Data supplement for Visontay et al., Moderate Alcohol Consumption and Depression: A Marginal Structural Model Approach Promoting Causal Inference. Am J Psychiatry (doi: 10.1176/appi.ajp.22010043)

CONTENTS

1. Measurement occasions

The NLSY-79 introduced a suite of additional measurements to be completed by participants at ages 40 and 50, i.e., at the scheduled wave coinciding with the year when they reach the relevant age. While the CES-D-SF was initially tied to year-based measurement i.e., the 1992 and 1994 surveys, it was then moved to the suite of health measurements at ages 40 (conducted between 1998 and 2006) and 50 (conducted between 2008 and 2016). Because CES-D-SF score is one of the time-varying covariates in our model, we decided to change all time-varying covariates after 1994 to be those measured at waves coinciding with age 40 measurement (as opposed to a particular calendar year). To mitigate the impacts of this change, age was included as a covariate in the IPTW models for 1994 and 2002 alcohol consumption, and participants were considered censored if their designated ‘age 40’ or ‘age 50’ questionnaire occurred when their actual age at measurement deviated from the target by more than one year (i.e., only those aged 39-41, and 49-51 respectively were eligible).

2. Variable derivation: baseline time-fixed variables

Time-fixed historical alcohol consumption was based on pre-baseline measurement occasions at 1983, 1984, 1985, 1988, 1989 and 1992, and condensed into a single four-category variable consisting: some past occasional drinking, some past moderate drinking, some past above-guidelines drinking, or lifetime abstention (abstinent at all waves *and* reported no past lifetime consumption). Individuals missing more than one wave of pre-1994 data, and with no confirmed above-guidelines drinking, were excluded. No information on heavy episodic drinking was available at the 1992 wave.

Sex was answered at the 1979 initial interview, with male and female the only options. The binary ever-smoked variable was based on 1992 measurement, as was the ever-use of illicit drugs variable (derived from answers to separate questions about cocaine, cannabis, and crack cocaine. Frequent previous religious service attendance was also a binary variable, coded as yes if the individual attended services at least monthly in either 1979 or 1982. Race, as assessed in 1979, was a categorical variable with three options: Hispanic, Black, or Non-Black and Non-Hispanic. Educational attainment was a continuous variable based on response in 1992, and average parental educational attainment was a continuous variable averaging mother and father's years of education based on response in 1992.

3. Distribution of CES-D-SF scores at age 50

Given a large number of 0 values on the CES-D-SF (i.e., complete absence of depressive symptoms), a Shapiro-Wilk test of normality in the baseline sample was performed. This revealed a non-normal distribution ($W=0.81$, $p<0.001$), but this skew value is not considered large enough to warrant transformation or a non-parametric statistical test (1).

4. Covariate selection process

A 2020 meta-analysis of studies evaluating the relationship between alcohol consumption and depressive symptoms was used to guide covariate selection for the current study (2). Covariates controlled for in the meta-analysis' included papers were extracted and tallied (see Web Figure 2). Of the 16 variables most frequently controlled for (controlled for in at least four of the meta-analysis' included studies), the only ones we were unable to derive adequate variables for from our own dataset were those related to physical activity, diet, childhood adversities, and familial history of alcohol or mental health problems. Given other variables may help in predicting exposure group membership, in addition to variables corresponding to the 12 remaining from the above process, we also included illicit drug use (given known comorbidities with alcohol use), household size, health insurance status, urban/rural-dwelling, receipt of welfare, prior religious attendance, and parental education.

5. Variable derivation: time-varying variables

The age variables at each wave were calculated based on age given in 1981, as recommended by the NLSY79 itself. Self-reported health limitation was a binary variable reflecting subjective assessment of whether a health limitation impacted the individual's working ability (excluding cases for whom this limitation is pregnancy only). The marital status variables had three categories: never married, married (both of these options were present in the original question), and then a condensed category of separated/divorced/widowed. Smoking status (N/A at age 40) was a binary current smoking or not variable, as was illicit drug use (derived from answers to separate questions about cocaine, cannabis, and crack cocaine). BMI was a continuous variable and was winsorized at the 95th percentile. Employment status was a binary variable (currently employed or unemployed), as was health insurance status (currently possess or not), urban/rural-dwelling, and current receipt of any form of welfare. Household size was a continuous variable indicating number of individuals in the household (including the participant). Income was a continuous variable representing the individual's income, or if they had a spouse, an average of combined income, and was winsorized at the 95th percentile.

6. Assumptions for valid causal inference in marginal structural models

Drawing valid causal inference from marginal structural models rests on four assumptions: exchangeability, positivity, consistency, and correct specification of propensity score models used in weight generation (3). **Exchangeability** requires that the risk of an outcome in a particular alcohol consumption group X would have been the same for another consumption group Z, had those in X in fact consumed the same amount as those in Z (4). This means there must be no unmeasured confounders and no residual confounding remaining after weighting. Whether exchangeability is met is not strictly testable (3). **Positivity** requires that for any given combination of covariate values, there must be a positive probability of belonging to any of the alcohol consumption categories (as is the case in RCTs, where it is possible for any individual to be assigned to any condition) (4). This can be checked by inspecting descriptive cross-tabulations of exposure groups and categorical/categorized versions of continuous covariates (5). **Consistency** requires that the observed outcome under a given exposure is equal to the potential outcome under that same observed exposure (4). To satisfy the consistency assumption, there must not be multiple versions of a given exposure group (which may reasonably be questioned in the case of alcohol regarding, e.g., different drinking patterns or different beverages consumed). Finally, **correct model specification** of both the exposure model (for IPTWs) and the censoring model (for IPCWs) is assumed, which can be assessed by inspecting the distribution of stabilized weights (3).

