Methods

The data used in this study is from an imputed version of NHANES II study. The exposure in this study is alcohol intake (drinks/week) measured at the baseline. After exploring the distribution of alcohol intake, we decided to divide the continuous alcohol intake to categorical variables to reduce the influence of outliers. We found about 43.7% of people didn’t take alcohol per week. After excluding people with 0 alcohol intake, the minimum, median and maximum alcohol intake were 0.5, 2.0, and 77.0 per week. Then we divided people into 4 groups based on their alcohol intake: 0 per week, 0-0.5 drinks per week, 0.5-2 drinks per week and >2 drinks per week. The outcome is time to death of cancer during the follow-up, or time to censoring, or time to death from other causes, whichever comes first.

Descriptive statistics are utilized to measure the baseline characteristics in the study dataset. Number and percentages for categorical variables and mean and standard deviation for continuous variables are reported. Through these descriptive statistics, we measured the unbalanced distribution for the covariates between alcohol intake groups, and thus, explore potential confounding in the association between alcohol assumption and cancer mortality.

Given the cross-sectional nature of the study, the alcohol intake can be perceived as prevalent intake, which could induce the “prevalent user” bias. To test the robustness of the findings given this potential bias, in the main analysis, we chose to conduct Cox proportional hazard models through two approaches. The first approach is setting the age 21 (which is the legal age to buy alcohol in the US) to be the time origin and choosing age as the time-scale. In this approach, we conducted crude model and fully-adjusted model. No model adjusting for baseline age was conducted since age was set as the time scale. The fully-adjusted model now only adjusted for sex, and other socio-economic status (SES) related covariates, because all the biomarkers, biometrics, and disease status at baseline might be the consequence of the alcohol intake before. The second approach is setting baseline time as the time origin, 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 adjusted for confounders which are reasonable common causes of prevalent alcohol intake as well as future risk of death because of cancer. Also the confounders were chosen based on the descriptive statistics mentioned above. The conceptual framework as well as DAGs are shown in figure 1.

图示

Description automatically generated

Fig 1. Conceptual framework for Cox proportional hazard models in the primary analysis

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 effects of alcohol consumption 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) including the product term in the fully-adjusted model, 2) conducting stratified Cox regression by sex and adding an interaction term by sex and alcohol intake, 3) conducting Cox regressions in males and female separately (called subgroup analysis). Wald-tests were utilized to test the significance of effect modification.

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 fully-adjusted model non-linearity was also checked first by comparing the category-specific hazard ratio. Then, we fit a natural cubic spline replacing the linear term of alcohol consumption to explore potential non-linear relationship.

In the sensitivity analysis, we adopted the first approach which using time since baseline as the time scale, and conducted logistic and Poisson regression to check the influence of model selection on our findings.

**Results**

**Xinyi**

We assessed the association between categorical alcohol consumption through two approaches using 5 models. Across all the models, no significant associations are found for alcohol consumption 0-0.5 drinks per week, 0.5-2 drinks per week compared to 0 drinks per week. For the category >2 drinks per week, the first approach yields a HR of 1.57 (95% CI, 1.29, 1.91) in the crude model and 1.24 (95% CI, 1.00, 1.54) in the fully-adjusted model. Using the second approach, the magnitude of the association is stronger in the fully-adjust model than in the crude model and in the model adjusting for age assessing the association between continuous alcohol consumption and cancer mortality, all the models yield very weak association with HR’s ranging from 1.01 and 1.04. The fully-adjusted model using the first approach yields marginal significant results (95% CI 1.01, 1.04) but the fully-adjusted model using the first approach yields nonsignificant results (95% CI 0.99, 1.02). However, the estimated association between two approaches using the fully-adjusted model are pretty close

With the findings from the main analysis, we explore the potential modification of sex. First, we compared the estimated HRs comparing >2 drinks per week to 0 drinks per week across males and females. The HR’s estimated in the females are greater than in the males in all of the models, indicating the alcohol consumption’s harmful effects are more pronounced in the females than in the males. Only 1 model using the interaction term through the first approach reported a significant effect modification. Then we treated the alcohol consumption as a continuous variable. Then all the six models yield similar results for males and females, both are very close to the estimates in the overall sample analysis. The effect modification is of no statistical significance.

