Harvard Medical School: Essentials of the Profession I

Clinical Epidemiology and Population Health (CEPH) Syllabus

Session 1 - Study Design and Causality

Format: Lecture and Small Groups

Session Overview

Welcome to Clinical Epidemiology and Population Health (CEPH)! We are very excited to meet you and learn together.

This first session will comprise a 45-minute lecture followed by a 75-minute small group discussion. We will discuss the framework for interpreting the literature that serves as the underlying structure for this portion of Essentials. We will review and discuss criteria for understanding whether associations we observe are likely to be causal. We will debate strengths and limitations of different study designs. Armed with this knowledge, you will break out to design a research study.

Learning Objectives

At the end of this session, students will be able to:

- Apply a structured framework for assessing evidence from research to clinical decisions.
- Consider how we move from research evidence to understanding of causality in science and medicine, including use of defined criteria and more complex causal models.
- Become familiar with the study designs most commonly found in clinical research, and some of the strengths and weaknesses in each.
- Understand the difference between absolute and relative measures of effect and the use of 2 x 2 tables to calculate each.

Student Preparatory Instructions

1. **Read:** Learning Objectives, Case, and Population Health Context, and Thought Questions
2. **Watch Video:** Epidemiological Studies – made easy! (10 min)
3. **Watch Video:** Incidence and Prevalence (10 min)
4. **Watch Video**: Determining Causality: A Review of the Bradford Hill Criteria (4 min)
5. **Read:** Wakeford R. Association and causation in epidemiology – half a century since the publication of Bradford Hill's interpretational guidance. Journal of the Royal Society of Medicine 2015, Vol 108(1) 4-6. (2 pages)
6. **Read:** Morgan DJ et al. Clinician conceptualization of the benefits of treatments for individual patients. JAMA Network Open 2021;4(7): e2119747. (7 pages)
7. **Read:** Key Points: Study Design and Causality
8. **Complete** the Readiness Assessment Exercise

Additional (Optional) Resources

1. **Read:** Walvik L, Svensson AB, Friborg J, Lajer CB. The association between human papillomavirus and oropharyngeal squamous cell carcinoma: Reviewed according to the Bradford Hill criteria for causality. Oral Oncology 2016;63:61-5.

2. **Reading for history buffs:** Sir Austin Bradford Hill. The environment and disease: association or causation? Observational Studies 6 (2020) 1-9. This article was originally published in the Proceedings of the Royal Society of Medicine, May 1965, 58, 295-300. Includes new comments by the following researchers follow: Peter Armitage; Mike Baiocchi; Samantha Kleinberg; James O'Malley; Chris Phillips and Joel Greenhouse; Kenneth Rothman; Herb Smith; Tyler VanderWeele; Noel Weiss; and William Yeaton.
3. **Read:** Shapiro S. Causation, bias and confounding: a hitchhiker's guide to the epidemiological galaxy. Part 1. Principles of causality in epidemiological research: time order, specification of the study base and specificity. J Fam Plann Reprod Health Care 2008; 34(2):83-7. This article generally covers the same material as the videos. Read through Section 1b. (pages 83-5)

Session 2 - Frequency and Association

Format: Small Group

Session Overview

In this two-hour small group session, we will discuss a study on the impact of an HPV vaccination program among adolescent girls. This topic should sound familiar after your breakout at the end of the last session. We will use the time to solidify your understanding of calculations for measures of frequency and association (Incidence, Prevalence, AR, RR, OR), practice calculating measures of population impact (PAR, PAR%), and discuss internal and external validity of this study.