7. Weight derivation

Survey weights for all individuals were obtained from the NLSY79 website (https://www.nlsinfo.org/weights/nlsy79), and were then normalized. Inverse probability of treatment weights (IPTWs) were created for alcohol consumption category at 1994, 2002, and 2006. The purpose of the IPTWs was to reweight the sample at each timepoint such that covariates were no longer associated with exposure category. Residuals from generalized linear models regressing alcohol consumption on covariates indicated evidence of statistically significant non-linearity between household size and alcohol category, and also depressive symptoms and alcohol category (not at all waves for either), so quadratic terms were added in the IPTW models for these variables. Individuals' normalized survey weights were incorporated in the PS model (6,7). IPTWs were stabilized by incorporating prior time-varying alcohol exposure values in the numerator, for each person not yet censored.

The below gives the formula for the product of all IPTWs for an individual, where *R* is exposure category, *C* is attrition/censoring, $\bar{R}$ is exposure history, *L* is time-varying covariates, and *X* is baseline covariates:

$$W_i^{\bar{r}} = \prod_{t=1}^{t=3} \left\{ \frac{P(R_{i,t} | \bar{C}_t = 0, \bar{R}_{t-1} = \bar{R}_{i,t-1})}{P(R_{i,t} | \bar{C}_t = 0, \bar{R}_{t-1} = \bar{R}_{i,t-1}, \bar{L}_{t-1} = \bar{L}_{i,t-1}, X = x_i)} \right\}$$

Stabilized inverse probability of censoring weights (IPCWs) were also created for each of these waves, in order to reweight the sample at each timepoint so that covariates were no longer related to attrition status at the *following* wave (i.e., 2002, 2006, age 50).

The below gives the formula for the product of all IPCWs for an individual (same notation as above, and where $\bar{C}$ is attrition/censoring history):

$$W_i^c = \prod_{t=1}^{t=3} \left\{ \frac{P(\bar{C}_{t+1}=0 | \bar{C}_t=0, \bar{R}_t=\bar{R}_{i,t,})}{P(\bar{C}_{t+1}=0 | \bar{C}_t=0, \bar{R}_t=\bar{R}_{i,t}, \bar{L}_t=\bar{L}_{i,t}, X=x_i)} \right\}$$

The final weight for each individual was the product of their three IPTWs, three IPCWs, and survey weight. These final weights were trimmed at the 98th percentile to mitigate the impact of large weights, resulting in a set of weights in the final analyzed sample ranging from 0.07 to 6.19, with a mean of 1.16 and standard deviation of 1.27. As illustrated in Web Figure 3, the final weight greatly improved covariate balance between alcohol consumption groups at each timepoint, with mean differences between groups for most covariates falling within the conservative .1 rule-of-thumb after weighting.

8. R packages and scripts

The key R packages used for these analyses were *WeightIt* (for generating weights) (3,8,9), *cobalt* (for generating plots of covariate balance before and after weighting) (10), *multcomp* (for running contrast analyses) (11), *boot* (for bootstrapping results) (12,13), and *EValue* (for generating E-values) (14–16). An example script for generating IPTWs is provided below. Code for any other steps discussed in this paper is available from the authors on request.

```
```{r}

#Making propensity score formulae for use in weight generation

fixedVars<- c ("Historical_alcoholCategory", "Sex", "EverSmoked", " Everused_illicit", " Self_esteem_historical", " Relgious_attendance_historical", "RACE"," Education_years", "Parental_education_years")

