Marginal structural model

CMED6040 – Session 7

Tim Tsang (matklab@hku.hk)

School of Public Health
The University of Hong Kong

27 June 2023

Session 7 learning objectives

After this session, students should be able to

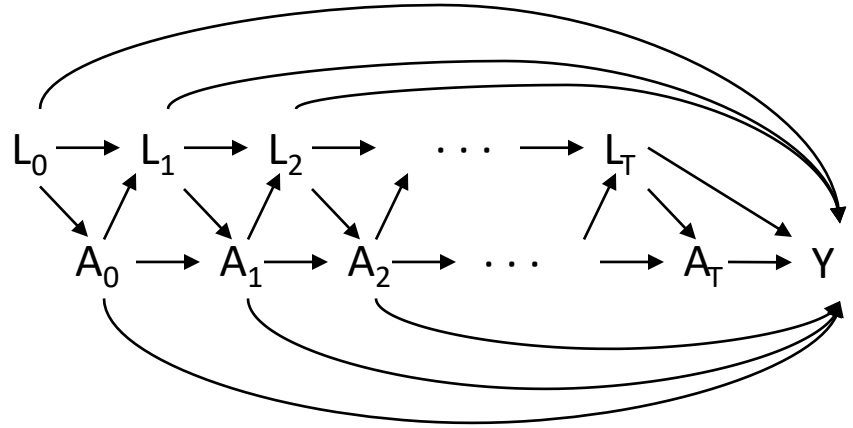
- Recognize the situation in longitudinal study with time-dependent confounder
- Analyse longitudinal data using Marginal Structural Models (MSM), accounting for endogenous confounding
- Recognize the assumptions of MSM
- Interpret the results from MSM

Longitudinal observational study

- Always more difficult to establish causality in observational study
- Need to deal with confounding (by both measured and unmeasured confounders)
- There may be time-dependent confounders which are at the same time affected by treatment (endogenous confounders)
- Estimating causal effects by adjusting for confounding may not be straightforward

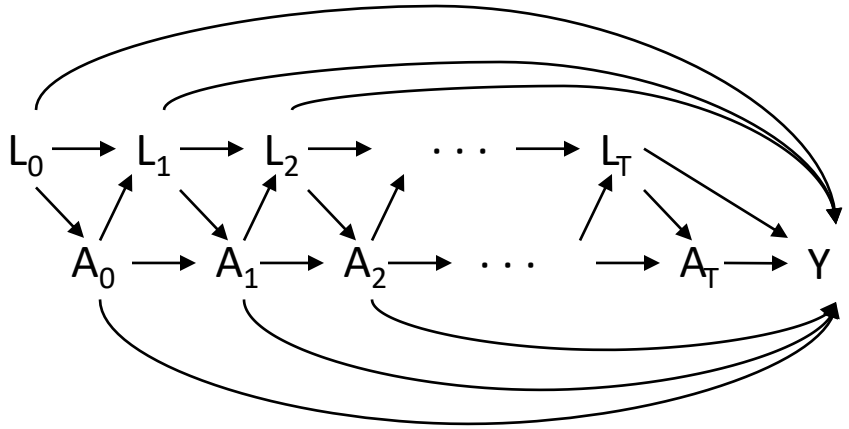
Endogenous confounding in observational study

- Endogenous confounding can be difficult to handle in standard methods:



- The above DAG assumed no unobserved confounder
- Suppose we want to estimate the causal effect of *A* on the outcome *Y*
- *L* is a time-varying confounder (e.g. $A_2 \leftarrow L_2 \rightarrow Y$)
- Suppose we adjust for *L*, however
  - *L* is also a mediator (e.g. $A_1 \rightarrow L_2 \rightarrow Y$)
  - Adjusting for *L* may affect the estimation of the causal effect of *A*

Endogenous confounding in observational study



- Example: Obesity (*A*), CVD (*L*), all-cause mortality (*Y*)
- Influenza vaccine (*A*), health status (*L*), influenza virus infection (*Y*)
- Depressive symptoms (*A*), health behavior (*L*), stroke (*Y*)

Strategy to control for endogenous confounding

- Imagine we can 'clone' the subject with identical characteristics, except taking the opposite treatment
- By comparing their response directly this will give us the causal treatment effect
- Note that for time-varying treatment, then a lot of clones are needed
  - e.g. treatment choice at n time points require 2^n-1 clones for comparison
- The idea is to weight the samples to form a pseudo-population that exposure is independent of the measured confounders (balance confounders across levels of exposure).
- Under such conditions, standard regression model can be used (without the measured confounders)

Counterfactuals

- Causal effect can be estimated by $Y_1 - Y_0$, if both were observable
- Suppose hypothetically the 'full data' is (Y_0, Y_1) for each subject
- Regard counterfactuals as missing data (cannot observe both in practice)
- Using IPW to reweight the dataset so that mean predictor levels are balanced between treated and untreated subjects
- After reweighting, a marginal analysis can be performed
- Weighting based on inverse probability
  - Propensity score for time-independent case
  - Treatment can also be time-dependent
  - At each time point, the probability was estimated conditional on past confounders and may also on past responses (depending on the assumed causal structure)
  - Can also handle dropouts similarly

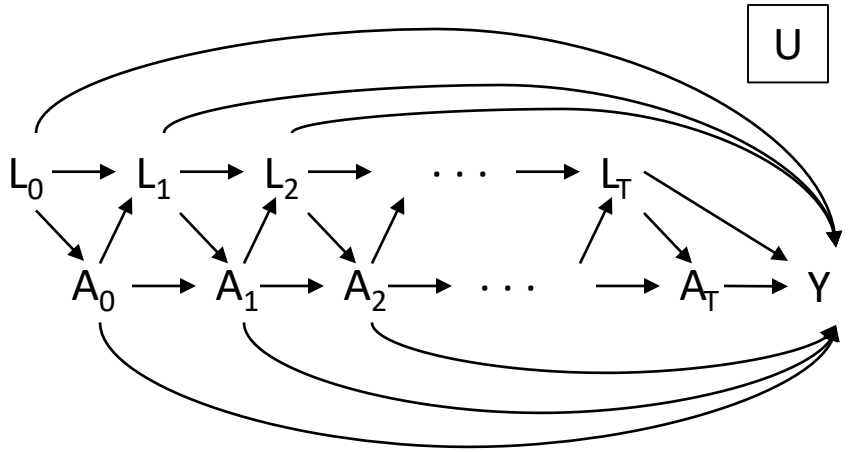
Marginal structural models (MSM)

- Handle the situation that endogenous confounders cannot be easily adjusted for in standard analysis
- 'Marginal' – do not model the correlation between observed outcome and counterfactuals
- 'Structural' – model the probabilities of counterfactual variables (terminology from econometrics or social sciences)
- Use inverse probability weighting to recover the 'missing' counterfactuals
- Treatment and predictors in the weighted dataset will be unconfounded (similar to propensity score method)
- Can be applied to GLM, Cox regression or GEE

Key assumptions of MSM

- No unmeasured confounders
  - Not testable
  - Also called conditional exchangeability

    $Y_k \perp\!\!\!\perp A_k \mid \bar{A}_{k-1}, \bar{L}_{k-1}$



- Experimental treatment assignment
  - Cannot perfectly predict treatment/exposure (probability of treatment ≠ 0 or 1)
  - Also called structural positivity