Learning Objectives

At the end of this session, students will be able to:

- Calculate and interpret measures of risk including relative risk, odds ratio, risk difference/attributable risk, and population attributable risk.
- Critically assess strengths and weaknesses of a retrospective cohort study design
- Consider factors affecting generalizability

Student Preparatory Instructions

1. **Read:** Learning Objectives, Case, Population Health Context, and Thought Questions for this session
2. **Watch Video:** Attributable Risk and Population Attributable Risk (5 min)
3. **Read:** Key Points: Frequency and Association
4. **Read:** Smith LM et al. The Early Benefits of Human Papillomavirus Vaccination on Cervical Dysplasia and Anogenital Warts. Pediatrics 2015;135(5):e1131.
    - The most important things are to understand the study question, basic study design, and what is the exposure (see Thought questions). Don't worry about the statistical methods or the "regression discontinuity" analysis. Look especially at Tables 1, 2, and 3, and Figure 3, which we will discuss in class.
5. **Complete** the Readiness Assessment Exercise.

Additional (Optional) Resources

1. **Read:** Pimple S, Mishra G, Shastri S. Global strategies for cervical cancer prevention. Curr Opin Obstet Gynecol 2016, 28:4–10.
2. **Read:** Ojha et al. Potential Overestimation of Racial Disparities in Response to the 8-Week Ledipasvir/Sofosbuvir Regimen for Hepatitis C Virus Genotype1 Infection. Gastroenterology 2018,155 (5):1646-7.
3. **Extra practice –** more questions for you to use for practice, with answer keys. Will be posted after class is over.

Session 3 - Interpreting Statistics in Clinical Research

Format: Lecture and Small Group

Session Overview

This session will include a 45-minute lecture followed by a 75 minute small group. We will perform a high-level review of descriptive and inferential statistics, and practice some of these concepts. While in practice most of us do not hand-calculate measures like z-scores and confidence intervals (instead, we use statistical software) we think it is helpful for you to do a few calculations by hand to give you a better sense of what inputs influence these values. At the end of the day, we want you to understand how these measures are applied and what variables will influence their magnitudes. We will walk through exercises together to interpret and critically discuss statistical calculations and tables from peer reviewed publications on autism spectrum disorder.

Learning Objectives

At the end of this session, students will be able to:

- Calculate and interpret descriptive statistics, including mean, median and z-score.
- Understand the underpinnings of statistical inference, including sampling, hypothesis testing, and risks of type 1 and 2 errors.
- Interpret one and two sample t-tests for continuous, normally distributed data.*
- Interpret chi squared tests of categorical data.*
- Calculate and interpret confidence intervals and understand their usefulness in portraying both the point estimate and level of certainty for a result.
- Interpret statistical results presented in tables of published studies.
- *Note that we will NOT expect you to calculate a one or two sample t-test or Chi-square test for the midterm or final. However, you will need to be able to calculate a 95% CI when given a mean and standard deviation or a proportion.

Student Preparatory Instructions

1. **Read:** Learning Objectives, Case, Population Health Context, and Thought Questions for this session
2. **Read:** Key Points: Descriptive Statistics
3. **Read:** Key Points: Sampling Distribution
    a. We will first learn about sampling distribution and why it matters for inferential statistics. This will include learning how to calculate a confidence interval of a mean.
4. **Read:** Key Points: Interpreting Statistics in Clinical Research
5. **Read:** Primer on Type 1 and 2 Errors. Effective Clinical Practice. 2001. (1.5 pages)
6. **Read:** Primer on Statistical Significance and P values. Effective Clinical Practice. 2001. (2 pages)
7. **Read:** Primer on 95% Confidence Intervals. Effective Clinical Practice. 2001. (2.5 pages)
8. **Inferential Statistics Videos** – Watch the following four videos. We will first learn about sampling distribution and why it matters for inferential statistics. This will include learning how to calculate a confidence interval of a mean. We will then learn about hypothesis testing and learn how to calculate one-sample and two-sample t-tests.
    a. **Watch Video:** Sampling Distribution (9 min) and complete this Tutorial (~5 min)

b. **Watch Video:** Confidence Intervals (6 min)
c. **Watch Video:** Hypothesis Testing (One-sample t-test) and p-Values (7:30 min)
d. **Watch Video:** Two-Sample t-test (15 min). This video will walk you through the theory and math behind a two-sample t-test.