Vars92<- c ("CESD_92", "Age92", "HealthLimit92", "Marital92", "SMOKE92", "Illicit92", "BMI92", "Labour92", "Insurance92", "Household_size92", "Urban92", "Welfare92", "Income92", "I(CESD_92^2)","I(Household_size92^2)")
```
```

```
Vars94<- c ("alcoholCategory_94","CESD_94","Age94", "HealthLimit94","Marital94","SMOKE94",
"Illicit94","BMI94","Labour94","Insurance94","Household_size94","Welfare94","Urban94", "Income94",
"I(CESD_94^2)","I(Household_size94^2)")

Vars40<- c ("alcoholCategory_02", "CESD_Age40",
"HealthLimit40","Marital40","BMI40","Labour40","Insurance40","Household_size40", "Welfare40", "Urban40","Income40",
"I(CESD_Age40^2)","I(Household_size40^2)")

fixedVars <- paste(fixedVars, collapse ="+")
Vars92 <- paste(Vars92, collapse ="+")
Vars94 <- paste(Vars94, collapse ="+")
Vars40 <- paste(Vars40, collapse ="+")

full94_IPTW<-paste("alcoholCategory_94 ~ ",paste(fixedVars,Vars92,sep="+"), collapse="")
full02_IPTW<-paste("alcoholCategory_02 ~ ",paste(fixedVars,Vars92,Vars94,sep="+"), collapse="")
full06_IPTW<-paste("alcoholCategory_06 ~ ",paste(fixedVars,Vars92,Vars94,Vars40,sep="+"), collapse="")

#Generating exposure weights for first wave
Tweights94_surv <- weightit(as.formula(full94_IPTW),
            data=IPTW_df,
            method="cbps",
            over=FALSE,
            stabilize=TRUE,
            s.weights=IPTW_df$NORMSurvWeight)

#Joining these weights back to the data frame
IPTW_df<-bind_cols(IPTW_df, Tweights94_surv[["weights"]])

#Generating exposure weights for second wave
IPTW_df_uncensored1<-IPTW_df %>% filter(CENSOR1==0)
IPTW_df_censored1<-IPTW_df %>% filter(CENSOR1==1)

full02_IPTW_list<-list(as.formula(full02_IPTW))
Tweights02_surv <- weightitMSM((full02_IPTW_list),
            data=IPTW_df_uncensored1,
            method="cbps",
            over=FALSE,
```

```
                    stabilize=TRUE,
                    num.formula=~alcoholCategory_94,
                    s.weights=IPTW_df_uncensored1$NORMSurvWeight)

#Joining these weights back to the data frame
IPTW_df_censored1<-IPTW_df_censored1 %>% mutate(Tweights02_surv=0)
IPTW_df_uncensored1<-bind_cols(IPTW_df_uncensored1, Tweights02_surv[["weights"]])
IPTW_df<-bind_rows(IPTW_df_censored1, IPTW_df_uncensored1)

#Generating exposure weights for third wave
IPTW_df_uncensored2<-IPTW_df %>% filter(CENSOR2==0)
IPTW_df_censored2<-IPTW_df %>% filter(CENSOR2==1)

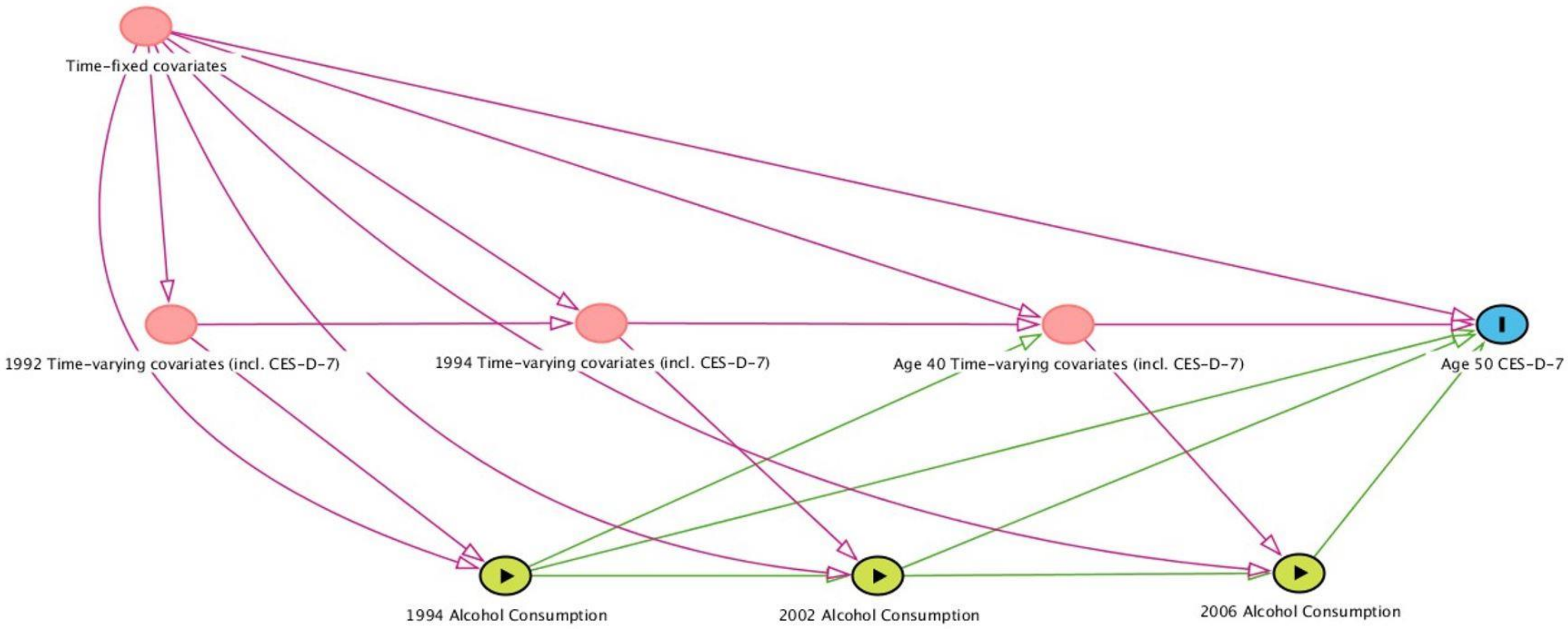
full06_IPTW_list<-list(as.formula(full06_IPTW))
Tweights06_surv <- weightitMSM((full06_IPTW_list),
                    data=IPTW_df_uncensored2,
                    method="cbps",
                    over=FALSE,
                    stabilize=TRUE,
                    num.formula=~alcoholCategory_94+alcoholCategory_02,
                    s.weights=IPTW_df_uncensored2$NORMSurvWeight)

#Joining these weights back to the data frame
IPTW_df_censored2<-IPTW_df_censored2 %>% mutate(Tweights06_surv=0)
#now bind weights from above to IPTW_df_censored using col bing
IPTW_df_uncensored2<-bind_cols(IPTW_df_uncensored2, Tweights06_surv[["weights"]])
#Then bind the 2 dfs together so you have both censored and uncensored together again
IPTW_df<-bind_rows(IPTW_df_censored2, IPTW_df_uncensored2)
```

