Are Marginal Structural Models useful to appropriately control the Healthy Worker Survivor Effect for occupational epidemiological studies?

Alex Keil

The University of North Carolina at Chapel Hill

Brit. J. prev. soc. Med. (1976), **30,** 225-230

Low mortality rates in industrial cohort studies due to selection for work and survival in the industry

A. J. FOX AND P. F. COLLIER

Office of Population Censuses and Surveys and the Employment Medical Advisory Service, Health and Safety Executive, London

Fox, A. J. and Collier, P. F. (1976). *British Journal of Preventive and Social Medicine,* 30, 225-230. **Low mortality rates in industrial cohort studies due to selection for work and survival in the industry.** Occupational groups are often described as being relatively healthy because their mortality rates are lower than those of the national average. Although

Healthy Worker Effect

Healthy Worker Effect has two main components

Healthy hire effect: initial selection into occupational cohort of healthy individuals

Healthy Worker Survivor Effect: continual selection over the time scale of an epidemiologic study.

Healthy Worker Effect

Healthy Worker Effect has two main components

Healthy hire effect
: initial selection into occupational cohort of healthy individuals

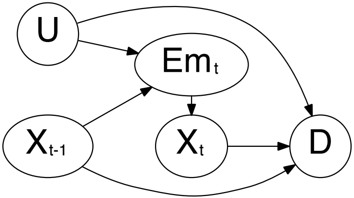
Healthy Worker Survivor Effect
: continual selection over the time scale of an epidemiologic study. Causes bias when

- Selection (employment status) depends on health status
- Selection (employment status) depends on occupational exposures or duration of employment

Healthy Worker Survivor Effect

For example, let

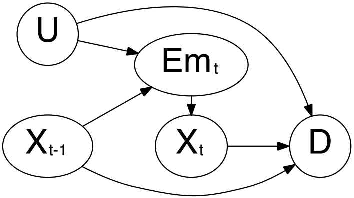
- X_t be total arsenic exposure in year t $(mg/m^3 - year)$
- Em_t be an indicator of work status (1=at work, 0=off work) and
- D be the time to death from lung cancer.
- U is an unknown baseline determinant of underlying health status (e.g. baseline smoking)



Healthy Worker Survivor Effect

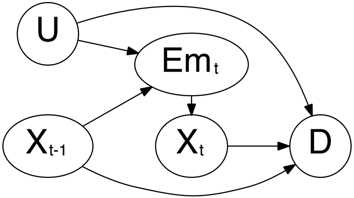
For example, let

- X_t be total arsenic exposure in year t $(mg/m^3 - year)$
- Em_t be an indicator of work status (1=at work, 0=off work) and
- D be the time to death from lung cancer.
- U is an unknown baseline determinant of underlying health status (e.g. baseline smoking)



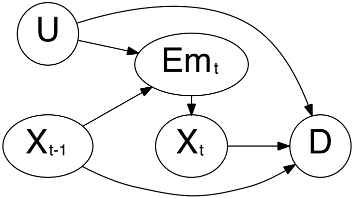
Healthy Worker Survivor Effect

- We want to measure the effect of cumulative arsenic exposure (X) on the time to die of lung cancer (D), but we have the time varying confounder Em_t



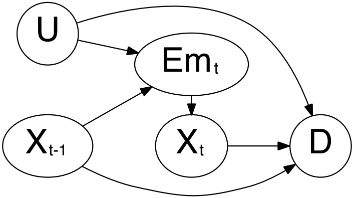
Healthy Worker Survivor Effect

- We want to measure the effect of cumulative arsenic exposure (X) on the time to die of lung cancer (D), but we have the time varying confounder Em_t
- Two proposed approaches to adjusting for time varying confounding:
  - We can include Em_t in a regression model
  - We can use Marginal Structural Models (G-methods) to adjust for Em_t



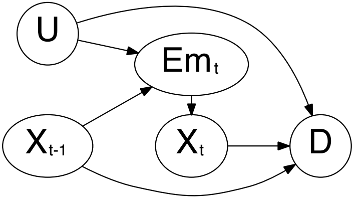
Marginal Structural Models

- Developed by James Robins[1]