9. **Complete** the Readiness Assessment Exercise
10. Be prepared to discuss the following **Thought Questions:**
    a. In what clinical context would it be useful to calculate a Z-score? How do you calculate a Z-score?
    b. Why do we use sampling distributions?
    c. What is the relationship between the standard deviation (SD) and standard error (SE) of a sample? Between the SE and the sample size?
    d. What is a p-value? Is $p<0.05$ an appropriate cut-off for statistical significance?
    e. Which method is preferable for demonstrating statistical significance – a p-value or 95% confidence interval?
    f. In what context would it be useful to calculate a two sample t-test vs. a chi-square test?

Additional (Optional) Resources

1. **Statistics Videos –** for statistics beginners, or those who need to brush up on the measures in the Descriptive Statistics Key Points.
    a. **Watch Video:** Descriptive Statistics (9 min)
        a. This video defines median, mean, mode, variance, and standard deviation.
    b. **Watch Video:** Z-score (5 min)
2. **Inferential Statistics Video**
    a. **Watch Video:** Chi-square square tests (25 min)
        i. **F**or those of you who want a more in-depth explanation of what goes into chi-square test calculations than we will have time to provide in the class

Session 4 - Bias, Confounding and Effect Modification

Format: Lecture and Small group

Session Overview

Up to now in this course, we have mainly been considering a single exposure and a single outcome. However, as you have seen and experienced, most clinical questions involve a complex set of potential relationships between many exposures (including personal characteristics, behaviors, health conditions, and many other factors) and an outcome. Typically, these many variables may not only influence the outcome of interest but also interrelate with each other. Teasing apart and quantifying these interrelationships is important for understanding risk.

This session will be a lecture followed by a small group in which we will do some exercises based on the lecture, followed by a journal club discussion of the Magnani et al. paper. We will begin to apply the epidemiologic concepts of bias, confounding, and effect modification in the context of risk factors for coronary heart disease. In the small groups, we will apply this knowledge to investigate racial differences in the associations of atrial fibrillation and its adverse outcomes (stroke, heart failure, coronary heart disease, and mortality). We know this material can be a lot to digest on the first pass, but do not fear, we will be coming back to these same concepts in the next 2 sessions and will discuss the Magnani et al. paper again in Session 6.

Learning Objectives

At the end of this session, in combination with Sessions 5 and 6, students will be able to:

- Apply their knowledge of bias, confounding and effect modification to interpret results presented in a published manuscript
- Calculate and interpret results from stratified analyses
- Assess exposure and outcome measures for possible sources of bias
- Discuss advantages and disadvantages of cohort studies examining cardiovascular disease outcomes
- Discuss the measurement and inclusion of race as a variable in population health research

Student Preparatory Instructions

1. **Read:** Learning Objectives, Case, and Population Health Context
2. **Watch Video**: Sketchy EBM 'What is Bias?' (4 min)
3. **Watch Video:** Clearing up Confounding (4 min)
4. **Watch Video:** Effect Modification (5 min)
5. **Read**: Key Points: Bias, Confounding, and Effect Modification
6. **Read:** Magnani et al., Racial Differences in Atrial Fibrillation-Related Cardiovascular Disease and Mortality: The Atherosclerosis Risk in Communities (ARIC) Study). JAMA Cardiol. 2016;1(4):433-441. ** Focus especially on the Introduction, Tables 1 and 2, and Discussion for today. In Session 6 we will be discussing the statistical methods, Table 3 and the Figures, so don't worry about those for today.
7. **Read:** Stamos TD and Darbar D. The "Double" Paradox of Atrial Fibrillation in Black Individuals. JAMA Cardiol 2016. 377-378.

