Appendix B. Equations of mixed model in meta-analysis on incidence or mortality risk of pancreatic cancer am and alcohol intake

The general formula for the fully-adjusted mixed model of pooled 37 incidence or mortality studies of PC for drinkers (vs nondrinkers) is as follows:

$$Y_{ij} = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \cdots + \beta_n X_{nij} + u_i + \varepsilon_{ij}$$

where Y_{it} is the natural log-transformed (to stabilize variance and make it more suitable for linear models) and weighted hazard ratio (using the inverse of variance to give more weights to the study estimates with high precision) in the i^{th} study at the j^{th} observation; β_0 is the fixed intercept for study i; β_1 is the coefficient of X_{1ij} which is the natural log-RR for drinkers (vs nondrinkers) and X_{1ij} is the category of drinkers (former drinkers, light – highest drinkers) or alcohol intake in grams per day among current drinkers; β_2 - β_n the coefficient of covariates X_{2iJ} - X_{nij}; u_i is the random effect of the i^{th}study; and ε_{ij} is the error term for j^{th} observation in the i^{th} study.

(i) Basic standard model

Basic standard model (Model I in Tables 3-5 and Tables C3-C5 in Appendix C) deals with skewed distribution by analyzing natural logged RR, weighting by inverse of variance in each estimate to adjust for sampling variability and adjusted for heterogeneity across studies The dependent variable was the natural-log of the RR estimated using the HR of each drinking group in relation to the abstainer category. All risk estimates were weighted by the inverse of the variance (IV) of the natural-log RR. Variance was estimated from reported standard errors or confidence intervals of HRs. This can be represented as follows:

$$Y_{ij} = \beta_0 + \beta_1 X_{1ij} + u_i + \varepsilon_{ij}$$

where Y_{it} is the natural log-transformed (to stabilize variance and make it more suitable for linear models) and weighted HR (using the inverse of variance to give more weights to the study estimates with high precision) in the i^{th} study at the j^{th} observation; β_0 is the fixed intercept for study i; β_1 is the coefficient of X_{1ij} which is the natural log-RR for drinkers (vs nondrinkers) and X_{1ij} is the category of drinkers (former drinkers, light – highest drinkers) or alcohol intake in grams per day among current drinkers; u_i is the random effect of the i^{th}study; and ε_{ij} is the error term for j^{th} observation in the i^{th} study.

(ii) *Covariate-adjusted Models*

Multivariable mixed effects models were used to estimate covariate-adjusted RRs for drinking categories or alcohol intake in grams per day among current drinkers by further

including covariates in the basic mixed models to adjust for their potential confounding effects (see Model II in Tables 3-5 and Tables C3-5 in Appendix C). The general formula of the covariate-adjusted mixed linear regression models is as follows:

$$Y_{ij} = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \cdots + \beta_n X_{nij} + u_i + \varepsilon_{ij}$$

This has the additional terms β_2 - β_n and includes also terms for the coefficient of covariates X_{2iJ} - X_{nij}; u_i is the random effect of the i^{th}study; and ε_{ij} is the error term for j^{th} observation in the i^{th} study.

To be considered for inclusion in models of 37 pooled studies, a covariate had to be theoretically related to the exposure and outcome, be significantly associated with the RRs of PC in bivariate analysis, and lack multicollinearity with other included study-level covariates. Bivariate tests were conducted to test the significance of the relationships between each potential study level covariate and log-adjusted RR estimates of PC incidence/mortality outcomes ($P<0.05$) (Hosmer & Lemeshow, 2000). We tested for multicollinearity of covariates by exploring the correlation matrix using the correlation, the Tolerance and the Variance Inflation Factor (VIF) methods and the covariates were included in the models to adjust for their potential confounding effects when the correlation coefficients of any two covariates were lower than 0.4, no Tolerance values fell below 0.1 and the VIF values were below 2 (Allison, 2012; Schreiber-Gregory, 2017). According to the above criterion, covariates included in multivariable mixed linear regression models (Model II in Tables 3-4) of the pooled 37 studies are (i) mean cohort age of cohorts/follow-up years (cohort age<56 and follow-up=10+ vs others), (ii) former drinker bias strictly defined and less strictly defined (yes or no), (iii) study level differences in exclusion of baseline health conditions (exclusion or not), and (iv) if BMI controlled (yes or no) in individual studies.