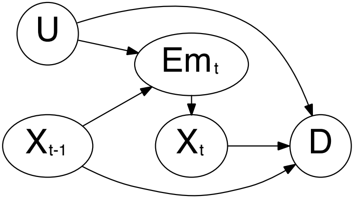
Marginal Structural Models

- Developed by James Robins[1]
- Consistently estimate causal effects under certain assumptions



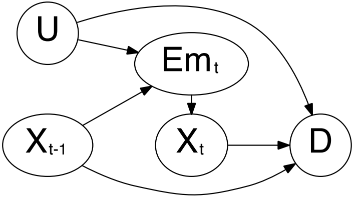
Marginal Structural Models

- Developed by James Robins[1]
- Consistently estimate causal effects under certain assumptions
- For example, MSMs consistently estimate causal effect $\bar{X} \rightarrow D$, adjusted for time-varying confounder Em_t without inducing additional bias by adjusting for a causal intermediate



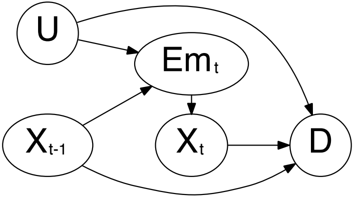
Marginal Structural Models

- Inverse Probability of Treatment (Exposure) Weighting can be used to estimate MSMs using standard statistical software
- For binary exposures, the weight at time t, $W_t = \Pr(X_t = x_t | \bar{X}_{t-1}, \bar{Em}_t)^{-1}$



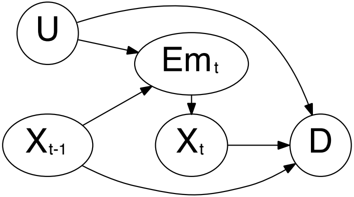
Marginal Structural Models

- Inverse Probability of Treatment (Exposure) Weighting can be used to estimate MSMs using standard statistical software
- For binary exposures, the weight at time t, $W_t = \Pr(X_t = x_t | \bar{X}_{t-1}, \bar{Em}_t)^{-1}$
- W_t is used as a subject- and time-specific weight in a regression model

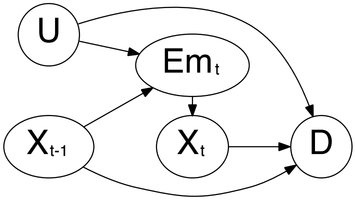


Marginal Structural Models

- Inverse Probability of Treatment (Exposure) Weighting can be used to estimate MSMs using standard statistical software
- For binary exposures, the weight at time t, $W_t = \Pr(X_t = x_t | \bar{X}_{t-1}, \bar{Em}_t)^{-1}$
- W_t is used as a subject- and time-specific weight in a regression model
- Weighting by W_t yields a "pseudo-population" in which $\Pr(X_t) = \Pr(X_t | Em_t)$, and $\Pr(D|X_t)$ is unchanged

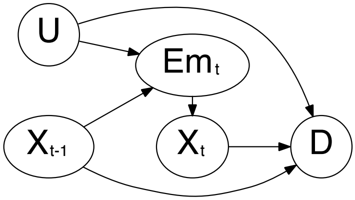


Marginal Structural Models

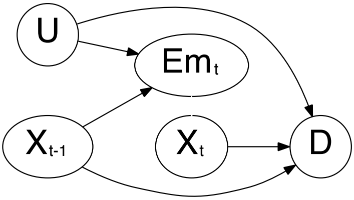


Original population

Marginal Structural Models



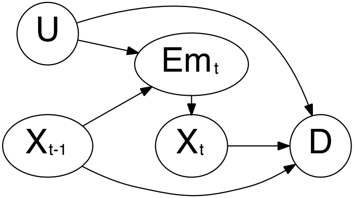
Original population



Weighted “Pseudo population”

