Mike Gruzynski
W241 – 2
Essay 1 - Final Draft
Link to article: https://www.eurekalert.org/pub_releases/2017-07/uosc-dcc070517.php

      The interesting conclusion of this article is that people who drink more coffee (could) live longer lives and have lower risk of disease. The article mentions that people who consumed one cup of coffee a day were 12 percent less likely to die compared to those who didn't drink coffee. The claim was even stronger with people who drank two to three cups of coffee; who had a 18 percent reduced chance of dying in comparison to non-coffee drinkers. This article discusses this study is from the National Cancer Institute. The data is from a multiethnic cohort study with University of Hawaii Cancer Center in conjunction with the Keck School of Medicine of USC. The study had more than 185,855 participants among multiple ethnicities including African-Americans (17%), Native Hawaiians (7%), Japanese-Americans (29%), Latinos (22%) and whites (25%) with ages ranging from 45 to 75 at the start of the study. The participants in the study answered questionnaires about diet, lifestyle, and family medical history. The questionnaire was distributed every five years with an average length of the study being at 16 years and included questions that reflected likert-scale like questions and whether the coffee type was regular or decaf. The data was adjusted for age, sex, ethnicity, smoking habits, education, preexisting disease, amount of physical exercise and alcohol consumption, although the actual mechanism for the variable adjustments is not explained in detail.

      The proposed causal effect should not be accepted without some scrutiny. There are a number of instances that could lead to an overestimation of the effects of drinking coffee and living longer lives. This study was an observational study with a lot of data points, however the study should not be accepted as fact. There are countless number of omitted variable bias among other biases inside this study. It is known through this study that the subjects were binned or grouped by their ethnicities and the data was adjusted for a long list of characteristics (shown above in the end of the first paragraph).

      One subject that the observational study did not touch on is income level or quality of life which is a huge driver for overall health and longevity of an individual. The average cup of coffee according to usnews.com report [1] is $2.70 purchased at a business. If an individual purchases one cup of coffee every day of the year the total cost would be $985.5. Now scale that to two, three and four cups of coffee and the total cost of coffee purchased from a business is respectively $1971, $2956.5, and $3942. The costs start to add up, and looking at the mean household income from the US census is 56,516 [2]. So coffee purchase alone can range from ~2 percent all the way up to ~7 percent of total income if a participant is at the national average income of the United States. That is just coffee alone, not counting additional costs of taxes, living, transportation, food, and other luxury items. This points to the theme that coffee is a luxury item that can be afforded by those who have more money to spend on the good. In addition, wealth can also lead to more leisure time, more stress reducing activities, and avoidance poor quality of life situations. Another big factor that was an omitted variable from the study for data adjustment was type of diet for each individual. Another article which studied their own coffee and longevity observational study, had a statement about the diet be consistent around a Mediterranean style of diet for as a requirement for the observational study. This is closely related to income level as well because people with more money can afford higher quality food and drink which are better overall for consumption vs cheaper canned or frozen type foods which yield less efficient foods. In addition, intrinsic metabolic response is not taken account for in this study. Researching around other coffee studies, showed the correlations of consumption based on health for participants categorized as "high metabolic response" or "low metabolic response", or in other words how well a participant can process what goes into the body. The results from the other studies show diverging responses based on the individual's intrinsic metabolic capability [3]. Also, location can play a large role in the drinking habits of

the the subjects. For example the drinking habits of people who live in Seattle (a coffee drinking city) on average probably are different than drinking habits of coffee from people who live in Tokyo (a tea drinking city). There is an additional cultural and geographical variable that will lead to omitted variable bias if not properly handled with experimental mechanisms (i.e random selection).

In addition to potential omitted variable bias present in observational studies, the study could have biases associated with the architecture of the study. Since there is a survey questionnaire aspect to the study, there could be response bias present (the participant's answers are not fully accurate or truthful based on either foggy memory or no incentive to be honest about the results). Another type of study biasness could be convenience sampling. There is no note of where exactly these participants came from. They could have come from similar regions of the US or from one type of clinic or hospital. Convenience sampling is a non-probabilistic sampling technique where subjects are chosen because of their proximity (or ease of interaction) with the study. This could lead to clustering of data inputs or outputs. For example, compare the West coast to the southern part of US. There are very different cultural and political differences, along with socio-economical and standard of life between the two groups, even though they are all categorized as Americans. One set of group may be more equipped with better standard of living characteristics (such as health, disposable income, leisure time, stress level, family life, etc) than the other. In addition, not all coffee is the same. A latte and a simple drip coffee have very different dietary characteristics associated with them (drip = 0-5 calories and latte = 300 - 1000 calories). There is no mention of what type of coffee was purchased by the participants.

The main assumptions, which are hard to believe is that there is no clustering of data from selection biases and that we are comparing on average apples to apples. Because of the assumption that we have selection biases, we have no idea exactly where these observation participants come from. It is hard to perform a "ceteris paribus" study on coffee effects if we have no idea of the overall standard of life and diet (among other omitted variables of the study). This is just a very small sample from the overall population and have no information what the skew of the data looks for the adjusted variables or characteristics of the participants (i.e. age, sex, ethnicity, smoking habits, education, preexisting disease, amount of physical exercise and alcohol consumption), or the possible skew of very large drivers of omitted variables (i.e standard of living, diet, etc.). The data ideally should be normally distributed (or equally spaced explanatory variable) in order to get un biased statistics from the data.

The study notes no difference between decaf and regular coffee drinks. The claim that individuals who drink more coffee (could) live longer lives and have lower risk of disease needs to be experimentally tested. Professional's have an hypothesis that coffee contains a lot of antioxidants and phenolic compounds that play an important role reducing cancer and improving duration of life. An experiment worth conducting is to select a random sample of patients (this will help smear out highly skewed data) and to create four groups with varying levels of amount of daily intake of phenolic compounds and antioxidants (None, low, medium and high intake of concentrated coffee ingredients). The study will go on for 30 years with yearly checkups for panel test data retrieval that monitor health statistics like weight, heart health, vitals,and among other tracked values (objective responses instead of subjective survey responses). The patients will be able to still have caffeine intake (since caffeine vs no caffeine has little effect on the subject) and coffee intake (must be documented accurately and ideally no additional coffee intake is experienced). This would be an ideal experiment, but probably not feasible in reality since the amount of money and time investment in both the experiment study and the participants as well as the ability for the participants to stick to the experimental procedure for 30 years.

References:
1. https://www.usnews.com/news/blogs/data-mine/2015/09/29/americans-pay-an-average-270-for-coffee-while-tipping-20-percent
2. http://www.kff.org/other/state-indicator/median-annual-income
3. http://www.sandiegouniontribune.com/business/biotech/sd-me-coffee-longevity-20170710-story.html