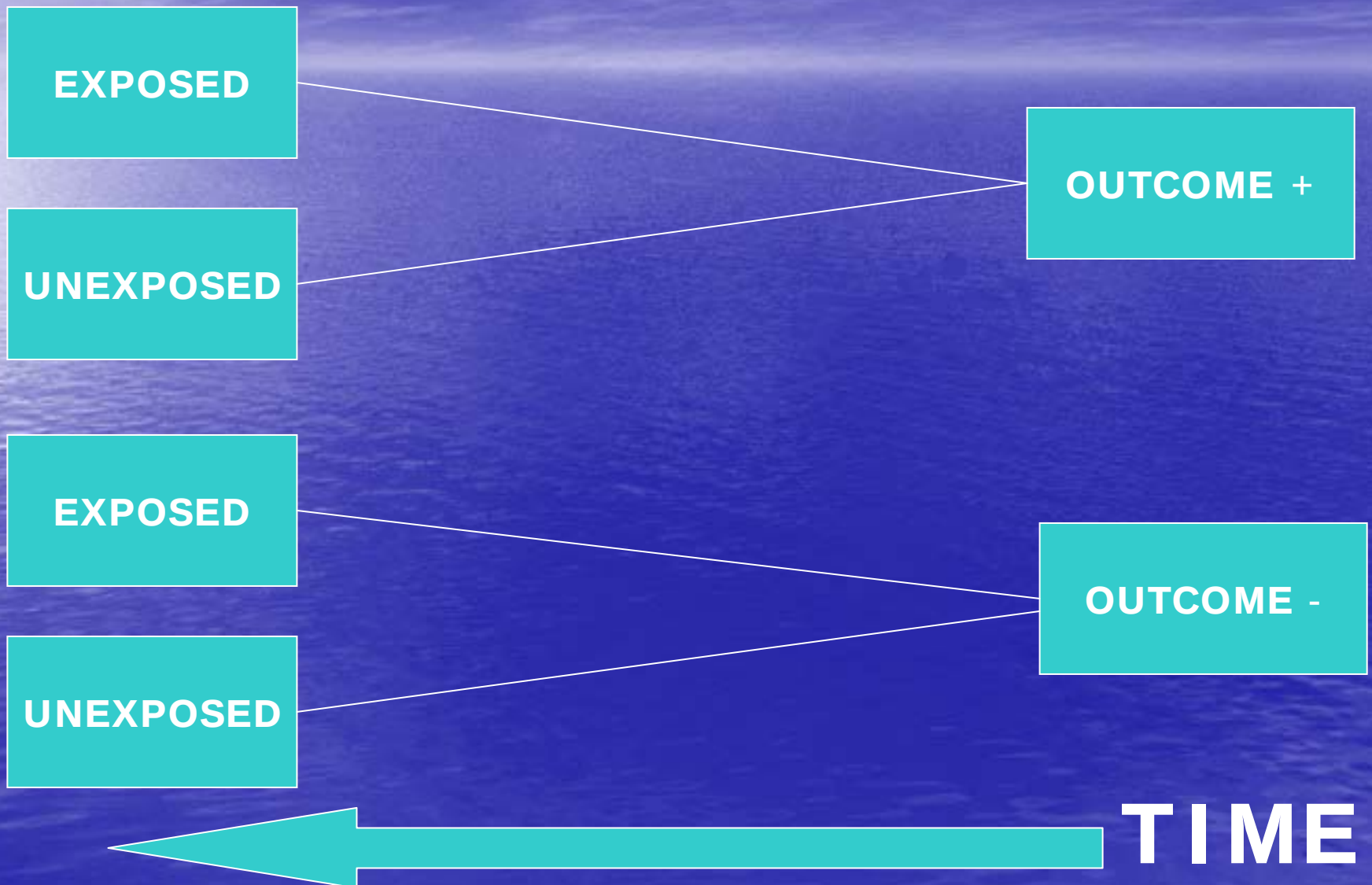
	Lung Cancer Positive	Lung Cancer Negative	
Smoker	63	99937	100000
Nonsmoker	7	99993	100000
	70	199930	200000