8. **Read:** Kaplan JB, Bennett T. Use of race and ethnicity in biomedical publication [published correction appears in JAMA. 2004 Sep 1;292(9):1022]. JAMA. 2003;289(20):2709-2716. doi:10.1001/jama.289.20.2709
9. **Complete** the Readiness Assessment Exercise

Additional (Optional) Resources

1. **Read:** "On Racism: A New Standard For Publishing On Racial Health Inequities, " Health Affairs Blog, July 2, 2020. DOI: 10.1377/hblog20200630.939347
2. **Read:** "Many Medical Decision Tools Disadvantage Black Patients"
3. **Read**: Kuban KCK, Boynton RJ, Heeren T, O'Shea TM. A Consideration of Racism in Pediatric Epidemiologic Studies. J Pediatr. 2021 Aug 11:S0022-3476(21)00759-9.
4. **Read**: Shapiro S. Causation, bias and confounding: a hitchhiker's guide to the epidemiological galaxy. Part 2. Principles of causality in epidemiological research: confounding, effect modification and strength of association. J Fam Plann Reprod Health Care 2008; 34(3):185-190. Read sections 2a and 2b (pages 185-187). Feel free to read through the end if you have time and/or interest.
5. **Read**: Norton EC et al. Odds Ratios—Current Best Practice and Use. JAMA, 2018.

Session 5- Randomized Controlled Trials and Power

Format: Lecture and Small Group

Session Overview

This session will consist of a 45-minute lecture followed by a 75-minute small group. In the lecture, Westyn Branch-Elliman, MD, MPH a physician investigator and infectious disease specialist at the VA, will discuss her experience designing and conducting randomized controlled trials (RCT) for COVID-19 therapeutics during the pandemic.

In the small group that follows, we will conduct a faculty-led journal club on the RCT for Moderna's COVID-19 vaccine. You will focus on the choice of study outcomes and study population, key RCT design features (such as random assignment and blinding) and analytic approach and discuss how these elements impact the internal validity and external validity (i.e., generalizability) of the study.

Learning Objectives

At the end of this session, students will be able to**:**

- Consider design elements of an RCT, and the strengths and weaknesses of this design.
- Apply their knowledge of bias, confounding, and effect modification and how these are addressed in the RCT study design
- Understand how selection of study population might affect both internal validity and study generalizability.
- Understand the interplay between type 1 and type 2 error, and how they relate to sample size when planning a study, and interpretation of results (particularly of "negative" studies).
- Use an online calculator to estimate required sample size and see how it changes under different assumptions regarding desired power and risk of type 1 error.
- Understand (and be able to calculate) Number Needed to Treat (NNT)

Student Preparatory Instructions

1. **Read:** Learning Objectives and Population Health Perspective
2. **Read:** Key Points: Randomized Controlled Trials
3. **Read:** Key Points: Power and Sample Size Estimation
4. **Read:** Baden LR, et al. Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. NEJM. 2021;384(5): 403-415. Plus pages 44-45 of study protocol (on Inclusion/Exclusion criteria).
    a. This is the article we will discuss in detail during small group – please read it thoroughly!
5. **Read:** Simon GE, et al. Evidence from Pragmatic Trials during Routine Care – Slouching toward a Learning Health System. NEJM. 2020;382(16):1488-1491).
    a. The concepts of pragmatic trials and the learning health care system are relevant to today's lecture.
6. **Read:** Spong CY. Improving Public Health Requires Inclusion of Underrepresented Populations in Research. JAMA. 2018;319(4): 337-338.
7. **Watch Video:** Introduction to Experiment Design (10 min)
8. **Watch Video:** Introduction to Power in Significance Tests (10 min)
9. **Complete** the Readiness Assessment Exercise

