**EPI : Homework #1**

**Question 1:** (source: <https://biolincc.nhlbi.nih.gov/teaching/>)

The Framingham Heart Study (FHS) is a prospective cohort study that began in 1948. Participants have been examined biennially, and all participants are continuously followed through regular surveillance. Clinic examination data have included cardiovascular disease risk factors and markers of disease such as blood pressure, smoking history, health behaviors, blood chemistry, lung function, echocardiography, and medication use. Through regular surveillance of area hospitals, participant contact, and death certificates, the FHS reviews and adjudicates events for the occurrence of cardiovascular outcomes.

The NHLBI teaching dataset is a subset of the data from the FHS and includes data on 4,434 participants. Participant clinic data were collected during 3 examination periods, approximately 6 years apart, from roughly 1956 to 1968. Each participant was followed for up to 24 years. Although the teaching dataset contains Framingham data “as collected”, specific methods were used to ensure an anonymous dataset that protects patient confidentiality; therefore, this dataset is inappropriate for publication purposes.

This question uses the following variables from the teaching dataset:

CURSMOKE = 1 if current smoker at the 1956 exam

= 0 otherwise

PREVCHD = 1 if diagnosed with coronary heart disease (CHD) before the 1956 exam

= 0 otherwise

DIABETES = 1 if diabetic according to criteria at the 1956 exam (either treated for diabetes or casual glucose ≥ 200 mg/dL)

= 0 otherwise

DEATH = 1 if died (from any cause) during the 24 years of follow-up after the 1956 exam

= 0 otherwise

TIMEDTH = follow-up time (in years) from the 1956 exam to:

1. time of death for participants who died during the 24-year follow-up period
2. end of study (24 years) for participants who did not die during the follow-up period

ANYCHD = 1 if developed CHD during the 24 years of follow-up after the 1956 exam

= 0 otherwiseTIMECHD = follow-up time (in years) from the 1956 exam to:

1. time of developing CHD for participants who had documentation of the development CHD during the 24-year follow-up period,
2. time of death for participants who had no documentation of the development of CHD during the follow-up period but who died from another cause,
3. time of last contact for participants who had no documentation of the development of CHD by their last contact and were lost to follow-up, or
4. end of study (24 years) for participants who were followed for 24 years and did not develop CHD or die during the follow-up period.

The following table (Table 1) shows the relationship between current smoking and prevalent CHD in 1956.

**Table 1.**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **PREVCHD** | |  |
| **CURSMOKE** | **1** | **0** | **Total** |
| **1** | 86 | 2095 | 2181 |
| **0** | 108 | 2145 | 2253 |
| **Total** | 194 | 4240 | 4434 |

1. What is the prevalence of CHD at the 1956 exam among all participants?

1. What is the prevalence of CHD at the 1956 exam among current smokers? What is the prevalence among non-smokers?
2. What is the prevalence ratio for CHD, comparing current smokers to non-smokers? Give at least two plausible explanations for this value.

The following table (Table 2) shows the relationship between diabetes at the 1956 exam and mortality (death) during the 24 years of follow-up. Note that there were no losses to follow-up for the outcome of death, because mortality status was obtained for all participants through a search of the National Death Index or through the Social Security Death Index.

**Table 2.**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **DEATH** | |  |
| **DIABETES** | **1** | **0** | **Total participants** |
| **1** | 94 | 27 | 121 |
| **0** | 1456 | 2857 | 4313 |
| **Total** | 1550 | 2884 | 4434 |

The following table (Table 3) shows the total amount of follow-up time in years until death (for those who died) or until the end of the study, i.e., 24 years (for those who survived), among the 4434 participants in Table 2, according to whether they had diabetes in 1956.

**Table 3.**

|  |  |
| --- | --- |
|  | **TIMEDTH** |
| **DIABETES** | **Number person-years** |
| **1** | 1752.48 |
| **0** | 89363.10 |

1. If you wanted to estimate the effect of diabetes on the incidence of death, which of the following measures of association could you calculate validly from the data in Tables 2 and 3? **Calculate** the value for the measure(s) that you can calculate validly, and then briefly **interpret** the meaning of each value in words. If you do not feel that the measure can be calculated validly, then state in one sentence why you do not feel that it is valid.

Risk Ratio:

Odds Ratio:

Rate Ratio:

The following table (Table 4) shows the relationship between diabetes at the 1956 exam and the incidence of coronary heart disease (CHD) during the 24 years of follow-up. Note that this table includes only 4,240 participants, rather than all 4,434 participants in the dataset.

**Table 4.**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **ANYCHD** | |  |
| **DIABETES** | **1** | **0** | **Total Participants** |
| **1** | 55 | 54 | 109 |
| **0** | 991 | 3140 | 4131 |
| **Total** | 1046 | 3194 | 4240 |

1. Why are the 194 participants with prevalent CHD at the 1956 exam not included in this table?

The following table (Table 5) shows the total amount of follow-up time in years until the development of CHD (for those who developed CHD during the follow-up period) or until death (for those who died of causes other than CHD), last contact (for those who were lost to follow-up), or the end of follow-up (for others), among the 4,240 participants in Table 4.

**Table 5.**

|  |  |
| --- | --- |
|  | **TIMECHD** |
| **DIABETES** | **Number person-years** |
| **1** | 1443.30 |
| **0** | 79481.85 |

1. If you wanted to estimate the effect of diabetes on the incidence of CHD, which of the following measures of association could you calculate validly from the data in Tables 4 and 5? **Calculate** the value for the measure(s) that you can calculate validly, and then briefly **interpret** the meaning of each value in words. If you do not feel that the measure can be calculated validly, then state in one sentence why you do not feel that it is valid.

Risk Ratio:

Odds Ratio:

Rate Ratio:

**Question 2:** A hypothetical cohort study included 1000 primary care patients with no prior history of diabetes or prediabetes (prediabetes is a condition characterized by elevated blood sugar, but not high enough to be diagnosed as type 2 diabetes). Of these 1000 patients, 650 were classified as overweight or obese (body mass index ≥ 25 kg/m2). All 1000 patients were followed for 2 years; assume that there were no competing risks or losses to follow-up during this period. Among the 650 overweight or obese patients, 60 were diagnosed with prediabetes during the follow-up period; among the other 350 patients, 10 were diagnosed with prediabetes.

1. What is the 2-year Risk Ratio for prediabetes in the patients with overweight or obesity, as compared to the other patients? State the meaning of this value in words.
2. What proportion of prediabetes among the overweight/obese patients is associated with overweight or obesity?
3. What proportion of prediabetes among all of the patients is associated with overweight or obesity?

**Question 3:** Use data from the internet to investigate the relationship between political party (Conservative vs. Labour) and the mortality rate (incidence rate of death) after starting office (until death or 2022 if still alive) for Prime Ministers of the United Kingdom who have served from 1940 until the present. Create and display the dataset by completing the table below.Two of the entries (first and last) are already inserted, but you will need to fill in the others. Below is a suggested reference: <https://en.wikipedia.org/wiki/List_of_Prime_Ministers_of_the_United_Kingdom>

Calculate the mortality rate from starting office for each party and an appropriate measure of association. Summarize your findings in a sentence. (Note: For Prime Ministers who served more than once, their follow-up should start from the year of their first term.)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Prime Minister** | **Party (C or L)** | **Year started** | **Death (1=yes, 0=no)** | **Year of death or 2022 if alive** | **Follow-up (years)** |
| Churchill | C | 1940 | 1 | 1965 | 25 |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
| Johnson | C | 2019 | 0 | 2022 | 3 |