- Models are specified correctly

Inverse probability weighting in MSM

- We want to recover the 'full' dataset (including counterfactuals)
  - At each time point for longitudinal analysis
- Inverse probability of treatment weighed (IPTW) estimator
- Cannot have complete confounding (probability = 0 or 1), similar to propensity score analysis
  - need to review the weighting (to make sure they are away from 0 or 1)
  - called "experimental treatment assignment"

Recap on inverse probability weighting (IPW)

- Consider the case that we have a point treatment
- We want to estimate $E[Y_1-Y_0]$ using inverse probability weighting (IPW)
- Observed outcome is actually conditioned on the assigned treatment, i.e., observed $E[Y_1|T=1]$ and $E[Y_0|T=0]$
- estimate $E[Y_i]$ by $E[Y_i\delta(T=i)/P(T=i)]$, i=0, 1 (δ: indicator function)
- The inverse probability weights are given by $w_i = \frac{1}{P(A_i=a_i|L_i=l_i)}$, L is time-varying confounding, A is the exposure
- Average treatment effect = $E[Y_1-Y_0]$ =

$$\frac{1}{N}\left(\sum_{T_i=1}\frac{Y_{1i}}{p_{1i}} - \sum_{T_i=0}\frac{Y_{0i}}{1-p_{1i}}\right) = \frac{1}{N}\left(\sum_{T_i=1} w_{1i}Y_{1i} - \sum_{T_i=0} w_{0i}Y_{0i}\right)$$

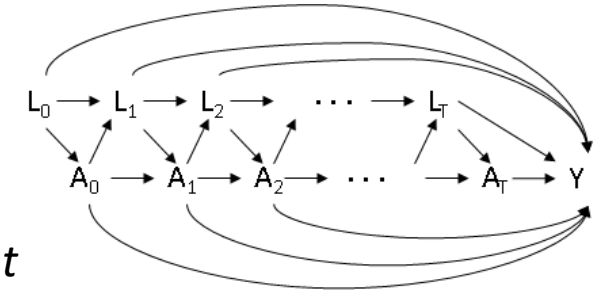
Stabilized weights

- IPW for individual *i*: $w_i = \frac{1}{P(A_i=a_i|L_i=l_i)}$ can be unstable
  - Extremely large weighting may dominate the further weighted analysis and increase the variability
- Stabilized weight was also proposed:

$$sw_i = \frac{P(A_i=a_i)}{P(A_i=a_i|L_i=l_i)}$$

- *P*(*A*) appearing in both numerator and denominator helps to reduce the variation of the weights
- Always use stabilized weight for continuous exposure variable

IPW in longitudinal analysis



- The treatment/exposure *A* may change over time
- Exposure measured at A_0, A_1, A_2, … A_T
  - $\bar{A}_t = (A_0, A_1, \ldots, A_t)$ to represent the history of *A* up to time *t*
- Similarly for confounder $\bar{L}_t = (L_0, L_1, \ldots, L_t)$
- To create pseudo-population with no confounding at each time point
- The stabilized weights for individual *i*, time point t_{ij}, are given by:

$$sw_{ij} = \prod_{k=0}^{j} \frac{P(A_{ik} = a_{ik} | \bar{A}_{ik-1} = \bar{a}_{ik-1})}{P(A_{ik} = a_{ik} | \bar{A}_{ik-1} = \bar{a}_{ik-1}, \bar{L}_{ik} = \bar{l}_{ik})}$$

- Include baseline / time-independent variables in the numerator model and both baseline and time-dependent variables in the denominator model

Standard error for MSM estimates

- Need to account for the correlation induced by weighting
- For longitudinal data there are also within-subject correlations
- Robust standard error estimator is used to account for such correlation (e.g. sandwich estimator)
- May also use bootstrap method to obtain the standard error

MSM in R

- package: ipw
- ipwpoint, ipwtm to generate inverse probability weighting for situations with time-independent and time-varying exposure/confounder
- ipwpoint(*exposure*, *family*, *link*, *numerator = NULL*, *denominator*, *data*, …)
  - *exposure* specifies the exposure/treatment variable of interest
  - *family* specifies the family of the relation, e.g. “binomial”, “multinomial”, “ordinal”, “gaussian”
  - *link* specifies the link function, e.g. “logit” for logistic regression, “gaussian” for linear regression, other links are also possible
  - *numerator* is a formula which specifies the predictors in the numerator of IPW. If unspecified, unstabilized weights will be used (using 1 in the numerator of IPW)
  - *denominator* is a formula which specified the predictors in the denominator of IPW. Confounders should be included.

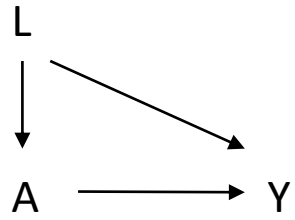
Example – simulated data

- Example from van der Wal et al. 2011
- Simulate data from the following process:

$$L \sim N(10, 5^2)$$

$$logit(P(A = 1)) = -10 + L$$

$$Y = 10A + 0.5L - 10 + \varepsilon, \text{where } \varepsilon \sim N(0, 5^2)$$



- The true causal effect of *A* on *Y* is then 10
- This can be estimated by standard linear regression by adjusting for *L*
- An alternative is to use IPTW estimator

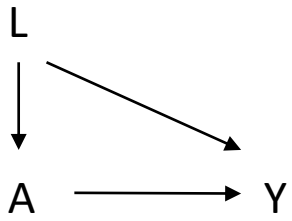
Example – simulated data

- Simulating the dataset:

$$L \sim N(10, 5^2)$$

$$logit(P(A = 1)) = -10 + L$$

$$Y = 10A + 0.5L - 10 + \varepsilon, \text{where } \varepsilon \sim N(0, 5^2)$$




```
set.seed(111)
n <- 1000
simdat <- data.frame(l = rnorm(n, 10, 5))
a.link <- simdat$l - 10
pa <- exp(a.link)/(1 + exp(a.link))
simdat$a <- rbinom(n, 1, prob = pa)
simdat$y <- 10*simdat$a + 0.5*simdat$l - 10 + rnorm(n, 0, 5)
simdat[1:4,]
```

```
          l a           y
1 11.176104 1  12.9941002
2  8.346321 0  -0.5416105
3  8.441881 0 -11.8720489
4 -1.511728 0  -9.6681398
```