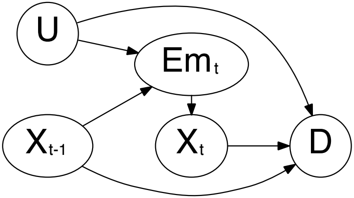
Positivity Assumption

- But if Arsenic exposure (X_t) cannot happen off work ($Em_t = 0$), then $\Pr(X_t = 1 \,|\bar{X}_{t-1}, \bar{Em}_t) = 0$



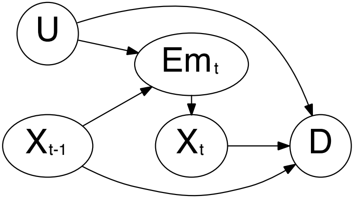
Positivity Assumption

- But if Arsenic exposure (X_t) cannot happen off work ($Em_t = 0$), then $\Pr(X_t = 1 \mid \bar{X}_{t-1}, \bar{Em}_t) = 0$, and $W_t = \frac{1}{0} = \infty$



Positivity Assumption

- But if Arsenic exposure (X_t) cannot happen off work ($Em_t = 0$), then $\Pr(X_t = 1 \mid \bar{X}_{t-1}, \bar{Em}_t) = 0$, and $W_t = \frac{1}{0} = \infty$
- The probability of exposure must be positive at all levels of confounders, hence it is known as the positivity assumption



Positivity Assumption - 1 time point

- Single time point
  (Y: 1=dead, 0=alive)

Em	X	Y	N Observed Population	$pr(X\|Em)$	W
0	0	1	240	1	1
0	0	0	160	1	1
1	1	1	20	0.5	2
1	1	0	30	0.5	2
1	0	1	40	0.5	2
1	0	0	10	0.5	2

Adapted from Robins et al 2000

Positivity Assumption - 1 time point

- Single time point (Y: 1=dead, 0=alive)
- Positivity violated for those exposed off work (N=0)

Em	X	Y	N Observed Population	$pr(X\|Em)$	W
0	1	1	0	0	
0	1	0	0	0	
0	0	1	240	1	1
0	0	0	160	1	1
1	1	1	20	0.5	2
1	1	0	30	0.5	2
1	0	1	40	0.5	2
1	0	0	10	0.5	2

Adapted from Robins et al 2000

Positivity Assumption - 1 time point

- Single time point
  (Y: 1=dead, 0=alive)
- Positivity violated for those exposed off work (N=0)
- Weights are infinite for those exposed off work

Em	X	Y	N Observed Population	$pr(X\|Em)$	W
0	1	1	0	0	∞
0	1	0	0	0	∞
0	0	1	240	1	1
0	0	0	160	1	1
1	1	1	20	0.5	2
1	1	0	30	0.5	2
1	0	1	40	0.5	2
1	0	0	10	0.5	2

Adapted from Robins et al 2000

Positivity Assumption - 1 time point

- Single time point (Y: 1=dead, 0=alive)
- Positivity violated for those exposed off work (N=0)
- Weights are infinite for those exposed off work
- Cannot estimate MSM because the positivity assumption is violated

Em	X	Y	N Observed Population	$pr(X \mid Em)$	W
0	1	1	0	0	∞
0	1	0	0	0	∞
0	0	1	240	1	1
0	0	0	160	1	1
1	1	1	20	0.5	2
1	1	0	30	0.5	2
1	0	1	40	0.5	2
1	0	0	10	0.5	2

Adapted from Robins et al 2000

Positivity Assumption - 1 time point

- What if we treat $\Pr(X = 1|Em = 0)$ as random zeroes due to sparse data?

Em	X	Y	N Observed Population	$\Pr(X|Em)$	W

Adapted from Robins et al 2000

Positivity Assumption - 1 time point

- What if we treat $\Pr(X = 1|Em = 0)$ as random zeroes due to sparse data?
- All weights are non-infinite - i.e. regression software can estimate them

Em	X	Y	N Observed Population	$\Pr(X\|Em)$	W
0	1	1	0	~ 0	*big*
0	1	0	0	~ 0	*big*

Adapted from Robins et al 2000

Positivity Assumption - 1 time point

- What if we treat $\Pr(X = 1|Em = 0)$ as random zeroes due to sparse data?
- All weights are non-infinite - i.e. regression software can estimate them
- Then we can use W to estimate the $X \rightarrow Y$ association in a marginal structural model