Additional (Optional) Resources

1. **Watch Video:** RCT vs. Cohort study video (11 min)
2. **Read:** Asch DA, et al. Rethinking ethical oversight in the era of the learning health system. Healthcare. 2020. 8(4):100462. https://pubmed.ncbi.nlm.nih.gov/32992106/
3. **Read:** Chambers. et al. Convergence of Implementation Science, Precision Medicine, and the Learning Health Care System. JAMA. 2016;315(18):1941-1942.
    a. While this article focuses on the topic of genomics, and not COVID-19, the concepts of implementation science and the learning health care system are relevant to today's lecture
4. **Read:** Schultz et al. CONSORT 2010 Statement: Updated Guidelines for Reporting Parallel Group Randomized Trials. BMJ 2010;340:c332.
5. **Read:** P. Parfrey and B. Barrett (eds.), Methods of Molecular Biology, Clinical Epidemiology, vol.473. Chapter 5: Design of Randomized Controlled Trials.
    a. This chapter covers similar but more in-depth content as the prep video

Session 6 - Multivariable Modeling

Format: Lecture and Small Group

Session Overview

As you have seen, most clinical questions involve a complex set of potential relationships between exposures and an outcome. Typically, many exposures can influence each other and the outcome of interest and teasing apart and quantifying this relationship is important for understanding risk.

This session will be a ~40-minute lecture followed by small groups. In the lecture, we will learn about multivariable modeling, a set of statistical approaches that allow us to determine the individual contributions of multiple variables to a single outcome, with cases focused on heart disease. We will review three types of multivariable models: linear regression (for continuous outcomes), logistic regression (for binary outcomes) and Cox proportional hazard models (for survival analysis).

Then we will apply our knowledge of modelling, in combination with what we know about bias, confounding, and effect modification, to interpret some published results.

Learning Objectives

At the end of session, students will be able to:

- Understand why multivariable models are used in clinical and population health research.
- Know which type of model – linear regression, logistic regression, or proportional hazards regression - is most appropriate given the study question and outcome variable of interest.
- In research studies, identity the type of model used and interpret the parameter estimates (i.e., beta coefficients), their associated p-values and confidence intervals.
- Understand how to use multivariable models to identify confounding and effect modification

Student Preparatory Instructions

1. **Read:** Learning Objectives, Case, and Population Health Context
2. **Watch Video:** Simple Regression (8 min):
3. **Watch Video:** Interpreting Correlation Coefficient and Simple Linear Regression – Examples from Medical Literature (14 min)
4. **Read**: Key Points: Multivariable Modelling
5. **Read:** Magnani et al., Racial Differences in Atrial Fibrillation-Related Cardiovascular Disease and Mortality: The Atherosclerosis Risk in Communities (ARIC) Study). JAMA Cardiol. 2016;1(4):433-441. (Same paper as session 4 – focus now on the figures and Table 3)
6. **Complete** the Readiness Assessment Exercise

Additional (Optional) Resources

1. **Read:** Katz MH. Multivariable analysis: a primer for readers of medical research. Ann Intern Med. 2003;138(8):644-650. doi:10.7326/0003-4819-138-8-200304150-00012
2. **Watch Video**: "Measuring Racial/Ethnic Health Care Disparities" by Ben Cook. (Slides here). This video provides details on how to measure health care disparities using multivariable regression.

Session 7 - Diagnostic Testing and Screening I

Format: Lecture and Small Group

Session Overview

In today's session, we will learn about test characteristics (sensitivity and specificity) and the predictive value of test results. We will have a 45-minute lecture on COVID-19 testing followed by a 75-minute small group session to walk through a clinical case on pulmonary embolism.