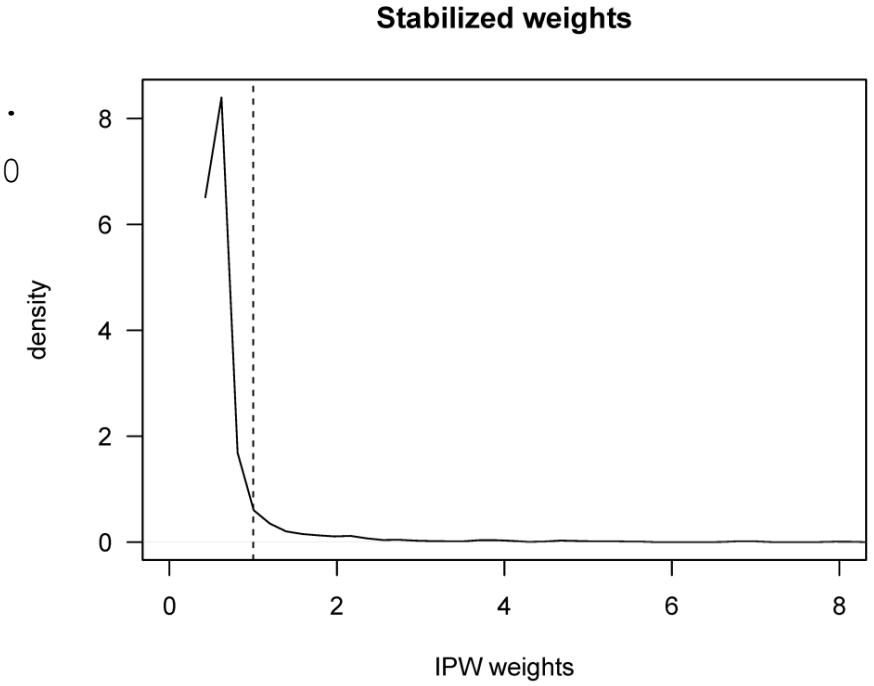
Simulated data – estimating the weights

```
library(ipw)
temp <- ipwpoint(exposure = a, family = "binomial", link =
"logit", numerator = ~ 1, denominator = ~ 1, data = simdat)
summary(temp$ipw.weights)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.4890  0.5062  0.5198  1.0080  0.6257 98.8700
```

- ipwplot() to show the distribution of the weights

```
ipwplot(weights = temp$ipw.weights,
logscale = FALSE,
main = "Stabilized weights",
xlim = c(0, 8))
```



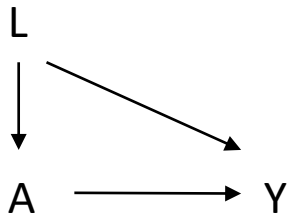
Simulated data – IPW models

- To examine the IPW models:

$$L \sim N(10, 5^2)$$

$$logit(P(A=1)) = -10 + L$$

$$Y = 10A + 0.5L - 10 + \varepsilon, \text{where } \varepsilon \sim N(0, 5^2)$$




```
summary(temp$num.mod)
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.04401    0.06326   0.696    0.487
```

```
summary(temp$den.mod)
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -10.41319    0.74004  -14.07   <2e-16 ***
l             1.04125    0.07257   14.35   <2e-16 ***
```

- Note that for this example, the numerator can only take two values: P(*A*=0) and P(*A*=1)

Simulated data – obtaining IPTW estimates

```
simdat$sw <- temp$ipw.weights

require(survey)
```

- svyglm() (in package 'survey') to fit the MSM model applying the IPW
  - svydesign(*id*, *weights*, *data*) to specify the weighting (id ~ 1 for no clusters)
  - svyglm will give the robust estimator for the standard error, accounting for the clustering effect introduced by weighting
- The IPW will balance the confounder (*L*) and hence is not adjusted for in the model:

```
msm <- svyglm(y ~ a, design = svydesign(~ 1, weights = ~ sw,
data = simdat))
```

MSM in R

```
summary(msm)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -6.3031     0.2921  -21.58   <2e-16 ***
a            10.8823     0.8397   12.96   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
confint(msm)[2,]
    2.5 %    97.5 %
 9.236574 12.527945
```

- The estimate causal effect of *A* on *Y* is 10.9 (95% CI = 9.2–12.5)

MSM using unstabilized weight

```
temp.uns <- ipwpoint(exposure = a, family = "binomial", link =
"logit", denominator = ~ 1, data = simdat)
```

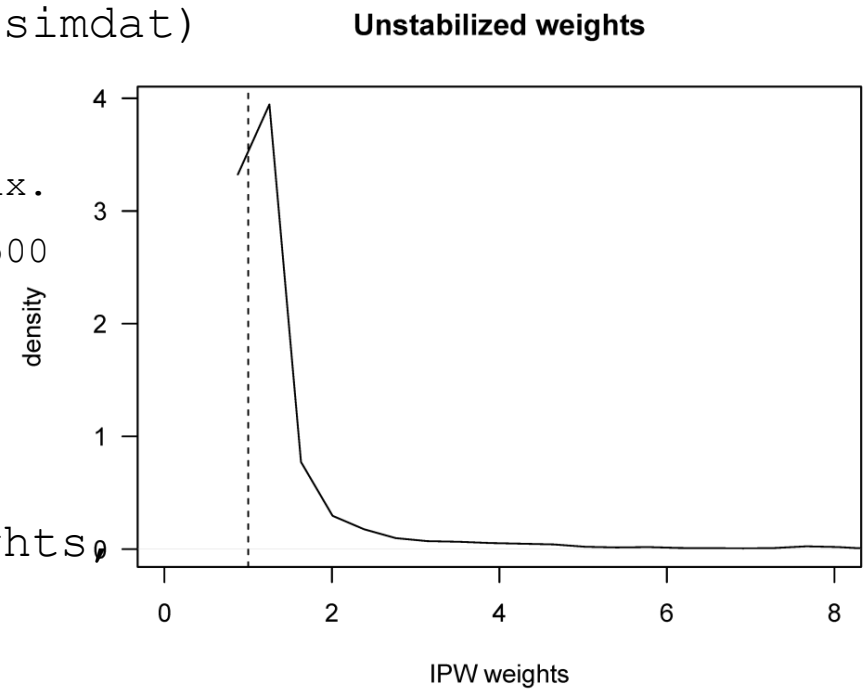
```
summary(temp.uns$ipw.weights)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.000	1.003	1.032	2.013	1.247	193.500

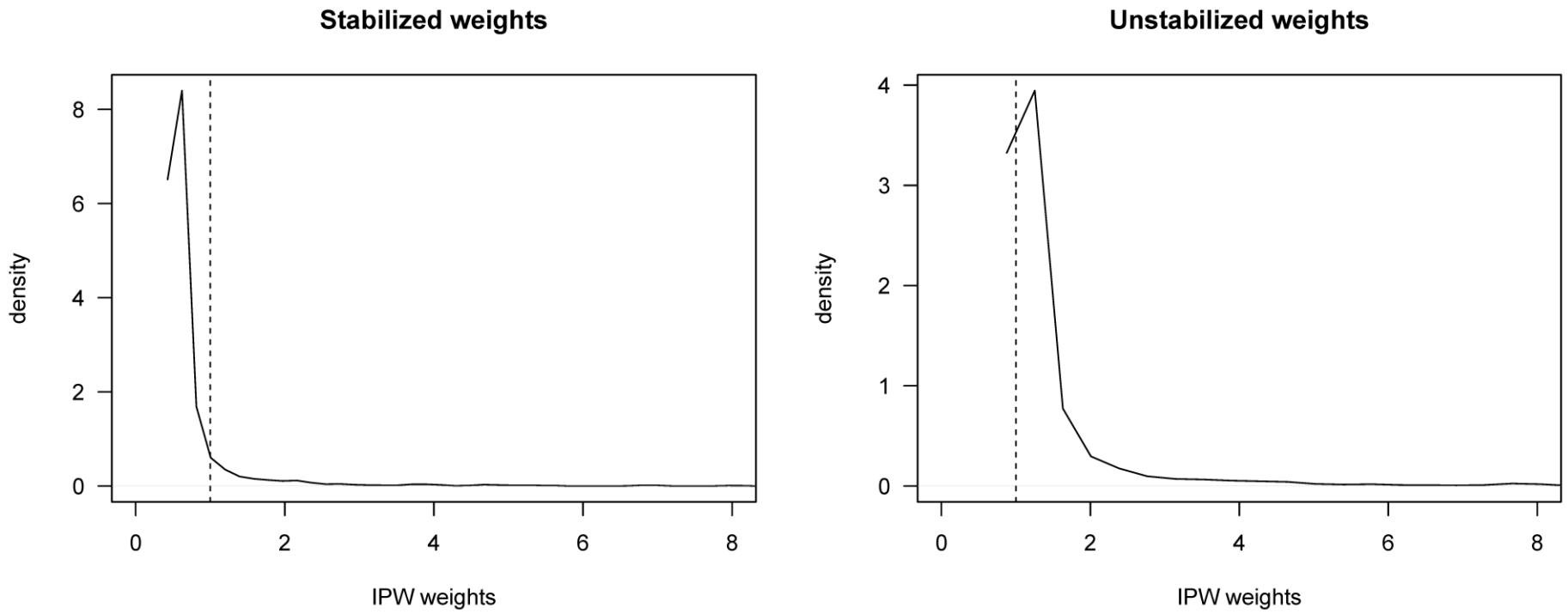
- ipwplot() to show the distribution

of the weights