We checked the proportional hazard assumption in all the fully-adjusted models using Scofield’s residuals. The assumption satisfied for the alcohol consumption and for the overall model in all the models. We also explored potential non-linear relationships by using a natural cubic spline with 3 degrees of freedom. The findings shows that the dose-response curve for the log hazard ratio is near horizontal when the alcohol consumption is less than 5 drinks per week, and becomes linear thereafter.

Table 1 Baseline characteristics of study population by alcohol intake(N=) .

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Alcohol intake per week | | | |  |
|  | 0/week | 0-0.5/week | 0.5-2/week | >2/week | p-value |
|  | N=1052 | N=209 | N=319 | N=530 |
| Sex |  |  |  |  | <0.001 |
| Male | 515 (49.0%) | 110(52.6%) | 205(64.3%) | 407(76.8%) |  |
| Female | 537 (51.0%) | 537(51.0%) | 114(35.7%) | 123(23.2%) |  |
| Mean Age at entry (SD) | 65.2 (7.81) | 64.9 (8.32) | 62.2 (9.43) | 62.1 (9.67) | <0.001 |
| Race |  |  |  |  | <0.001 |
| white | 109 (10.4%) | 21 (10.0%) | 51 (16.0%) | 87 (16.4%) |  |
| black | 388 (36.9%) | 93 (44.5%) | 121(37.9%) | 234(44.2%) |  |
| other | 555 (52.8%) | 95 (45.5%) | 147(46.1%) | 209(39.4%) |  |
| Mean year of schooling (SD) | 9.24 (3.69) | 10.4 (3.42) | 10.3 (3.57) | 10.9 (3.35) | <0.001 |
| Marital status |  |  |  |  | . |
| Married | 699 (66.4%) | 143 (68.4%) | 225 (70.5%) | 379 (71.5%) |  |
| Widowed | 231 (22.0%) | 36 (17.2%) | 49 (15.4%) | 67 (12.6%) |  |
| Divorced | 44 (4.18%) | 12 (5.74%) | 13 (4.08%) | 12 (9.76%) |  |
| Separated | 24 (2.28%) | 7 (3.35%) | 10 (3.13%) | 11 (2.08%) |  |
| Never married | 53 (5.04%) | 9 (4.31%) | 21 (6.58%) | 32 (6.04%) |  |
| Blank | 1 (0.10%) | 2 (0.96%) | 1 (0.31%) | 5 (0.94%) |  |
| Mean BMI (SD) | 26.3 (5.58) | 25.8 (4.51) | 26.1 (5.21) | 25.1 (4.17) | <0.001 |
| Mean smoke per day (SD) | 5.60 (11.8) | 6.74 (11.9) | 9.78 (15.2) | 12.0 (15.0) | <0.001 |
| Size of place | 5.70 (2.57) | 5.07 (2.70) | 4.53 (2.69) | 4.44 (2.69) | <0.001 |
| Standard Metropolitan  Statistical Area |  |  |  |  | <0.001 |
| In central city | 265 (25.2%) | 59 (28.2%) | 105(32.9%) | 186(35.1%) |  |
| Not in central | 201 (19.1%) | 57 (27.3%) | 102(32.0%) | 183(34.5%) |  |
| Not in SMSA | 586 (55.7%) | 93 (44.5%) | 112(35.1%) | 161(30.4%) |  |
| Resides in urban area, Yes | 578 (54.9%) | 578(54.9%) | 235(73.7%) | 387(73.0%) | <0.001 |