Em	X	Y	N Observed Population	$\Pr(X\|Em)$	W
0	1	1	0	~ 0	*big*
0	1	0	0	~ 0	*big*
0	0	1	240	~ 1	~ 1
0	0	0	160	~ 1	~ 1
1	1	1	20	0.5	2
1	1	0	30	0.5	2
1	0	1	40	0.5	2
1	0	0	10	0.5	2

Adapted from Robins et al 2000

Methods

Goal: Estimate a Marginal Structural Cox Model when weights are infinite

Methods

Goal: Estimate a Marginal Structural Cox Model when weights are infinite

- Use Firth's logistic regression for IPTW estimation

Methods

Goal: Estimate a Marginal Structural Cox Model when weights are infinite

- Use Firth's logistic regression for IPTW estimation

Firth's logistic regression: Modification of the score function to estimate $\Pr(X_t|Em_t)$ in a logistic model for the weights. Converges to the same answer as the "add 0.5 to each cell" method in a 2X2 table.[2,3]

Methods

Goal: Estimate a Marginal Structural <u>Cox</u> Model when weights are infinite

- Use Firth's logistic regression for IPTW estimation

  Firth's logistic regression Modification of the score function to estimate $\Pr(X_t|Em_t)$ in a logistic model for the weights. Converges to the same answer as the "add 0.5 to each cell" method in a 2X2 table.[2,3]

- Monte-Carlo simulations (10000 iterations, N=1500)

Methods

Goal: Estimate a Marginal Structural Cox Model when weights are infinite

- Use Firth's logistic regression for IPTW estimation

  Firth's logistic regression: Modification of the score function to estimate $\Pr(X_t|Em_t)$ in a logistic model for the weights. Converges to the same answer as the "add 0.5 to each cell" method in a 2X2 table.[2,3]

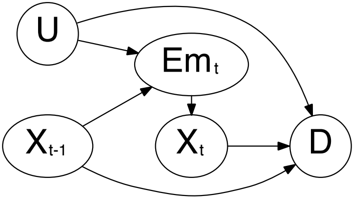
- Monte-Carlo simulations (10000 iterations, N=1500)
- Simulated HWSE by disallowing off-work exposure

Methods

Crude Crude Cox PH model for the association between exposure and the outcome with no covariates

Adjusted Crude model with the duration of employment included as a linear term

MSM Same as crude model, but performed in a population in which there are W_t copies of each individual at each time point



Results

Method	Time points	True HR	HR_{MC}	$Bias_{\beta}$	SD_{β}

Results

Method	Time points	True HR	HR_{MC}	$Bias_{\beta}$	SD_{β}
Crude	3	0.5	0.25	-0.69	0.24
Adjusted	3	0.5	0.76	0.42	0.28
Firth MSM	3	0.5	0.34	-0.38	0.25

Results

Method	Time points	True HR	HR_{MC}	$Bias_{\beta}$	SD_{β}
Crude	3	0.5	0.25	-0.69	0.24
Adjusted	3	0.5	0.76	0.42	0.28
Firth MSM	3	0.5	0.34	-0.38	0.25
Crude	3	1	0.48	-0.73	0.23
Adjusted	3	1	1.59	0.46	0.26
Firth MSM	3	1	0.65	-0.43	0.24

Results

Method	Time points	True HR	HR_{MC}	$Bias_{\beta}$	SD_{β}
Crude	3	0.5	0.25	-0.69	0.24
Adjusted	3	0.5	0.76	0.42	0.28
Firth MSM	3	0.5	0.34	-0.38	0.25
Crude	3	1	0.48	-0.73	0.23
Adjusted	3	1	1.59	0.46	0.26
Firth MSM	3	1	0.65	-0.43	0.24
Crude	5	0.5	0.06	-2.09	0.34
Adjusted	5	0.5	2.21	1.49	0.46
Firth MSM	5	0.5	0.14	-1.24	0.40