```
ipwplot(weights = temp.uns$ipw.weights,
logscale = FALSE,
main = "Unstabilized weights",
xlim = c(0, 8))
```



Stabilized weights



- Stabilized weights are able to reduce the variation

MSM using unstabilized weights

```
simdat$unsw <- temp.uns$ipw.weights
msm.uns <- svyglm(y ~ a, design = svydesign(~ 1, weights = ~
unsw, data = simdat))
summary(msm.uns)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -6.3031     0.2921  -21.58   <2e-16 ***
a            10.8823     0.8397   12.96   <2e-16 ***
confint(msm.uns)[2,]
    2.5 %    97.5 %
 9.236574 12.527945
```

- The estimate causal effect of *A* on *Y* is 10.9 (95% CI = 9.2–12.5)
- Very similar results as those using stabilized weights

MSM in R (time-dependent weights)

- ipwtm(*exposure*, *family*, *link*, *numerator = NULL*, *denominator*, *id*, *tstart*, *timevar*, *type*, *data*, …)
  - *exposure, family, link, numerator, denominator, data* same as ipwpoint
  - *id* identifies the cluster/subject where multiple measurements were made
  - *tstart* specifies the start time of the follow-up intervals, only needed when family = "survival" (*tstart* should be < 0 for the first interval)
  - *timevar* is the follow-up time. When family = "survival", it specifies the end time of the follow-up intervals
  - *type* specifies the type of exposure. Type="first" estimates weights up to the first switch of the exposure value (e.g. for death, from 0 to 1) then keep constant afterwards. Alternative is type="all" that weights were estimated at all time points.

Example – HAART and HIV survival

- Dataset from “ipw” package
- Can be loaded by data(haartdat)
- 1200 HIV patients was followed every 100 days since HIV seroconversion
- HAART therapy are initiated which may depend on CD4 counts
  - It is possible to initiate HAART therapy at time = 0, so the first interval was set to be < 0 to facilitate model fitting
  - *haartind* indicating whether HAART therapy was initiated in the follow-up interval
- *tstart, fuptime*: start and end of the follow-up interval
- *age*: age at the start of follow-up year
- cd4.sqrt: square root of the CD4 counts
- *event*: indicator of death at the end of the interval

Example – HAART and HIV survival

```
data(haartdat)

haartdat[haartdat$patient==18,]
```

	patient	tstart	fuptime	haartind	event	sex	age	cd4.sqrt	endtime	dropout
465	18	-100	0	0	0	0	49	25.67100	600	0
466	18	0	100	0	0	0	49	30.34798	600	0
467	18	100	200	0	0	0	49	20.85665	600	0
468	18	200	300	0	0	0	49	23.55844	600	0
469	18	300	400	0	0	0	49	24.06242	600	0
470	18	400	500	0	0	0	49	24.95997	600	0
471	18	500	600	0	0	0	49	26.58947	600	1

- This patient dropout of the study at day 600

Example – HAART and HIV survival

```
haartdat[haartdat$patient==63,]
```

	patient	tstart	fuptime	haartind	event	sex	age	cd4.sqrt	endtime	dropout
1030	63	-100	0	0	0	0	22	21.49419	1100	0
1031	63	0	100	0	0	0	22	16.88194	1100	0
1032	63	100	200	0	0	0	22	18.68154	1100	0
1033	63	200	300	0	0	0	22	23.00000	1100	0
1034	63	300	400	0	0	0	22	22.60531	1100	0
1035	63	400	500	0	0	0	22	22.75961	1100	0
1036	63	500	600	0	0	0	22	18.27567	1100	0
1037	63	600	700	0	0	0	22	20.68816	1100	0
1038	63	700	800	0	0	0	22	20.54264	1100	0
1039	63	800	900	0	0	0	22	18.08314	1100	0
1040	63	900	1000	0	0	0	22	16.88194	1100	0
1041	63	1000	1100	0	1	0	22	16.43168	1100	0

- This patient died at day 1100 (without initiation of HAART)

Example – HAART and HIV survival

- Note that the data format has a 'counting process' structure
- Each row specifies {N, Y, $\boldsymbol{Z}$}, count of events (0 or 1 for death), indicator for being at risk, and other predictors for the period $(t_1, t_2]$
- t_1, t_2 were chosen so that the subject were at risk during $(t_1, t_2]$, and the event may happen at t_2
- Different from interval censoring that the event may happen anywhere between $(t_1, t_2]$

Fitting a standard time-dependent Cox regression

```
cox0 <- coxph(Surv(tstart, fuptime, event) ~ haartind + age + sex
+ cd4.sqrt + cluster(patient), data = haartdat)

summary(cox0)

             coef exp(coef) se(coef) robust se      z Pr(>|z|)
haartind -0.61994   0.53798  0.43837   0.41971 -1.477 0.139657
age       0.05943   1.06123  0.01522   0.01657  3.588 0.000334 ***
sex       0.04408   1.04507  0.47194   0.45187  0.098 0.922282
cd4.sqrt -0.13125   0.87700  0.03440   0.04682 -2.803 0.005060 **
         exp(coef) exp(-coef) lower .95 upper .95
haartind     0.538     1.8588    0.2363    1.2247
age          1.061     0.9423    1.0273    1.0963
sex          1.045     0.9569    0.4310    2.5338
cd4.sqrt     0.877     1.1403    0.8001    0.9613
```

- The estimated hazard ratio of patients receiving HAART was 0.54 (95% CI 0.23–1.22), which is not statistically significant

HAART and HIV survival – estimating the weights

- To adjust for the endogenous confounder (CD4 counts), we use IPW:

$$sw_{ij} = \prod_{k=0}^{j} \frac{P(H_{ik}=h_{ik}|\bar{H}_{ik-1}=\bar{h}_{ik-1}, \boldsymbol{V_i}=\boldsymbol{v_i})}{P(H_{ik}=h_{ik}|\bar{H}_{ik-1}=\bar{h}_{ik-1}, \bar{L}_{ik}=\bar{l}_{ik}, \boldsymbol{V_i}=\boldsymbol{v_i})}$$

  which also include time-independent variables *V* (age and sex)

