Homework 1: 50pts

Assigned: August 30, 2021

Due: September 6, 2021, 8:30am

1. (3pts) A researcher is interested in investigating risk factors – specifically intravenous drug use – for new HIV infections among some population of interest. Would an experimental study or an observational study be more appropriate in this setting? Why?

Observational study, because we are comparing the two groups based on the outcome, looking retrospectively.

1. (7pts) A study was designed at Duke to compare body mass index (BMI) among those with prediabetes to BMI among those with glucose control (i.e., no prediabetes or diabetes). 200 individuals with prediabetes and 200 individuals with glucose control were identified through electronic health records (EHR data) and their last available BMI measurement was recorded.
2. (3pts) Out of “BMI” and “prediabetes status”, which one is the primary exposure (or risk factor) and which is the primary outcome? Why?

Prediabetes is the primary exposure, and BMI is the primary outcome. This is because the two groups we are comparing are the prediabetes and the glucose control. The measured outcome is the BMI.

1. (2pts) Based on your answers in (a) is this a cohort, case-control, or cross-sectional study? Explain why.

It is a cross-sectional cohort study. This is because we are splitting the two comparison groups on the exposure factor, not the ‘disease’ .

1. (2pts) Can the prevalence of prediabetes in the population of interest be estimated? Why or why not?

No it cannot, because we are sampling based on the exposure (prediabetes or not) and deliberately seeking out 200 individuals with this exposure and thus we don’t know how many individuals there are organically in a SRS of the population.

1. (4pts) A researcher is developing a novel medication reminder app that also tracks the user’s prescriptions and sends a reminder when it’s time for a refill. He would like your help to design an experimental study that shows his app improves prescription refill adherence better than having no app at all. He is worried that older people may not be as tech savvy and may not use the app correctly, thus biasing the results of the study. So he wants to assign people under 50 to use the app and people over 50 to having no app at all. Briefly (1-2 sentences) describe a possible consequence of this action. Give an example (1-2 sentences) of a more appropriate way to address the researcher’s concern.

A possible consequence of this study design would be if you split on age, the bias would be accentuated. You would not be able to tell if the observed difference in the groups is due to the age difference or the treatment vs no treatment.

A more appropriate way to do this would be to have 4 different groups. We can still split into two groups at the age of 50, but within each group we should have a treatment (app) and no treatment (no app) group. This way we are comparing similar individuals.

1. (6pts) Below is a summary of an article[[1]](#footnote-1):

Effective hypertension self-management interventions are needed for socially disadvantaged African American individuals, who have poorer blood pressure control compared to others. We studied the incremental effectiveness of contextually adapted hypertension self-management interventions among these individuals using a randomized comparative effectiveness trial. One hundred fifty-nine participants were recruited from an academically affiliated community-based primary care clinic in East Baltimore, Maryland. We identified potentially eligible patients by screening clinic electronic health records. Eligible patients were English-speaking adults aged 18 or greater who received care in the clinic, self-identified themselves as African American, and had uncontrolled hypertension (at least two measures of systolic BP (SBP) ≥ 140 mmHg or diastolic BP (DBP) ≥ 90 mmHg obtained at the clinic within the 6 months prior to screening and recruitment). We excluded patients from participating if they were pregnant for ethical, safety, and compliance reasons. Participants were randomly assigned to receive (1) a community health worker ("CHW") intervention; (2) the CHW plus additional training in shared decision-making skills ("DoMyPART"); or (3) the CHW plus additional training in self-management problem-solving ("Problem Solving"). We assessed study group differences in BP control (defined as systolic BP < 140 mmHg and diastolic BP < 90 mmHg), over 12 months using generalized linear mixed models.

1. (2pts) Is this a prospective or retrospective study? Justify your answer.

Prospective; they are starting at the beginning and then looking at it over the 12-month period.

1. (2pts) Describe the population from which the sample was drawn.

African American individuals who have poor blood pressure control.

1. (2pts) What is the primary outcome of the study? What is the primary exposure?

The primary outcome of the study is BP control, and the exposure is uncontrolled hypertension (defined in paragraph).