9. Sensitivity analyses: Stratification by sex

Analyses (as described in the Methods section of the main article and Web Methods in the supplement) were re-performed separately for the female and male sub-samples, beginning with separate weight generation in each sub-sample. Unlike for the main analysis, sex was not included as a covariate in IPTW/IPCW models, and survey weights not incorporated, as subsamples are not intended to be representative of the general population. For females, weights for the final analyzed sample ranged from 0.11 to 10.92, with a mean of 1.55 and standard deviation of 2.00. For males, weights in the final analyzed sample ranged from 0.13 to 7.69, with a mean of 1.34 and standard deviation of 1.51.

Figure S1. Directed acyclic graph (DAG) of assumed longitudinal relationships between variables over time



DAG made with the DAGitty web application (17). Pink bubbles indicate ancestors of exposure and outcome, green bubbles indicate exposures, and the blue bubble indicates outcome. Pink lines indicate biasing pathways while green lines indicate causal pathways.

Figure S2. Covariates controlled for in studies included in the Li et al. meta-analysis

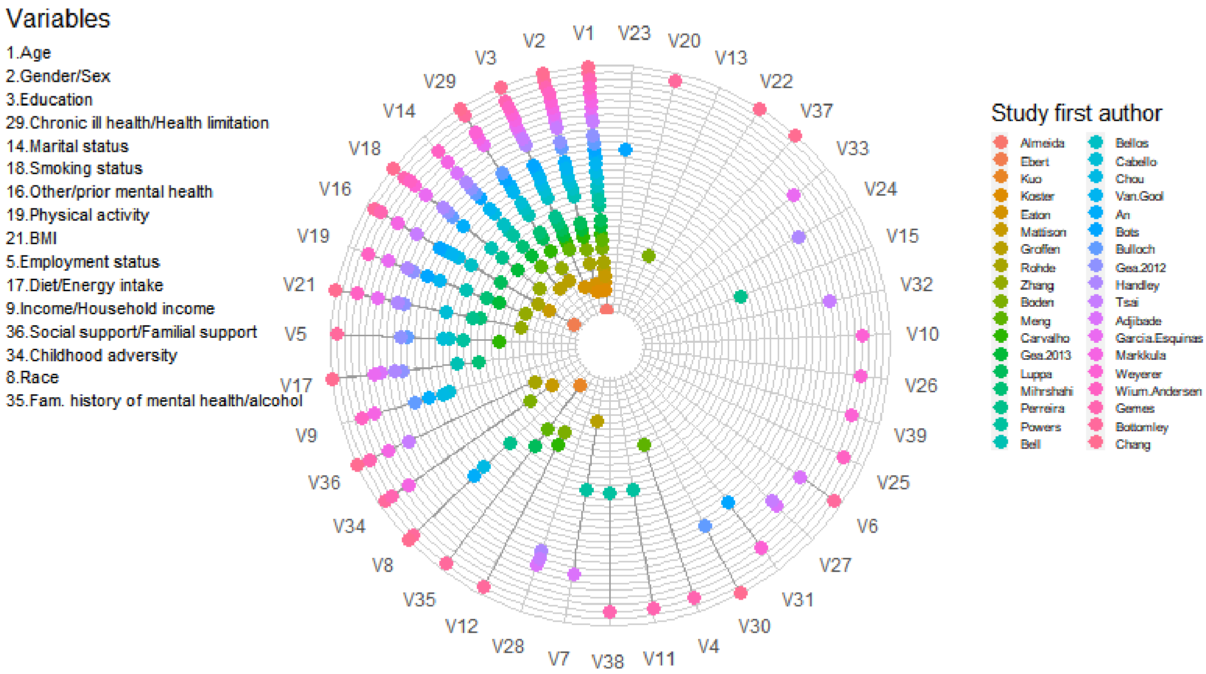
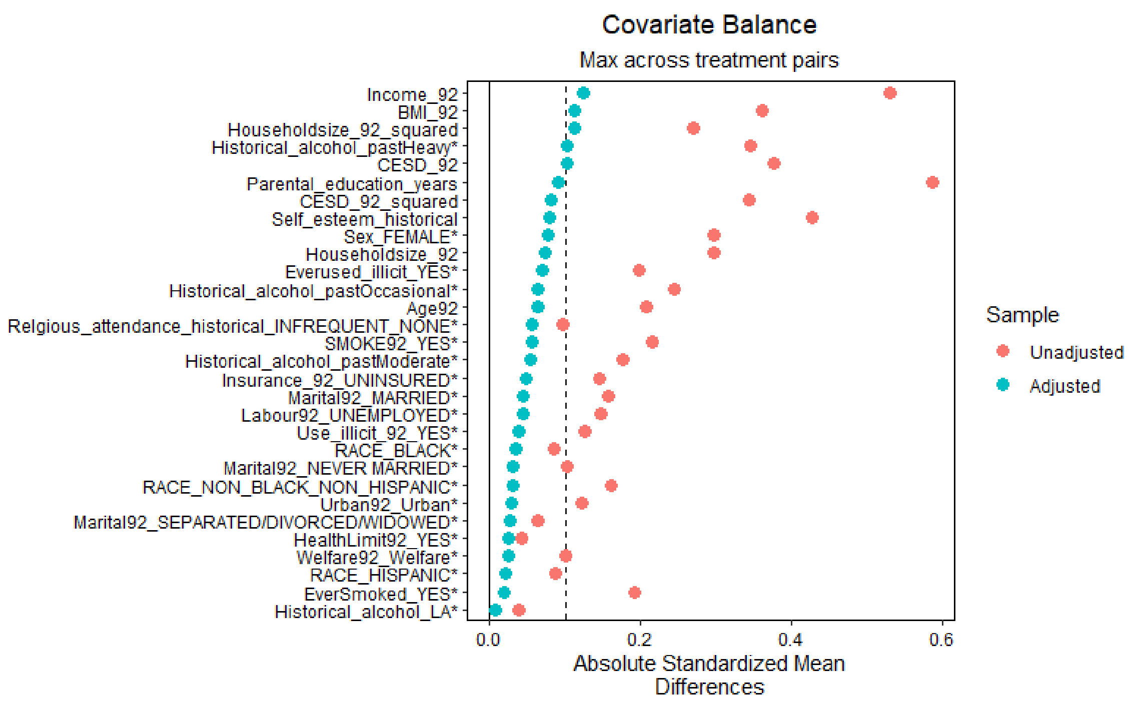
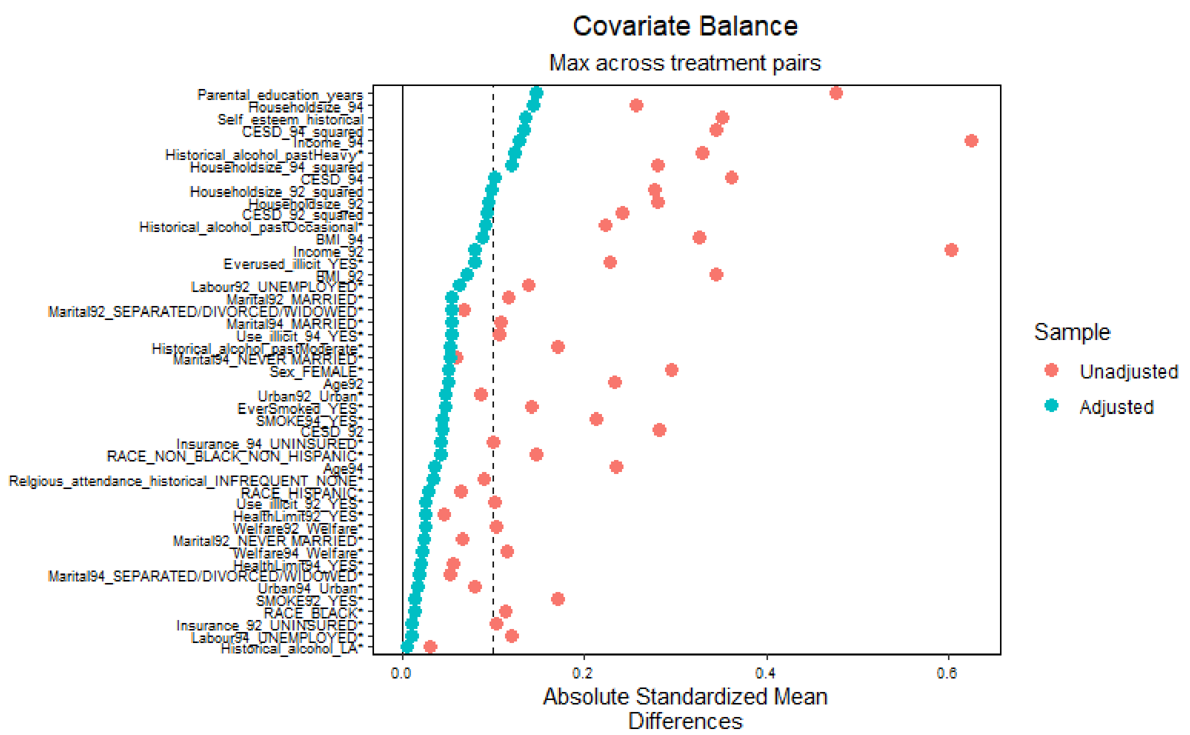


Figure S3. Covariate balance between exposure groups at 1994, 2002, and 2006 before and after weighting by the final weight[a]