```
temp <- ipwtm(exposure = haartind, family = "survival",
numerator = ~ sex + age, denominator = ~ cd4.sqrt + sex + age,
id = patient, tstart = tstart, timevar = fuptime, type = "first",
data = haartdat)
```

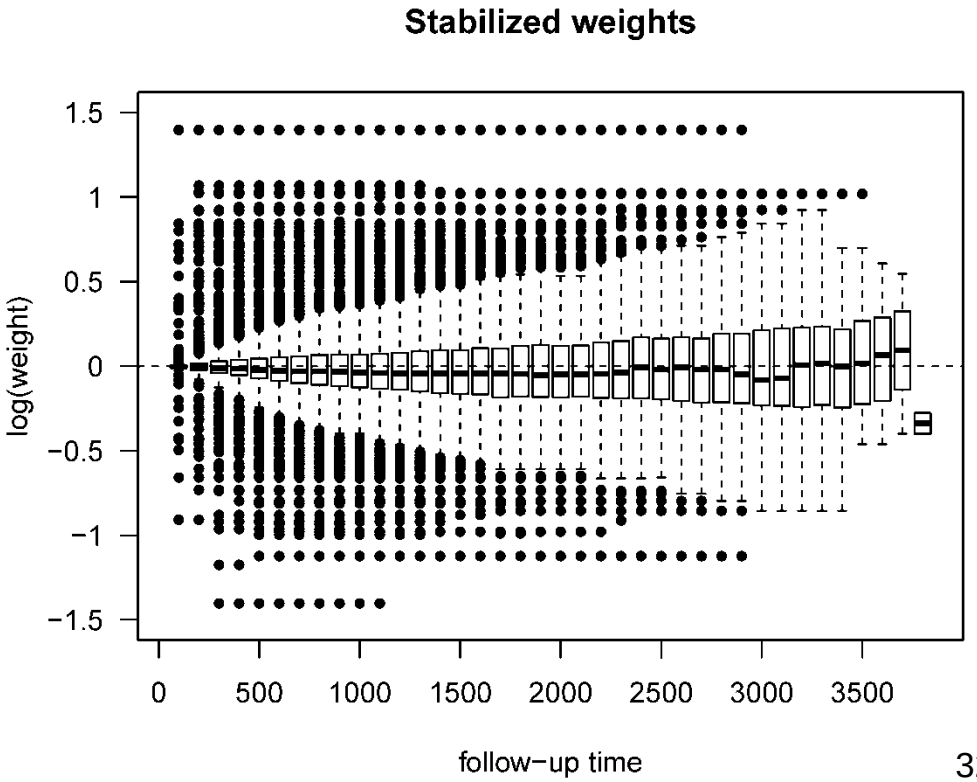
- Also estimate the unstabilized weights for comparison

```
temp.uns <- ipwtm(exposure = haartind, family = "survival",
denominator = ~ cd4.sqrt + sex + age, id = patient, tstart =
tstart,timevar = fuptime, type = "first", data = haartdat)
```

HAART and HIV survival – estimating the weights

- The distribution of the weights over time can be displayed by ipwplot()

```
ipwplot(weights = temp$ipw.weights, timevar = haartdat$fuptime,
binwidth = 100, ylim = c(-1.5, 1.5), main = "Stabilized weights",
xaxt = "n", yaxt = "n")
axis(side = 1, at = 0:7*5,
labels = 0:7*500)
axis(side = 2, at = -3:3*0.5,
labels = -3:3*0.5)
```



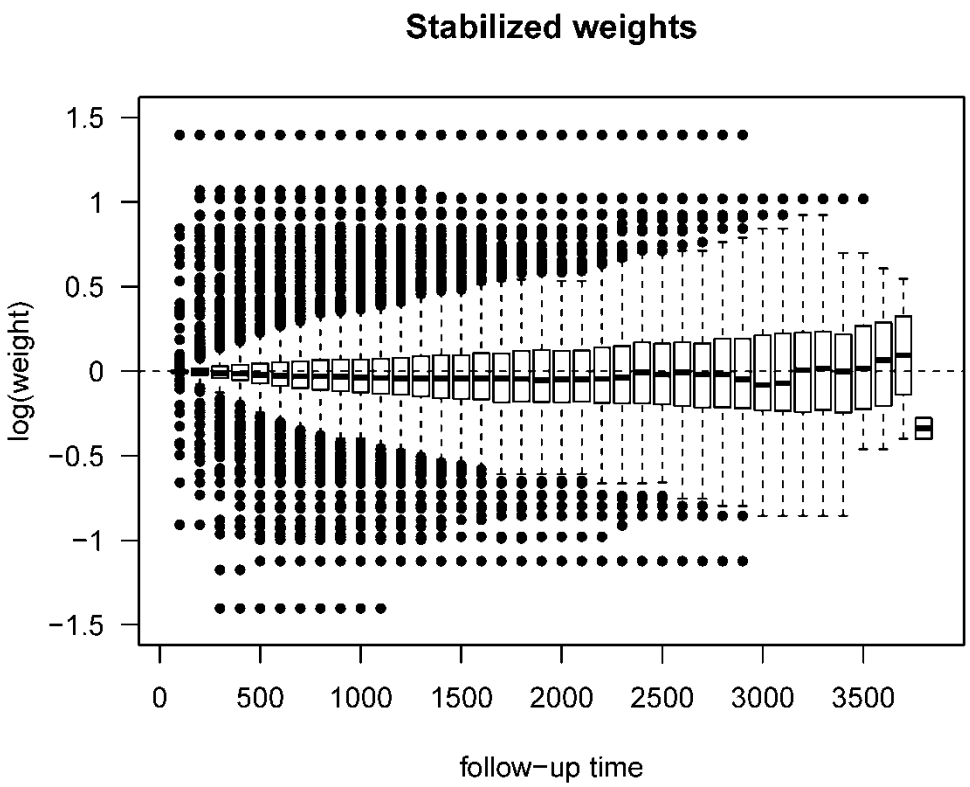
HAART and HIV survival – estimating the weights

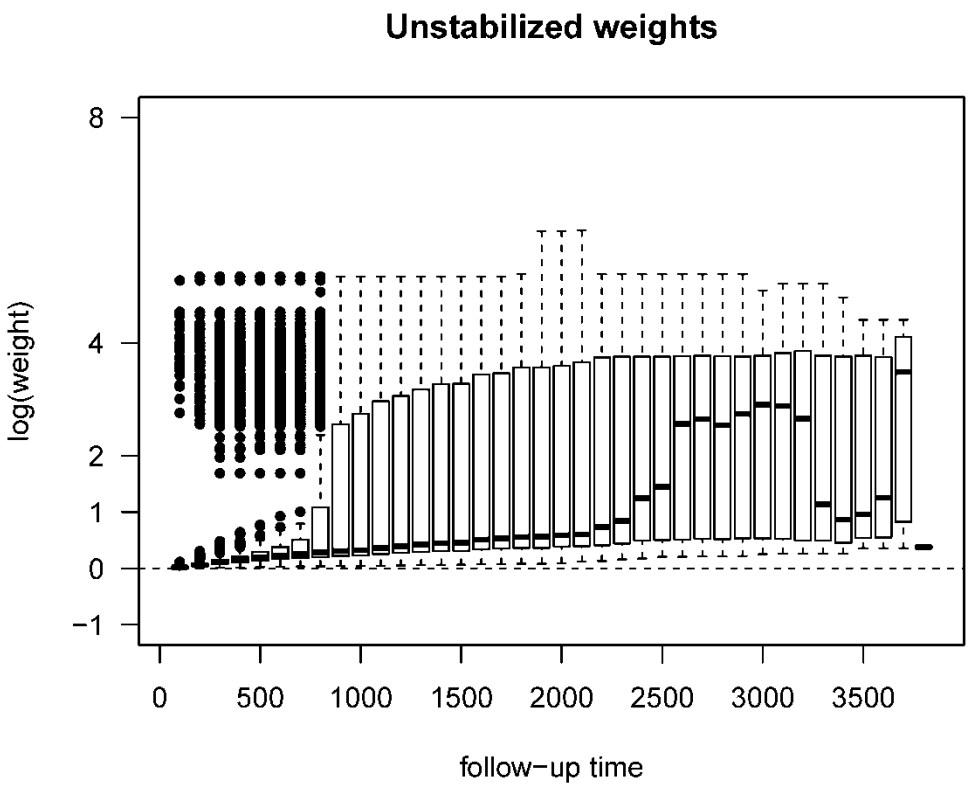
- Comparing stabilized vs unstabilized weights