Learning Objectives

At the end of the session, students will be able to:

- Understand that diagnostic tests provide information on the probability of disease.
- Calculate the specificity and sensitivity of a diagnostic or screening test.
- Understand how to use the results of a diagnostic test to revise pre-test probabilities and calculate post-test probabilities using two approaches:
    - Positive predictive value and negative predictive value
    - Likelihood ratios

Student Preparatory Instructions

1. **Read:** Learning Objectives, Case, Population Health Context, and Thought Questions
2. **Read:** Guglielmi, G. Fast Coronavirus Tests are Coming. Nature. September 24, 2020. (Note: This is an older article that came out before rapid antigen tests were approved for emergency use in the U.S., but it does a good job of explaining how rapid antigen tests work and how they differ from a PCR.)
3. **Read:** Fardy. Chapter 7. Evaluation of Diagnostic Tests. In Clinical Epidemiology: Practice and Methods. P. Parfrey and B. Barrett (eds.), Methods of Molecular Biology, Clinical Epidemiology, vol. 473. 2009. Humana Press. (9 pages)
4. **Read:** Key Points: Diagnostic Testing and Screening
5. **Watch Video:** Test Characteristics (10 min)
6. **Watch Video:** Likelihood Ratios (10 min)
7. **Complete** the Readiness Assessment Exercise

Additional (Optional) Resources

1. **Read:** Diagnostic test studies: assessment and critical appraisal. BMJ Best Practices.
2. **Read:** Grimes DA, Schulz KF. Refining clinical diagnosis with likelihood ratios. Lancet 2005; 365: 1500-1505.
3. **Read:** Rathbun S. The Surgeon General's Call to Action to Prevent Deep Vein Thrombosis and Pulmonary Embolism. Circulation. 2009;119:e480-e482.

Session 8 - Diagnostic Testing and Screening II

Format: Small Group

Session Overview

In this small group session, we will practice applying the principles of screening and characteristics of diagnostic and screening tests, which were introduced in Session 7, to the example of diabetes mellitus screening.

Learning Objectives

At the end of this session, students will be able to:

- Compute predictive values (positive and negative)
- Employ likelihood ratios to revise pre-test probabilities and odds into post-test probabilities and odds
- Appreciate how diagnostic cut-offs are determined using ROC curves
- Understand biases associated with evaluating whether screening impacts mortality and other health outcomes
- Apply concepts of shared decision making to management of pre-diabetes

Student Preparatory Instructions

1. **Read:** Session Learning Objectives, Case, and Population Health Context
2. **Read:** Dans LF, Silvestre MAA, Dans AL. Trade-off between benefit and harm is crucial in health screening recommendations. Part I: General Principles. Journal of Clinical Epidemiology 2011. 64(3):231-9.
    a. *Note: At the bottom of page 235, the authors cite USPSTF recommendations for lipid profile screening for primary prevention in adults with 3 or more risk factors for atherosclerosis. Updated guidelines published November 2016 (and which have understandably created debate) recommend universal lipid screening in adults aged 40 to 75 years.
3. **Read:** Feldman, AL, et al. Screening for type 2 diabetes: do screen-detected cases fare better? *Diabetologia*. 2017;60:2200-2209.
    a. Note: Also **read** the accompanying document "Screening for type 2 diabetes: do screen-detected cases fare better? – Explained" which explains the exposure groupings in a simpler framework.
4. **Watch Video:** Lead and Length Time Bias (4 mins)
5. **Read**: Breu AC. The Inaccuracy of Accuracy. South Med J. 2018 Mar;111(3):166-167.
6. **Read:** Key Points: Diagnosis and Screening
7. **Complete** the Readiness Assessment Exercise
8. **Consider:** Thought Questions for discussion in small groups

Additional (Optional) Resources

1. **Read:** Diabetes Prevention Program Research Group. Reduction in the Incidence of Type 2 Diabetes with Lifestyle Intervention or Metformin. *N Engl J Med.* 2002;346(6):393-403.

2. **Read:** CDC website with resources for diabetes prevention programs: https://www.cdc.gov/diabetes/prevention/index.html
3. **Read:** The International Expert Committee. International Expert Committee Report on the Role of A1c Assay in the Diagnosis of Diabetes. *Diabetes Care*. 2009;32(7):1327-1334.