1. (5pts) Below is a summary of an article[[2]](#footnote-2):

25-hydroxyvitamin D [25(OH)D] is lower in Black compared with white Americans but is not consistently associated with outcomes in this group, possibly due to genetic and other biological differences. We studied associations between baseline 25(OH)D with (i) annual rate of estimated glomerular filtration rate (eGFR) decline and (ii) incident chronic kidney disease (CKD) in a secondary analysis of data from the Jackson Heart Study, a prospective cohort of Black Americans. Plasma 25(OH)D levels were corrected for monthly variation in sunlight exposure using the residual method. We used adjusted generalized linear models to evaluate outcomes and assessed potential effect modification by diabetes mellitus, vitamin D binding protein genotype, obesity, dietary sodium intake, and use of renin-angiotensin-aldosterone system inhibitors.

1. (2pts) The study described above is retrospective. Explain why.

The data used was already collected, despite the Jackson Heart study being prospective, they used the data from it in another separate study.

1. (3pts) Identify the primary outcome(s) and exposure(s) of the study.

The primary outcome is 25(OH)D level. Exposures are annual rate of estimated glomerular filtration rate decline and incident chronic kidney disease.

1. (18pts) Suppose the true overall prevalence of a disease in a population is 18% and the prevalence of a specific risk factor for the disease is 1%. The prevalence of the disease among those with the risk factor is 37% and the prevalence among those without the risk factor is 17.8%. Researchers are interested in studying the association between the risk factor and the disease. (In reality, the true proportions given above would be unknown).
2. (3pts) What sampling technique (outcome-based, exposure-based, both, or neither) should the researchers use in designing a study to assess the association? Justify your answer. What is the “textbook” name of this type of study?

Exposure based sampling because what we want to look at is the association between the risk factor and disease, so if we sample people with risk factor and then without, we will have sufficient people with the disease in each group to compare. This is a cohort study.

Suppose 150 individuals were enrolled in a cross-sectional study based on the sampling technique specified in part (a). Of 33 people with the disease, 11 had the risk factor. Of 117 people without the disease, 19 had the risk factor.

1. (6pts) Estimate the following probabilities: , , , , , and . Do this by filling in the table below, and calculating appropriate proportions.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Disease | |  |
| Risk Factor | NO | YES | Total |
| NO | 98 | 22 | 120 |
| YES | 19 | 11 | 30 |
| Total | 117 | 33 | 150 |

P(D|E) = .367

P(D|E-) = .183

P(E|D) = .333

P(E|D-) = .162

P(D) = .22

P(E) = .2

1. (4pts) Which of the estimated probabilities in part (b) are correct estimates of their respective population analogues and which are incorrect estimates? Justify your answer.

All of the estimates except for the P(D) appear to be good estimates. We are told that the true P(D) = .01, but in our sample it came out to be .22. All other estimates are reasonably close to what is given in the question.

1. (5pts) Let and be two events. Bayes’ theorem states that

(You will learn this in BIOS 701.) Use Bayes’ theorem to calculate the true using the percentages given at the beginning of the problem. Similarly, use Bayes’ theorem to calculate . Compare the population values to the estimates in part (c).

P(E|D) = [P(D|E)P(E)]/[P(D)] 🡪 [.37\*.01]/[.18] = .021

P(E|D-) = [P(D-|E)P(E)]/[P(D-)] 🡪 [.63\*.01]/[.82] = .008

These values are very different from the population to the sample.

(2pts) FINAL PROJECT

Read Section 1 and 2 of the SAP on Sakai. Critically review Sections 3.1-3.3.

Write out the research hypotheses for Aims 2-4 in terms of the primary exposures and outcomes described in Sections 3.1-3.3. Specify the direction you think the association will go, based on the overview in Section 1. (I’ve partly done Aim 1 below.)

Research Hypothesis 1: Higher levels of serum K is associated with lower risk of CKD.

Research Hypothesis 2: Higher levels of dietary K is associated with lower risk of CKD.

Research Hypothesis 3: Higher levels of serum K is associated with lower risk of kidney function.

Research Hypothesis 4: Higher levels of dietary K is associated with lower risk of kidney function.

1. Boulware, et al. (2020), Hypertension Self-management in Socially Disadvantaged African Americans: the Achieving Blood Pressure Control Together (ACT) Randomized Comparative Effectiveness Trial, *Journal of General Internal Medicine*, 35, 142-152. <https://doi.org/10.1007/s11606-019-05396-7> [↑](#footnote-ref-1)
2. Lunyera J, et al. (2019), Modifiers of Plasma 25-Hydroxyvitamin D and Chronic Kidney Disease Outcomes in Black Americans: The Jackson Heart Study, *The Journal of Clinical Endocrinology & Metabolism*, 104, 2267-2276. <https://doi.org/10.1210/jc.2018-01747> [↑](#footnote-ref-2)