Table 2. Association between alcohol intake and hazard of death from cancer in overall NHANES II (N=).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 0/week | 0-0.5/week | 0.5-2/week | >2/week | Continuous | Ptrend |
| First approach, defining the exposure as alcohol intake age 21-entry, time scale as age | | | | | | |
| Crude model | Ref | 1.08 (0.80, 1.46) | 1.11 (0.87, 1.42) | 1.57 (1.29, 1.91) | 1.04 (1.03, 1.05) | <0.001 |
| Age-adjusted | Ref |  |  |  |  |  |
| MV-adjusted | Ref | 1.04 (0.77, 1.41) | 0.97 (0.76, 1.26) | 1.24 (1.00, 1.54) | 1.02 (1.01, 1.04) | 0.087 |
| Second approach, defining the exposure as prevalent alcohol intake at baseline, time scale as years since entry | | | | | | |
| Crude model | Ref | 0.93 (0.69, 1.26) | 0.83 (0.65, 1.06) | 1.15 (0.94, 1.39) | 1.03 (1.02, 1.04) | 0.374 |
| Age-adjusted | Ref | 1.01 (0.81, 1.48) | 1.12 (0.87, 1.43) | 1.59 (1.30, 1.94) | 1.07 (1.06, 1.08) | <0.001 |
| MV-adjusted | Ref | 1.05 (0.78, 1.42) | 0.97 (0.75, 1.25) | 1.24 (1.00, 1.54) | 1.01 (1.01, 1.04) | 0.095 |

Table 3. Associations between alcohol intake and hazard of death from cancer in males and females from NHANES II (N=).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | >2/week vs 0/week | | Continuous | |
| First approach, defining the exposure as alcohol intake age 21-entry, time scale as age | | | | |
| Adding interaction term | | | | |
|  | MV-adjusted | Heterogeneity | MV-adjusted | Heterogeneity |
| Female | 1.05 (0.81, 1.36) | 0.044 | 1.02 (1.01, 1.04) | 0.710 |
| Male | 0.68 (0.37, 1.24) |  | 1.02 (0.97, 1.06) |  |
| Stratified Cox | | | | |
|  | MV-adjusted | Heterogeneity | MV-adjusted | Heterogeneity |
| Female | 1.52 (1.07, 2.18) | 0.120 | 1.03 (0.99, 1.07) | 0.880 |
| Male | 1.08 (0.84, 1.40) |  | 1.02 (1.01, 1.04) |  |
| Subgroup analysis | | | | |
| Female | 1.36 (0.94-1.98) | 0.207 | 1.02 (0.97-1.06) | 0.303 |
| Male | 1.13 (0.87-1.47) |  | 1.03 (1.01-1.04) |  |
| Second approach, defining the exposure as prevalent alcohol intake at baseline, time scale as years since entry | | | | |
| Adding interaction term | | | | |
|  | MV-adjusted | Heterogeneity | MV-adjusted | Heterogeneity |
| Female | 1.60 (1.12-2.29) | 0.054 | 1.03 (0.99-1.07) | 0.757 |
| Male | 1.05 (0.81-1.36) |  | 1.02 (1.01-1.04) |  |
| Stratified Cox | | | | |
|  | MV-adjusted | Heterogeneity | MV-adjusted | Heterogeneity |
| Female | 1.60 (1.12-2.28) | 0.057 | 1.03 (0.99-1.07) | 0.771 |
| Male | 1.06 (0.82-1.37) |  | 1.02 (1.01-1.04) |  |
| Subgroup analysis | | | | |
| Female | 1.37 (0.94-1.99) | 0.218 | 1.01 (0.97-1.06) | 0.710 |
| Male | 1.14 (0.87-1.48) |  | 1.03 (1.01-1.04) |  |

Table 4. Tests for the proportional hazard assumption throughout the models (N=).

|  |  |  |  |
| --- | --- | --- | --- |
|  | Chi-squared statistic | p-value for alcohol consumption | Global p-value |
| First approach, defining the exposure as alcohol intake age 21-entry, time scale as age | | | |
| MV-adjusted categorical model | 7.28 | 0.06 | 0.15 |
| MV-adjusted continuous model | 0.22 | 0.64 | 0.28 |
| Second approach, defining the exposure as prevalent alcohol intake at baseline, time scale as years since entry | | | |
| MV-adjusted categorical model | 3.40 | 0.33 | 0.67 |
| MV-adjusted continuous model | 0.37 | 0.54 | 0.71 |

Fig

图表, 折线图

Description automatically generated