Covariate-adjusted models (Model II in Table 3 and Table C3 in Appendix C) further adjust for potential confounding effects of mean age of cohorts/follow-up years (cohort age<56 and follow-up=10+ or others) (X_{2ij}), former drinker bias strictly defined and less strictly defined (X_{3ij}), baseline conditions (exclusion or not) (X_{4ij}), and control body mass index-BMI (yes or not) in individual studies (X_{5ij}). The equation is as follows:

$$Y_{ij} = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \beta_3 X_{3ij} + \beta_4 X_{4ij} + \beta_5 X_{5ij} + u_i + \varepsilon_{ij}$$

(iii) Quality weighted Models

We also used a quality weighting method by giving larger weights to risk estimates in the studies with the mean age of cohorts younger than 56 years (+0.4), no former drinker bias (+0.4) or reduced former drinker bias (+0.2) and follow-up for at least 10 years (+0.1). The

quality weights were incorporated into the sampling weights, i.e., the inverse of variance. Other covariates were also adjusted as in Model II in Table 3 and Table C3 in Appendix C; results are presented in Table 4 and Table C4 in Appendix C (Model II). The equation is as follows:

$$Y_{ij} = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \beta_3 X_{3ij} + u_i + \varepsilon_{ij}$$

(iv) Models stratified by former drinker bias

Analyses of studies stratified by presence or absence former drinker bias were performed. As only three fully met strict criteria for being free from abstainer bias, the criteria were relaxed to allow inclusion of 7 additional studies that made partial attempts to remove former drinker bias (Table 5). This stratification approach was supported empirically by a significant interaction between level of alcohol use and presence/absence of former drinker bias (F-test value$_{(df=266)}$=2.68 and P=0.0215) in the model without any adjustment. In Table 5, the stratified analysis on 10 studies with reduced formed drinker bias and 28 studies with former drinker bias did not include drinker bias variable in Model II. In Nakamura et al (2011), RR for men included in the analysis of 28 studies with former drinker bias and RR for women included in the analysis of 10 studies with reduced former drinker bias. Table C5 presents the stratified analysis on 25 incidence and 13 mortality studies. The mortality risk estimates in Jayasekara et al were included in the analysis of mortality from PC when the RR estimates of mortality in this study were excluded in the pooled analysis.

We also performed an analysis comparing incidence and mortality studies to examine how the RR estimates differ by the outcomes (incidence and mortality) in the included studies (see Table C5 in Appendix C).

The SAS MIXED procedure (SAS Institute Inc., 2013) was used to model the log-transformed and weighted hazard ratios (HRs) of PC incidence or mortality for alcohol use per day to investigate the relationship between the risk of PC and alcohol use. The SAS proc mixed procedure, for example, for modeling the meta-data of PC to estimate the mean RR for drinkers (versus nondrinkers) in pooled 37 studies (279 HR estimates, Model II in Table 3) is as follows:

SAS proc mixed procedure for modeling the meta-data of hazard ratio of pancreatic cancer due to alcohol use

proc mixed data=meta_data method=reml covtest;
class study X_{1ij} X_{2ij} X_{3ij} X_{4ij} X_{5ij};
model RR=X_{1ij} X_{2ij} X_{3ij} X_{4ij} X_{5ij} / s cl;
random intercept / subject=study type=un;
run;

Note: RR: natural-log-transformed and weighted hazard ratio, X_{1ij}: categories of alcohol drinking, covariates

include mean age of cohorts/follow-up years (<10 and 10+) (X_{2ij}), former drinker bias strictly defined and less strictly defined (X_{3ij}), baseline conditions (exclusion or not) (X_{4ij}), and control for BMI (X_{5ij}) in individual studies.

References

Allison, P. (2012). Allison, P. When can you safely ignore multicollinearity? 2012. In (pp. 1-2): Statistical Horizons. Accessed September 10, 2022, https://statisticalhorizons.com/multicollinearity.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression*. John Wiley & Sons, Inc. https://doi.org/10.1002/0471722146

SAS Institute Inc. (2013). The GLIMMIX Procedure. In *SAS/STAT® 13.1 User's Guide*. SAS Institute Inc. . https://doi.org/Available from URL: https://support.sas.com/documentation/onlinedoc/stat/131/glimmix.pdf. Accessible February 15, 2024

Schreiber-Gregory, D. N. (2017). Multicollinearity: what is it, why should we care, and how can it be controlled? The SAS Global Forum 2017 Conference, Orlando, Florida.