```
summary(temp$ipw.weights)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.2459  0.9036  0.9862  1.0390  1.0611  7.1257
summary(temp.uns$ipw.weights)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.003   1.160   1.372  13.161  15.401 406.338
```

- Unstabilized weights have a much larger variation

HAART and HIV survival – stabilized weights





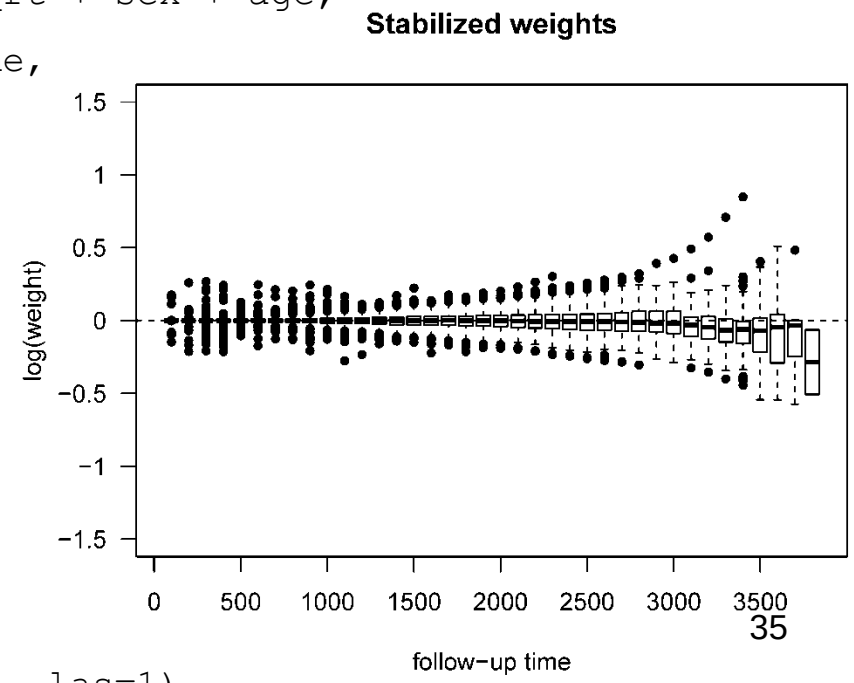
- Unstabilized weights have a much larger variation

HAART and HIV survival – estimating the weights

- The previous weights adjusted for endogenous confounding
- However, there may also be informative censoring according to CD4 counts
- An additional weight can be used to account for that

```
temp2 <- ipwtm(exposure = dropout, family = "survival",
numerator = ~ sex + age, denominator = ~ cd4.sqrt + sex + age,
id = patient, tstart = tstart, timevar = fuptime,
type = "first", data = haartdat)

ipwplot(weights = temp2$ipw.weights,
timevar = haartdat$fuptime, binwidth = 100,
ylim = c(-1.5, 1.5),
main = "Stabilized weights", xaxt = "n",
yaxt = "n", xlab="follow-up time",
ylab="log(weight)")
axis(side = 1, at = 0:7*5, labels = 0:7*500)
axis(side = 2, at = -3:3*0.5, labels = -3:3*0.5, las=1)
```



HAART and HIV survival – fitting MSM

```
require(survival)
summary(coxph(Surv(tstart, fuptime, event) ~ haartind +
cluster(patient), data = haartdat,
weights = temp$ipw.weights*temp2$ipw.weights))
           coef exp(coef) se(coef) robust se      z Pr(>|z|)
haartind -0.9378    0.3915   0.4300    0.4524 -2.073   0.0382 *

         exp(coef) exp(-coef) lower .95 upper .95
haartind    0.3915      2.554    0.1613    0.9501
```

- Robust (sandwich) standard error was used to account for the correlation among patient
  - Provided when using coxph with cluster() to specify correlated observations
- The hazard ratio of patients receiving HAART was 0.39 (95% CI 0.16–0.95)
  - The estimated effect is stronger than that from the Cox regression

Fitting MSM using the unstabilized weights

- Compare the results from the model using stabilized weights

```
temp2.uns <- ipwtm(exposure = dropout, family = "survival",
denominator = ~ cd4.sqrt + sex + age, id = patient, tstart =
tstart, timevar = fuptime, type = "first", data = haartdat)
summary(coxph(Surv(tstart, fuptime, event) ~ haartind +
cluster(patient), data = haartdat, weights =
temp.uns$ipw.weights*temp2.uns$ipw.weights))
```

```
           coef exp(coef) se(coef) robust se     z Pr(>|z|)
haartind 0.2929    1.3403   0.1213    0.5953 0.492    0.623