Results

Method	Time points	True HR	HR_{MC}	$Bias_{\beta}$	SD_{β}
Crude	3	0.5	0.25	-0.69	0.24
Adjusted	3	0.5	0.76	0.42	0.28
Firth MSM	3	0.5	0.34	-0.38	0.25
Crude	3	1	0.48	-0.73	0.23
Adjusted	3	1	1.59	0.46	0.26
Firth MSM	3	1	0.65	-0.43	0.24
Crude	5	0.5	0.06	-2.09	0.34
Adjusted	5	0.5	2.21	1.49	0.46
Firth MSM	5	0.5	0.14	-1.24	0.40
Crude	5	1	0.11	-2.20	0.33
Adjusted	5	1	4.89	1.59	0.46
Firth MSM	5	1	0.26	-1.36	0.38

Example - revisited

- N Pseudo = (N Observed) * W

Em	X	Y	N Observed Population	$\Pr(X \mid Em)$	W	N Pseudo Population

Adapted from Robins et al 2000

Example - revisited

- N Pseudo = (N Observed) * W

Em	X	Y	N Observed Population	$\Pr(X \mid Em)$	W	N Pseudo Population
0	0	1	240	~ 1	~ 1	
0	0	0	160	~ 1	~ 1	
1	1	1	20	0.5	2	
1	1	0	30	0.5	2	
1	0	1	40	0.5	2	
1	0	0	10	0.5	2	

Adapted from Robins et al 2000

Example - revisited

- N Pseudo = (N Observed) * W

Em	X	Y	N Observed Population	$\Pr(X \mid Em)$	W	N Pseudo Population
0	1	1	0	~ 0	*big*	
0	1	0	0	~ 0	*big*	
0	0	1	240	~ 1	~ 1	
0	0	0	160	~ 1	~ 1	
1	1	1	20	0.5	2	40
1	1	0	30	0.5	2	60
1	0	1	40	0.5	2	80
1	0	0	10	0.5	2	20

Adapted from Robins et al 2000

Example - revisited

- N Pseudo = (N Observed) * W
- Pseudo population $(0, 1, y)$: N = 0

Em	X	Y	N Observed Population	$\Pr(X \mid Em)$	W	N Pseudo Population
0	1	1	0	~ 0	*big*	0
0	1	0	0	~ 0	*big*	0
0	0	1	240	~ 1	~ 1	~ 240
0	0	0	160	~ 1	~ 1	~ 160
1	1	1	20	0.5	2	40
1	1	0	30	0.5	2	60
1	0	1	40	0.5	2	80
1	0	0	10	0.5	2	20

Adapted from Robins et al 2000

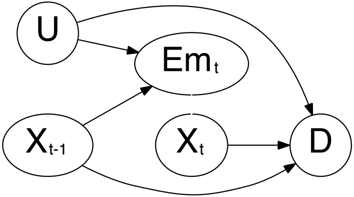
Example - revisited

- N Pseudo = (N Observed) * W
- Pseudo population $(0, 1, y)$: N = 0
- We haven't created the appropriate pseudo population

Em	X	Y	N Observed Population	$\Pr(X \mid Em)$	W	N Pseudo Population
0	1	1	0	~ 0	*big*	0
0	1	0	0	~ 0	*big*	0
0	0	1	240	~ 1	~ 1	~ 240
0	0	0	160	~ 1	~ 1	~ 160
1	1	1	20	0.5	2	40
1	1	0	30	0.5	2	60
1	0	1	40	0.5	2	80
1	0	0	10	0.5	2	20

Adapted from Robins et al 2000

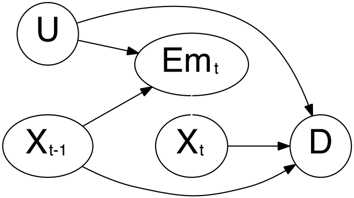
Simulations - revisited



Weighted "Pseudo population"

- Estimation of the MSM can be done in a pseudo-population in which time-varying confounding has been "weighted out"

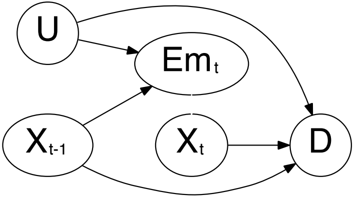
Simulations - revisited



Weighted “Pseudo population”