Attributable Risk

[A/A+B – C/C+D] =
0.00063-0.00007 = 0.00056

PAR = [PE(RR-1)] [PE(RR-1)+1]
= [.5 (9 – 1)] [.5 (9 – 1) +1]
= 4/5 = .8 or 80%

Retrospective Design



2X2 TABLE CC Study

	OUTCOME Present	OUTCOME Absent	
Exposure Present	a	b	a+b
Exposure Absent	c	d	c+d
	a+c	b+d	a+b+c+d Population

ODDS RATIO

- Pertains to a retrospective or case control study.
- a/c b/d= a/b c/d = ad/bc
- 3 possibilities
- (N/D)>1, (N/D) =1, (N/D)<1
- >1 implies a positive association.
- <1 implies a negative association.
- =1 implies no association.
- About equal numbers of cases (a+b) and controls (c+d) are usually selected.
- An estimate of relative risk.

ODDs Ratio Example

	Lung Cancer	No Lung cancer	
Smokers	80	30	110
Nonsmokers	20	70	90
	100	100	200

Odds Ratio example

$a/b \div c/d = 80 \times 70 / 30 \times 20 =$

$= 5600/600 = 9.3$

PAR [Population Attributable Risk]

PAR = [P(OR – 1)]/[P(OR – 1) +1]

= [.3(9.3 – 1)]/[.3(9.3 – 1) +1]

= 2.49/3.49 = 71%

Relative Risk, Attributable Risk Ratios Example

	Asbestosis Exposure	No Asbestosis	
Miners working in Area A	190	10	200
Miners not working in Area A	130	70	200
	320	80	400

ODDs Ratio Example

	Asbestosis	No Asbestosis	
Miners working in Area A	190	10	200
Miners not working in Area A	130	70	200
	320	80	400

Sources of Bias in Studies

- Bias – Favoring a particular outcome or conclusion.

- Many Types include
- Selection Bias
- Interviewer or Data Collection Bias
- Recall Bias

- Can occur in any study

The Confounding Variables

- Outside or extraneous factor exists that is associated with the exposure of interest and is an independent cause of the outcome of interest or disease observed.
- Confounder can explain an apparent association when one really doesn’t exist or conversely can explain why an association is not made when one really does exist.
- Can occur in any type of study.

Making Cause-Effect Conclusions

- Strength of the association?
- Dose –Response Relationship?
- Consistency i.e. Complements other existing findings i.e. other studies?
- Temporality i.e. Time sequence exists? (Cause Effect)?
- Intervention effect on risk?
- Biological mechanism is plausible?
- Coherence with existing knowledge?

Cause-Effect Conclusions

REMEMBER!

Invalid associations may be drawn secondary to Random error or Chance, and Systematic Error (Bias, or Confounding)?

CLINICAL RELEVANCE?

Reference List for Medical Research compiled by David Dayya, D.O., M.P.H.

Introductory Level References

- **Statistics/Biostatistics**

- Hinton PR. Statistics Explained. 2nd. Ed. 2004.
- Glantz SA. Primer Of Biostatistics. 6th Ed. 2005.
- Salkind NJ. Statistics For People Who Think They Hate Statistics. 2nd Ed. 2003.
- Moore. Introduction to the Principles of Statistics. 4th Ed. 2004.
- Kleinbaum DG, Klein M, Pryor ER. Logistic Regression: A Self learning Text. 2nd Ed. 2005.
- Kleinbaum DG, Klein M. Survival Analysis: A Self learning Text. 2nd Ed. 2005.
- Bluman AG. Probability Demystified. 1st Ed. 2005
- Gonick L, Smith W. Cartoon Guide to Statistics. 1st Ed. 1994.

References

- Scholarly Research Paper Publication/Bibliography Software

- Agrawal A. EndNote 1-2-3 Easy! Reference Management For the Professional. 1st ed. 2005.
- Maran R. Microsoft Office 2000 Simplified. 1999.
- Maran R. Maran Illustrated Office 2003. 1st Ed. 2005.