         exp(coef) exp(-coef) lower .95 upper .95
haartind     1.381      0.724    0.4314     4.422
```

- The 95% CI is much wider

Model diagnostics

- Unconfoundedness between treatment and predictors
  - Not testable
  - More likely by identifying and including relevant confounders from the literature (diligently)
  - Perform sensitivity analysis by including other potential confounders
- Experimental treatment assignment / structural positivity
  - Plot the distribution of the stabilized weight
  - Extreme values indicate potential violation of positivity
  - Mean value far away from 1 indicates model mis-specification

Review

- Marginal structural model estimates the causal effects of treatment/exposure for longitudinal data, accounting for endogenous confounders
- Inverse probability of treatment weighed (IPTW) estimator is used to recover the ‘full’ dataset including counterfactuals
- Robust standard error are used to account for correlation induced by IPW and/or within-subject correlation

MSM example – Obesity and smoking effect on mortality

American Journal of Epidemiology

Vol. 183, No. 2
DOI: 10.1093/aje/kwv168
Advance Access publication:
December 10, 2015

Original Contribution

Estimating the Time-Varying Joint Effects of Obesity and Smoking on All-Cause Mortality Using Marginal Structural Models

Hailey R. Banack* and Jay S. Kaufman

* Correspondence to Hailey R. Banack, Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, 1020 Pine Avenue West Montreal, Quebec, Canada H3A 1A2 (e-mail: hailey.banack@mail.mcgill.ca).

- Banack et al., AJE, 2016
- Objective: to estimate the joint effects of obesity and smoking on all-cause mortality
- Study design: prospective longitudinal study
- Subjects: 15,792 men and women aged 45–64 years old in the US

MSM example – Obesity and smoking effect on mortality

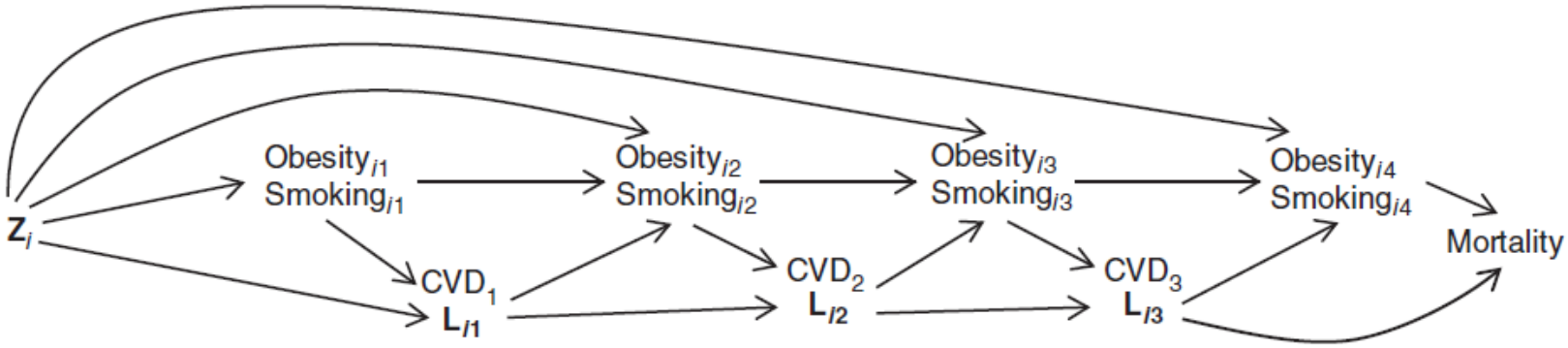


Figure 1. Directed acyclic graph for the effect of obesity and smoking on all-cause mortality. $\mathbf{Z}_i$ is a vector of baseline covariates; $\mathbf{L}_{i1}$, $\mathbf{L}_{i2}$, and $\mathbf{L}_{i3}$ represent vectors of time-varying covariates for individual *i* at cohort visits 1, 2, and 3, respectively. CVD, cardiovascular disease.

- Exposure: Obesity and smoking measured at 1987, 1990, 1993 and 1996
- Covariates measured at baseline: age, race, sex, number of years of education completed, and smoking history
- Other time-varying covariates: current alcohol drinking status, CVD status, diabetes status, and physical activity level
- Note that CVD is a confounder (e.g. Smoking$_4$ ← CVD$_3$ → Mortality) as well as a mediator (e.g. Smoking$_3$ → CVD$_3$ → Mortality)

MSM example – Obesity and smoking effect on mortality

Statistical analysis

To examine the obesity-mortality relationship among smokers and nonsmokers, we used 2 unweighted, stratified Poisson regression models. These models were stratified by smoking status to explore whether the obesity-mortality relationship differs between ever smokers and never smokers. We then fit a crude regression model and a standard adjusted Poisson regression model. Finally, we fit a joint marginal structural model to estimate the joint effects of obesity and smoking on all-cause mortality. Fitting this model is a 2-step process.

Step 1: fitting inverse probability of treatment weights

$$\mathrm{SW}^{X1}{}_{(t)} = \prod_{k=0}^{j} \frac{f\left[X_1(k_{(m)}) | X_1(k_{(m-1)}), \overline{X}_2(k_{(m)}), \mathbf{Z}_1, \overline{C}_{i(k(m))=0}\right]}{f\left[X_1(k_{(m)}) | X_1(k_{(m-1)}), \overline{X}_2(k_{(m)}), \overline{\boldsymbol{L}}_1(k_{(m-1)}), \overline{C}_{i(k(m))=0}\right]}$$

