Methods

The data used in this study is from an imputed version of NHANES II study. The exposure in this study is whether the participant’s alcohol intake (drinks/week) measured at the baseline. The outcome is time to death of cancer during the follow-up. Other baseline covariates are measured. Baseline characteristics based on quantiles of alcohol intake is shown in table 1.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Alcohol intake =0 | Alcohol intake(0, 0.5] | Alcohol intake(0.5, 2.0] | Alcohol intake(2.0, ∞ |
| Mean Alcohol, SD |  |  |  |  |
| Age at entry, SD |  |  |  |  |
| BMI |  |  |  |  |
| Grades |  |  |  |  |
| … |  |  |  |  |

Given the cross-sectional nature of the study, the alcohol intake can be perceived as prevalent intakes, which may lead to “prevalent user” problems. Taking this potential bias into consideration, in the main analysis, we chose to conduct Cox proportional hazard models through two approaches. The first approach is to set the time origin to be baseline, and time since entry is set to be the time scale. We conducted crude model, models adjusting for baseline age, and the fully-adjusted model In the fully-adjusted model, we adjust for the confounders which are reasonable common causes of prevalent alcohol intake as well as future risk of death because of cancer. The second approach is to set the age 21 (which is the legal age to buy alcohol in the US) and choose the age as the time-scale. In this approach, we conduct crude model and fully-adjusted model. No model adjusting for baseline age is conducted since we treat the age at entry as the left truncation threshold. The fully-adjusted model now only adjusted for sex, and other SES related covariates, because all the biomarkers, biometrics, and disease status at baseline might be the consequence of the alcohol intake before. The conceptuall framework as well as DAGs are shown in figure 1.

[Figure 1]

The interpretations for the exposure effect are different for the two approaches. The interpretation for the first approach, is the baseline prevalent alcohol intake’s effect on the hazard for time to the event (dying from cancer) since the measured baseline. The interpretation for the second approach, is the alcohol intake’s effects after age 21 on the hazard for the age when dying from cancer.

To explore potential effect modification by sex, we conducted the fully-adjusted models using two method 1) conducting stratified Cox regression by sex 2) including the product term in the fully-adjusted model. Both two methods are utilized in each of the two approach mentioned above, so there are 4 models.

We checked the proportional hazard assumption using Schofield residuals and reported the corresponding chi-squared p-values for the exposure as well as for the whole model.

In the sensitivity analysis, we adopted the first approach which using time since baseline as the time scale, and conducted logistic and Poisson regression as sensitivity analysis.