- Supplementary and Advanced Level References

- Statistics/Biostatistics

- Glantz SA, Slinker BK. Primer of Applied Regression and Analysis of Variance. 2nd Ed. 2000.
- Kleinbaum DG, Kupper LL, Nizam A, Muller KE. Applied Regression Analysis and Multivariable Methods. 4th Ed.. 2007.
- Winer BJ, Brown DR, Michels KM. Statistical Principles in Experimental Design, 3rd Ed.
- Snedecor GW, Cochran WG. Statistical Methods. 8th Ed. 1989.
- Maxwell SE, Delaney HD. Designing Experiments and Analyzing Data: A Model Comparison Approach. 2nd Ed.
- Keppel G, Wickens TD. Design And Analysis. A Researchers Handbook. 4th Ed. 2004.
- McMahon D. Linear Algebra Demystified. 1st Ed. 2005.
- Lay DC. Linear Algebra and its Applications. 3rd Ed. 2005.
- Clark-Carter D. Quantitative Psychological Research: A Students Handbook. 2004.
- Russo R. Statistics for the Behavioral Sciences. 2003.

- Epidemiology/Research Methods

- Evans JS, Evans BT. How To Do Research. 2005.
- Boynton PM. The Research Companion. A Practical Guide for the Social and Health Sciences. 2005.
- Medical Health Informatics
- Englebardt SP. Health Care Informatics: An Interdisciplinary Approach

References

- **Epidemiology / Research Methods**

- Gehlbach SH. Interpreting the Medical Literature. Practical Epidemiology For Clinicians. 5th Ed. 2006.
- Gordis L. Epidemiology. 3rd Ed. 2004.
- Hulley SB, Designing Clinical Research. An Epidemiologic Approach. 3rd Ed. 2006.

- **Dataset / Database Management**

  - Maran R. Microsoft Office 2000 Simplified. 1999.

- Maran R. Maran Illustrated Office 2003. 1st Ed. 2005.
- Maran R. Maran Illustrated Access 2003. 1st Ed. 2005.
- Maran R. Maran illustrated Excel 2003. 1st Ed. 2005.
- George D, Mallory P. SPSS For Windows: Step-by-Step. 7th Ed. 2006
- Hinton PR. Brownlow C, McMurray I. et. al. SPSS Explained. 1st Ed. 2004.
- Delwiche LD, Slaughter SJ. The Little SAS Book: A Primer. 3rd Ed. 2003
- Acock AC. A Gentle Introduction to Stata. 1st. Ed. 2005.

- **Scholarly Research Paper Publication / Bibliography Software**

- Agrawal A. EndNote 1-2-3 Easy! Reference Management For the Professional. 1st ed. 2005.
- Maran R. Microsoft Office 2000 Simplified. 1999.
- Maran R. Maran Illustrated Office 2003. 1st Ed. 2005.

References

- **Supplementary and Advanced Level References**

– Statistics/Biostatistics

- Glantz SA, Slinker BK. Primer of Applied Regression and Analysis of Variance. 2nd Ed. 2000.
- Kleinbaum DG, Kupper LL, Nizam A, Muller KE. Applied Regression Analysis and Multivariable Methods. 4th Ed.. 2007.
- Winer BJ, Brown DR, Michels KM. Statistical Principles in Experimental Design, 3rd Ed.
- Snedecor GW, Cochran WG. Statistical Methods. 8th Ed. 1989.
- Maxwell SE, Delaney HD. Designing Experiments and Analyzing Data: A Model Comparison Approach. 2nd Ed.
- Keppel G, Wickens TD. Design And Analysis. A Researchers Handbook. 4th Ed. 2004.
- McMahon D. Linear Algebra Demystified. 1st Ed. 2005.
- Lay DC. Linear Algebra and its Applications. 3rd Ed. 2005.
- Clark-Carter D. Quantitative Psychological Research: A Students Handbook. 2004.
- Russo R. Statistics for the Behavioral Sciences. 2003.

– Epidemiology/Research Methods

- (1)Evans JS, Evans BT. How To Do Research. 2005.
- (2) Boynton PM. The Research Companion. A Practical Guide for the Social and Health Sciences. 2005.

– Medical Health Informatics

- Englebardt SP. Health Care Informatics: An Interdisciplinary Approach
- Medical Informatics: Knowledge Management and Data Mining in Biomedicine

References

- **Useful WWW Online Resources**

- Rice University Virtual Lab in Statistics (online multimedia tutorial and textbook) www.onlinestatbook.com/rvls/
- UCLA Statistical Computing Online Tutorial on SAS, STATA, and SPSS. www.ats.ucla.edu/stat/overview.htm
- Practice Datasets www.vetmed.wsu.edu/appliedregression/

- **Video Instruction Resources**

- Against All Odds: Inside Statistics www.learner.org/resources/series65.html
- Statistics www.videoaidedinstruction.com/