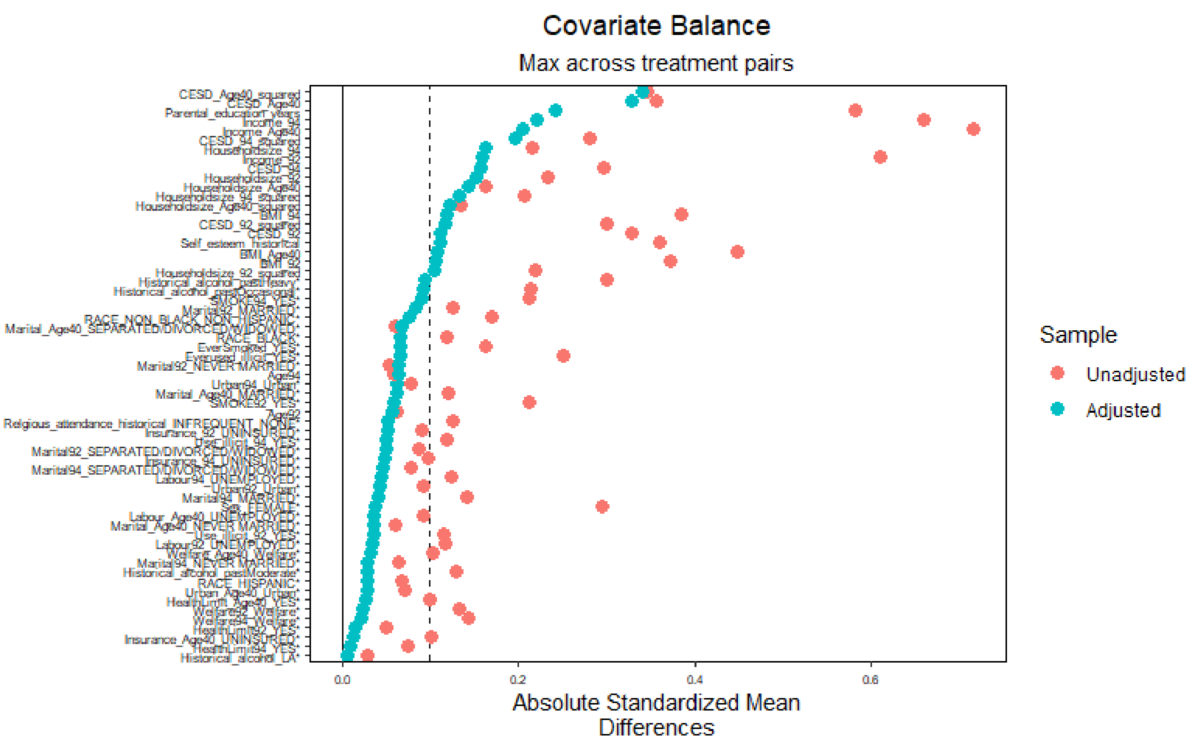
A) 1994



B) 2002

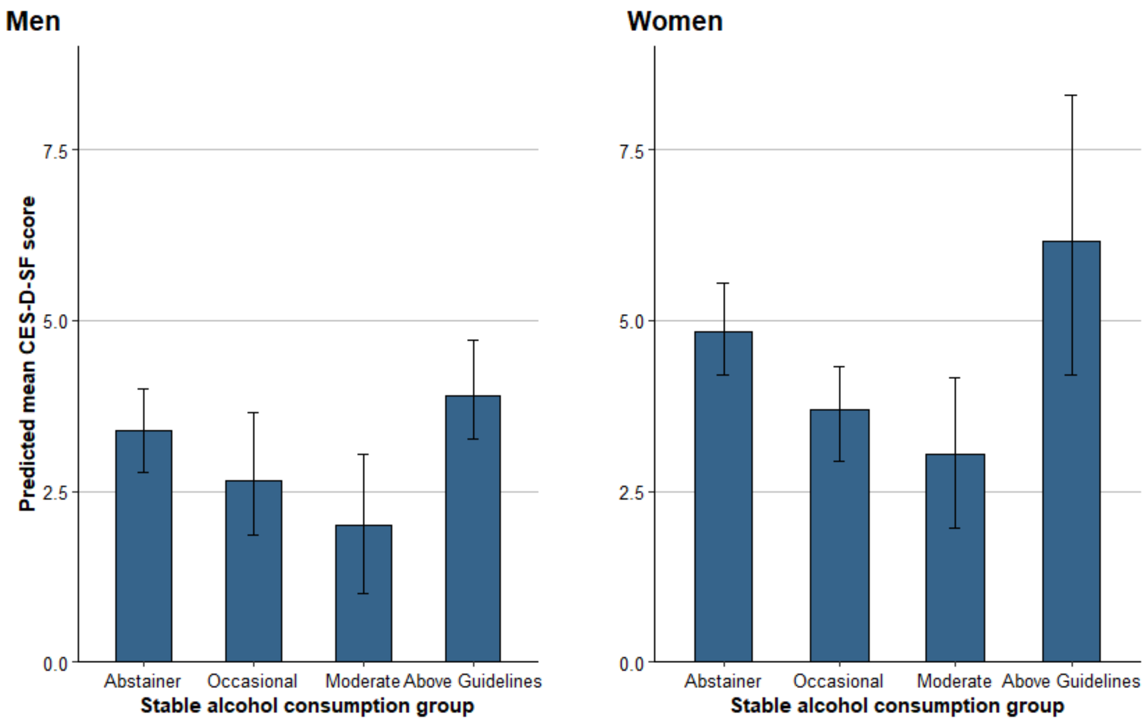


C) 2006



[a]Figures display differences in means for the two exposure groups for which these differences are largest. Variables followed by an asterisk are unstandardized categorical variables. Covariate names (and their categories) are abbreviated for brevity: ‘illicit’ stands for illicit drugs; ‘HealthLimit’ stands for self-reported health limitations; ‘Historical_alcohol_LA’ stands for lifetime abstainer at baseline.

Figure S4. Predicted mean CES-D-SF scores at age 50 for those consistently abstaining, drinking occasionally, moderately or above guidelines over the 1994, 2002 and 2006 measurement occasions, stratified by sex[a]



[a]Whiskers indicate bootstrapped 95% CIs. Groups do not possess a specific number of subjects as they represent hypothetical trajectories.

Figure S5. Predicted probability of probable depression at age 50 for those consistently abstaining, drinking occasionally, moderately or above guidelines over the 1994, 2002 and 2006 measurement occasions, stratified by sex[a]

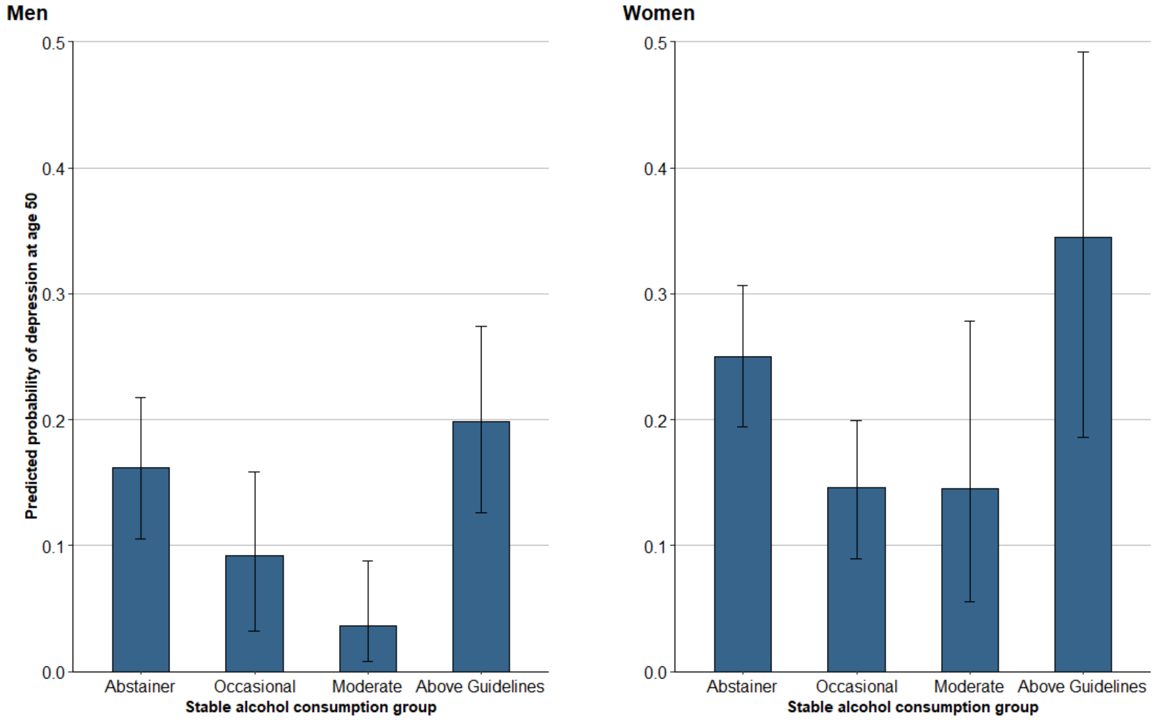


Table S1. Ways in which MSMs can improve causal inference

Common limitations of past research	**MSMs:**
Traditional confounder adjustment relies on making accurate modelling assumptions about confounder-outcome relationships(18–20)	Instead require modelling confounder-treatment relationships, with the ability to include multiple functional forms
Adjusting only for baseline covariates ignores time-varying confounders, while standard control for time-varying covariates will block mediated effects and/or induce collider bias(21,22)	Can incorporate time-varying covariates that act as both mediators and confounders
Difficulty in isolating effects in the direction of interest (i.e., alcohol's effect on depression isolated from depression's effect on alcohol consumption)	Can incorporate multiple measurements of both the exposure and outcome over time, accounting for time-varying exposures being affected by past outcome levels
Including only baseline alcohol consumption induces exposure misclassification, ignoring variability in consumption over follow-up(23)	Can incorporate multiple measurements of exposure over time
Selection biases are common,(24) including over-representation of healthy drinkers and differential attrition	Can incorporate both survey weights (making the baseline sample representative of the population) and censoring weights (making the final sample representative of baseline)

