ibs.GRANADA
INSTITUTO DE INVESTIGACIÓN BIOSANITARIA

GOBIERNO DE ESPAÑA
MINISTERIO DE ECONOMÍA, INDUSTRIA Y COMPETITIVIDAD
Instituto de Salud Carlos III
UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
"Una manera de hacer Europa"

Pre-conference Workshop: Introduction to Causal Inference for Epidemiologists (II). XXXVII-SEE, Oviedo, Spain 2019

Miguel Angel Luque Fernandez

https://maluque.netlify.com/
https://ccci.netlify.com/

PART II: ESTIMATION and INFERENCE

ESTIMANDS, ESTIMATORS, and STATISTICAL INFERENCE

ESTIMANDS

Most common ESTIMANDS

https://migariane.github.io/DeltaMethodEpi.nb.html

ESTIMANDS

Most common ESTIMANDS

https://migariane.github.io/DeltaMethodEpi.nb.html

Estimands in Causal Inference

- RD = EY(1) - EY(0)

ESTIMANDS

Most common ESTIMANDS

https://migariane.github.io/DeltaMethodEpi.nb.html

Estimands in Causal Inference

- RD = EY(1) - EY(0)
- RR = EY(1) / EY(0)

ESTIMANDS

Most common ESTIMANDS

https://migariane.github.io/DeltaMethodEpi.nb.html

Estimands in Causal Inference

- RD = EY(1) - EY(0)
- RR = EY(1) / EY(0)
- MOR = EY(1)x E(1 - Y(0)) / EY(0)xE(1-Y(1))

ESTIMANDS

Most common ESTIMANDS

https://migariane.github.io/DeltaMethodEpi.nb.html

Estimands in Causal Inference

- RD = EY(1) - EY(0)
- RR = EY(1) / EY(0)
- MOR = EY(1)x E(1 - Y(0)) / EY(0)xE(1-Y(1))

G-Formula, (Robins, 1986)

G-Formula for the **identification** of the ATE with observational data

$$
\begin{aligned}
E(Y^a) &= \sum_y E(Y^a \mid W = w) P(W = w) \\
&= \sum_y E(Y^a \mid A = a, W = w) P(W = w) \text{ by consistency} \\
&= \sum_y E(Y = y \mid A = a, W = w) P(W = w) \text{ by ignorability}
\end{aligned}
$$

The **ATE**=

$$
\sum_{\mathbf{w}} \left[\sum_{\mathbf{y}} \mathbf{P(Y = y \mid A = 1, W = w)} - \sum_{\mathbf{y}} \mathbf{P(Y = y \mid A = 0, W = w)} \right] \mathbf{P(W = w)}
$$

$$
P(W = w) = \sum_{y,a} P(W = w, A = a, Y = y)
$$

G-Formula, (Robins, 1986)

G-Formula for the identification of the ATE with observational data

The **ATE**=

$$\sum_{\mathbf{w}} \left[\sum_{\mathbf{y}} \mathbf{P(Y = y \mid A = 1, W = w)} - \sum_{\mathbf{y}} \mathbf{P(Y = y \mid A = 0, W = w)} \right] \mathbf{P(W = w)}$$

$$P(W = w) = \sum_{y,a} P(W = w, A = a, Y = y)$$

G-Formula

- The sums is generic notation. In reality, likely involves sums and integrals (we are just integrating out the W's).
- The **g-formula** is a **generalization of standardization** and allow to estimate unbiased treatment effect estimates.

G-Formula, (Robins, 1986)

G-Formula for the identification of the ATE (Estimand) with observational data

The **ATE**=

$$\sum_{\mathbf{w}} \left[\sum_{\mathbf{y}} \mathbf{P(Y = y \mid A = 1, W = w)} - \sum_{\mathbf{y}} \mathbf{P(Y = y \mid A = 0, W = w)} \right] \mathbf{P(W = w)}$$

$$P(W = w) = \sum_{y,a} P(W = w, A = a, Y = y)$$

G-Formula

- The sums is generic notation. In reality, likely involves sums and integrals (we are just integrating out the W's).
- The **g-formula** is a non-parametric **generalization of standardization** and allow to estimate unbiased treatment effect estimates.