and

$$\mathrm{SW}^{X2}{}_{(t)} = \prod_{k=0}^{j} \frac{f\left[X_2(k_{(m)}) | X_2(k_{(m-1)}), \overline{X}_1(k_{(m-1)}), \mathbf{Z}_2, \overline{C}_{i(k(m))=0}\right]}{f\left[X_2(k_{(m)}) | X_2(k_{(m-1)}), \overline{X}_1(k_{(m-1)}), \overline{\boldsymbol{L}}_2(k_{(m-1)}), \overline{C}_{i(k(m))=0}\right]}.$$

$$\mathrm{SW}(t) = \mathrm{SW}^{X1}{}_{(t)} \mathrm{SW}^{X2}{}_{(t)}.$$

- A joint marginal structure model was fitted
- Two series of stabilized weights were used for obesity and smoking
- Final IPTW as the product of the two weights (assuming no interaction)

MSM example – Obesity and smoking effect on mortality

Step 2: joint marginal structural model

We estimated the parameters of a joint Poisson marginal structural model with a robust variance estimator. The number of days between cohort visits was included as an offset term:

$$\log\left(\Pr(Y|X_1, X_2\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \beta_4 Z.$$

Baseline confounders (Z) of the obesity-mortality and smoking-mortality relationships were also included in the final marginal structural model. Assuming no unmeasured confounding, informative censoring, structural positivity violations, or model misspecification, the estimates from this model correspond to the average causal incidence rate ratio, estimating the contrast in average outcomes between exposure groups (25, 27). This is a marginal effect with respect to the time-dependent covariates because it expresses the change in average outcome at the population level if everyone were set to one level of exposure versus if everyone were set to another level (35).

- Robust variance estimator was used
- Estimate the average causal effect if the assumptions are valid
- The marginal / population level effect was estimated

MSM example – Obesity and smoking effect on mortality

Table 3. Incidence Rate Ratios for the Joint Effects of Obesity and Smoking on Mortality Among Men and Women, Atherosclerosis Risk in Communities Study, 1987–2007

Smoking Status and Obesity Category	Crude Model		Adjusted Model		Weighted Model	
	IRR	95% CI	IRR	95% CI	IRR	95% CI
Never smoker						
Nonobese	1.00	Referent	1.00	Referent	1.00	Referent
Obese	1.50	1.32, 1.70	1.10	0.97, 1.26	1.31	1.13, 1.51
Smoker						
Nonobese	2.23	2.02, 2.45	2.09	1.89, 2.32	2.00	1.79, 2.24
Obese	2.31	2.05, 2.56	1.65	1.46, 1.85	1.97	1.73, 2.22
P value for product term		<0.001		<0.001		<0.001
RERI	−0.42	−0.68, −0.16	−0.54	−0.76, −0.31	−0.34	−0.60, −0.07

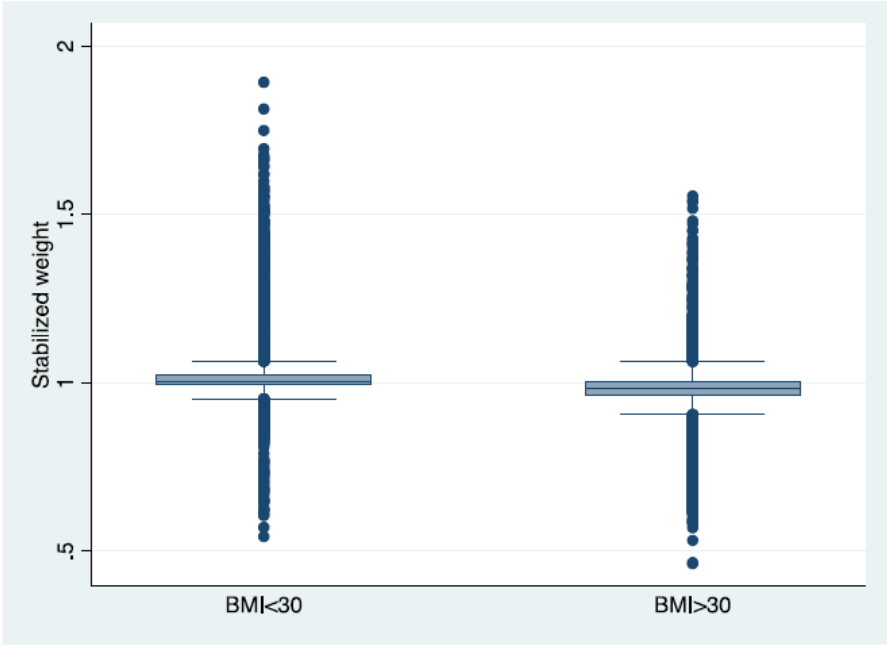
Abbreviations: CI, confidence interval; IRR, incidence rate ratio; RERI, relative excess risk of interaction.

- Weighted model refers to MSM
- MSM estimated a relatively higher effect of obesity than from adj. model
  - Possible reason: adjusted for CVD, but obesity → CVD → mortality

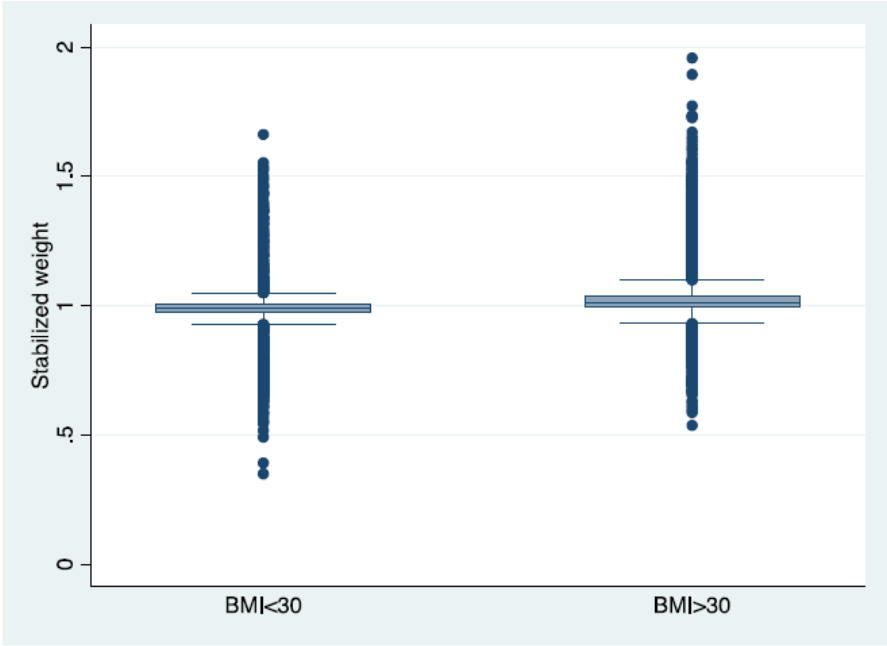
MSM example – Obesity and smoking effect on mortality

Web Figure 3: Boxplots of stabilized inverse probability of treatment weight by exposure group for obesity

Visit 2:



Visit 4:

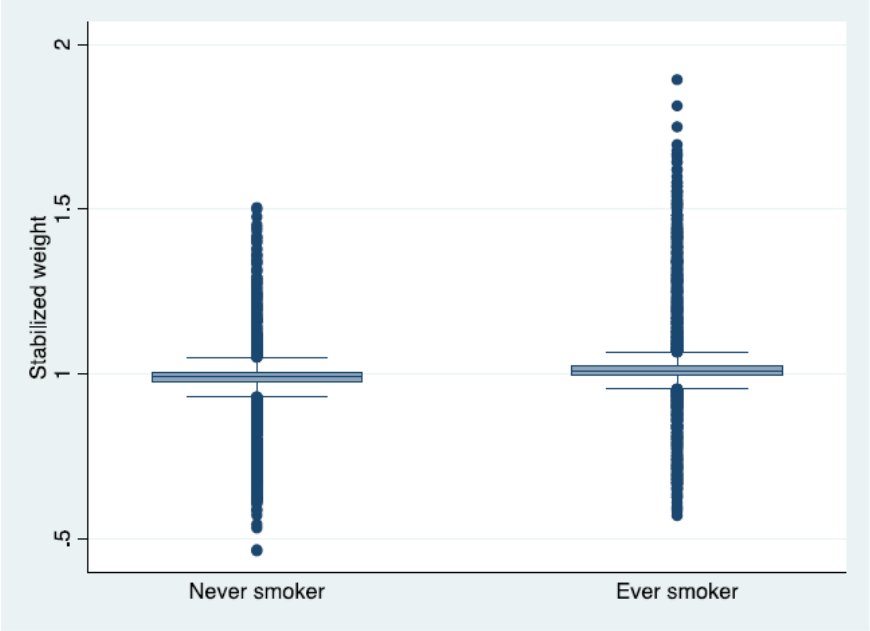


- No obvious association between obesity and treatment probability

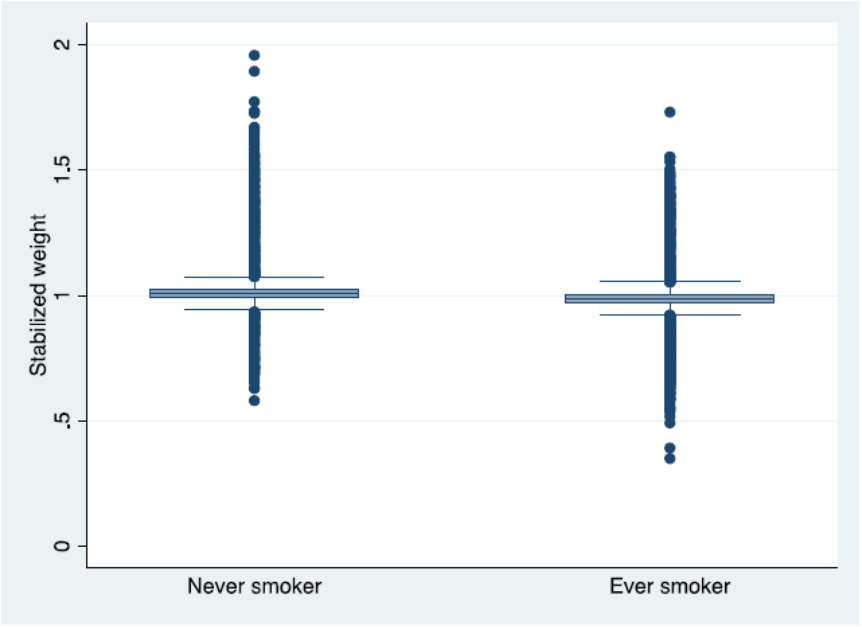
MSM example – Obesity and smoking effect on mortality

Boxplots of stabilized inverse probability of treatment weight by exposure group for smoking:

Visit 2:



Visit 4:



- No obvious association between smoking and treatment probability

References

- Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. Am J Epidemiol. 2008 Sep 15;168(6):656-64.

- Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology. 2000 Sep;11(5):550-60.