- Estimation of the MSM can be done in a pseudo-population in which time-varying confounding has been “weighted out”
- i.e. $\Pr(X_t) = \Pr(X_t|Em_t)$ in the pseudo-population

Simulations - revisited



Weighted "Pseudo population"

- Estimation of the MSM can be done in a pseudo-population in which time-varying confounding has been "weighted out"
- i.e. $\Pr(X_t) = \Pr(X_t|Em_t)$ in the pseudo-population
- In our pseudo population, $\Pr(X_t) \neq \Pr(X_t|Em_t)$

In our pseudo population:

$$\Pr(X = 1 | Em = 0)$$

$$\Pr(X = 1 | Em = 1)$$

$$\Pr(X = 1)$$

In our pseudo population:

$\Pr(X = 1|Em = 0)$ $(0 + 0)/(0 + 0 + \sim 240 + \sim 160) = \mathbf{0}$

$\Pr(X = 1|Em = 1)$

$\Pr(X = 1)$

In our pseudo population:

$$\Pr(X = 1|Em = 0) \quad (0+0)/(0+0+\sim 240+\sim 160) = \mathbf{0}$$

$$\Pr(X = 1|Em = 1) \quad (40+60)/(40+60+80+20) = \mathbf{0.5}$$

$$\Pr(X = 1)$$

In our pseudo population:

$\Pr(X = 1|Em = 0)$ $(0 + 0)/(0 + 0 + \sim 240 + \sim 160) = \mathbf{0}$

$\Pr(X = 1|Em = 1)$ $(40 + 60)/(40 + 60 + 80 + 20) = \mathbf{0.5}$

$\Pr(X = 1)$ $(0 + 0 + 40 + 60)/(0 + 0 + 40 + 60 + \sim 240 + \sim 160 + 20 + 80) = \mathbf{0.17}$

In our pseudo population:

$\Pr(X = 1|Em = 0)$ $(0 + 0)/(0 + 0 + \sim 240 + \sim 160) = \mathbf{0}$

$\Pr(X = 1|Em = 1)$ $(40 + 60)/(40 + 60 + 80 + 20) = \mathbf{0.5}$

$\Pr(X = 1)$ $(0 + 0 + 40 + 60)/(0 + 0 + 40 + 60 + \sim 240 + \sim 160 + 20 + 80) = \mathbf{0.17}$

$\Pr(X_t) \neq \Pr(X_t|Em_t)$ Inverse probability of treatment weighting does not work when all individuals off work are considered "unexposed"

Summary and future directions

- MSMs using sparse data methods to derive weights do not minimize the bias due to the Healthy Worker Survivor Effect
- This does not rule out other potential avenues for dealing non-positivity
- G-estimation of structural nested models is one alternative that does not rely on positivity

Summary and future directions

- MSMs using sparse data methods to derive weights do not minimize the bias due to the Healthy Worker Survivor Effect
- This does not rule out other potential avenues for dealing non-positivity
- G-estimation of structural nested models is one alternative that does not rely on positivity

Summary and future directions

- MSMs using sparse data methods to derive weights do not minimize the bias due to the Healthy Worker Survivor Effect
- This does not rule out other potential avenues for dealing non-positivity
- G-estimation of structural nested models is one alternative that does not rely on positivity
- Future directions:
  - Non-positivity in MSMs for continuous exposures
  - Developing non MSM based causal methods to adjust for HWSE drawing from literature on direct effects

References

[1] J.M. Robins. Marginal structural models. In *Proceedings of the American Statistical Association*, pages 1–10, 1997.

[2] D. Firth. Bias reduction of maximum likelihood estimates. *Biometrika*, 80(1):27, 1993.

[3] G. Heinze. The application of firth's procedure to cox and logistic regression. Technical report, Technical Report 10/1999, update in January 2001, Section of Clinical Biometrics, Department of Medical Computer Sciences, University of Vienna, 1999.

Acknowledgements

- David Richardson
- Stephen Cole

National Institute for Occupational Safety and Health
NIOSH