Table S2. Output from contrasts for analyses using continuous CES-D-SF scores as outcome

Contrasts	**Estimates** (*b*)	**Std. Error**	***P***	**Bootstrapped CI**
Consistent occasional vs Consistent abstinence	-0.84	0.26	0.001	**-1.63, -0.11**
Consistent moderate vs Consistent abstinence	-1.08	0.30	<0.001	**-1.88, -0.20**
Consistent above-guidelines vs Consistent abstinence	0.34	0.25	0.162	-0.62, 1.25

Table S3. Output from contrasts for analyses using binary probable depression as outcome

Contrasts	**Estimates** (OR)	***P***	**Bootstrapped CI**
Consistent occasional vs Consistent abstinence	0.58	<0.001	**0.36, 0.88**
Consistent moderate vs Consistent abstinence	0.59	0.005	0.26, 1.13
Consistent above-guidelines vs Consistent abstinence	1.06	0.683	0.66, 1.72

References

1. Kim H-Y. Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restor. Dent. Endod.* 2013;38(1):52.
2. Li J, Wang H, Li M, et al. Effect of alcohol use disorders and alcohol intake on the risk of subsequent depressive symptoms: a systematic review and meta-analysis of cohort studies. *Addiction*. 2020;115(7):1224–1243.
3. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. *Am. J. Epidemiol.* 2008;168(6):656–664.
4. Hernán MA, Robins JM. Causal Inference: What If. Boca Raton, FL: Chapman & Hill/CRC; 2020.
5. Zhong QY, Gelaye B, Weele TJV, et al. Causal model of the association of social support with antepartum depression: Amarginal structural modeling approach. *Am. J. Epidemiol.* 2018;187(9):1871–1879.
6. Dugoff EH, Schuler M, Stuart EA. Generalizing observational study results: Applying propensity score methods to complex surveys. *Health Serv. Res.* 2014;49(1):284–303.
7. Ridgeway G, Kovalchik SA, Griffin BA, et al. Propensity score analysis with survey weighted data. *J. causal inference*. 2015;3(2):237–249.
8. Greifer N. WeightIt: weighting for covariate balance in observational studies. *R*

Packag. version 0.9. 0. 2020;

9. Imai K, Ratkovic M. Covariate balancing propensity score. *J. R. Stat. Soc. Ser. B (Statistical Methodol.* 2014;76(1):243–263.

10. Greifer N. cobalt: Covariate balance tables and plots. *R Packag. version*. 2017;2(0).

11. Hothorn T, Bretz F, Ag P, et al. Simultaneous inference in general parametric models. *Biometrical J. J. Math. Methods Biosci.* 2008;50(3):346–363.

12. Canty A, Ripley B. boot: Bootstrap R (S-Plus) functions. *R Packag. version*. 2017;1:3–20.

13. Davison AC, Hinkley DV. Bootstrap methods and their application. Cambridge university press; 1997.

14. VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. *Ann. Intern. Med.* 2017;167(4):268–274.

15. Mathur MB, VanderWeele TJ. Sensitivity analysis for unmeasured confounding in meta-analyses. *J. Am. Stat. Assoc.* 2019;

16. Smith LH, VanderWeele TJ. Bounding bias due to selection. *Epidemiology*. 2019;30(4):509.

17. Textor J, Zander B Van Der, Gilthorpe MS. Robust causal inference using directed acyclic graphs : the R package ' dagitty .' *Int. J. Epidemiol.* 2017;45(6):1887–1894.

18. Brookhart MA, Wyss R, Layton JB, et al. Propensity score methods for confounding control in nonexperimental research. *Circ. Cardiovasc. Qual. Outcomes*. 2013;6(5):604–611.

19. Thoemmes F, Ong AD. A Primer on Inverse Probability of Treatment Weighting and Marginal Structural Models. *Emerg. Adulthood*. 2016;4(1):40–59.

20. Benedetto U, Head SJ, Angelini GD, et al. Statistical primer: Propensity score matching and its alternatives. *Eur. J. Cardio-thoracic Surg.* 2018;53(6):1112–1117.

21. Williamson T, Ravani P. Marginal structural models in clinical research: When and how to use them? *Nephrol. Dial. Transplant.* 2017;32(February):ii84–ii90.

22. Vandecandelaere M, Vansteelandt S, De Fraine B, et al. Time-Varying Treatments in

Observational Studies: Marginal Structural Models of the Effects of Early Grade Retention on Math Achievement. *Multivariate Behav. Res.* [electronic article]. 2016;51(6):843–864. (http://dx.doi.org/10.1080/00273171.2016.1155146)

23. Chikritzhs T, Fillmore K, Stockwell T. A healthy dose of scepticism: Four good reasons to think again about protective effects of alcohol on coronary heart disease. *Drug Alcohol Rev.* [electronic article]. 2009;28(4):441–444. (https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1465-3362.2009.00052.x)

24. Naimi TS, Stockwell T, Zhao J, et al. Selection biases in observational studies affect associations between ‘moderate’ alcohol consumption and mortality. *Addiction* [electronic article]. 2017;112(2):207–214. (https://onlinelibrary.wiley.com/doi/abs/10.1111/add.13451)