G-Formula, (Robins, 1986)

G-Formula for the identification of the ATE (Estimand) with observational data

The **ATE**=

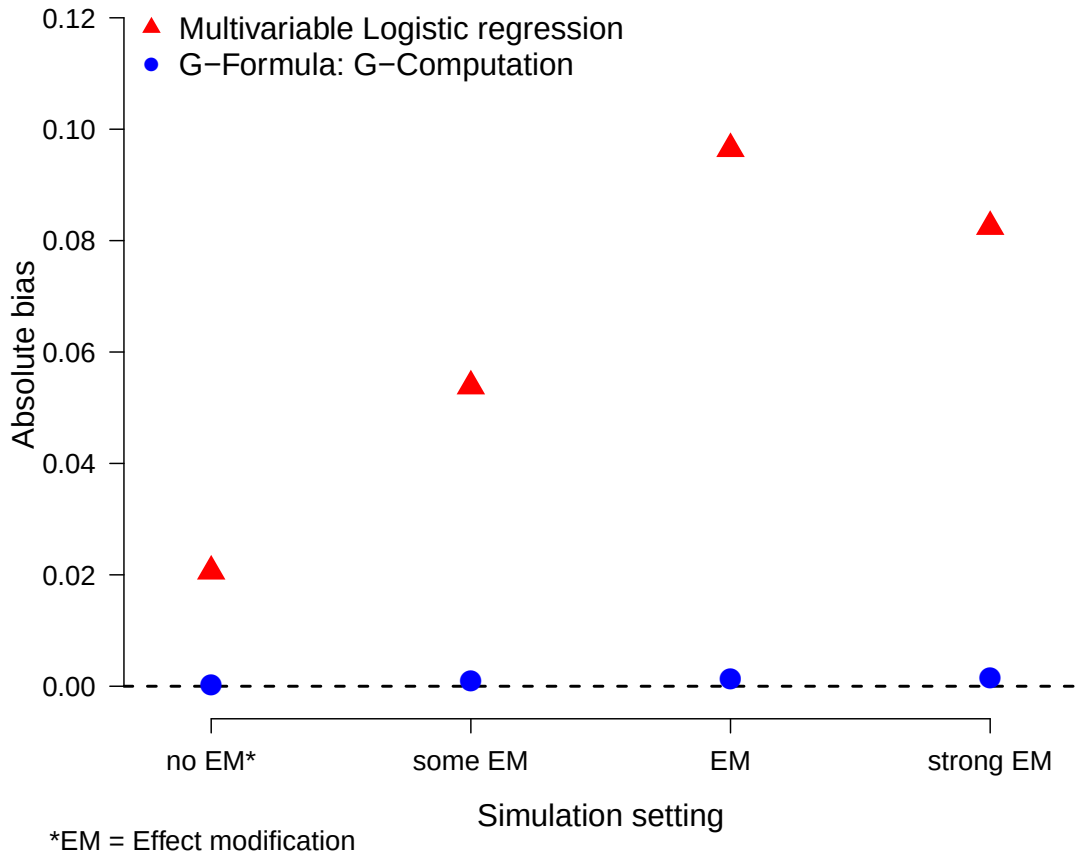
$$\sum_{\mathbf{w}} \left[\sum_{\mathbf{y}} \mathbf{P(Y = y \mid A = 1, W = w)} - \sum_{\mathbf{y}} \mathbf{P(Y = y \mid A = 0, W = w)} \right] \mathbf{P(W = w)}$$

$$P(W = w) = \sum_{y,a} P(W = w, A = a, Y = y)$$

G-Formula

- In the presence of **treatment or exposure effect heterogeneity** (i.e. effect modification) simple conditional estimate is biased.
- The **g-formula** allows to estimate the **marginal average treatment effect** (i.e., effect for the population averaged across the different level of covariates) allowing to estimate unbiased treatment effect estimates.

Treatment heterogeneity



Simulation treatment heterogeneity and estimation using OR and MOR, n = 1,000.

Luque-Fernandez et al.(2019) Effect modification and collapsibility when estimating the effect of public health interventions: A Monte Carlo simulation comparing classical multivariable regression adjustment versus the G-Formula, based on a cancer epidemiology illustration. AJPH https://maluque.netlify.com/project/hetmor/

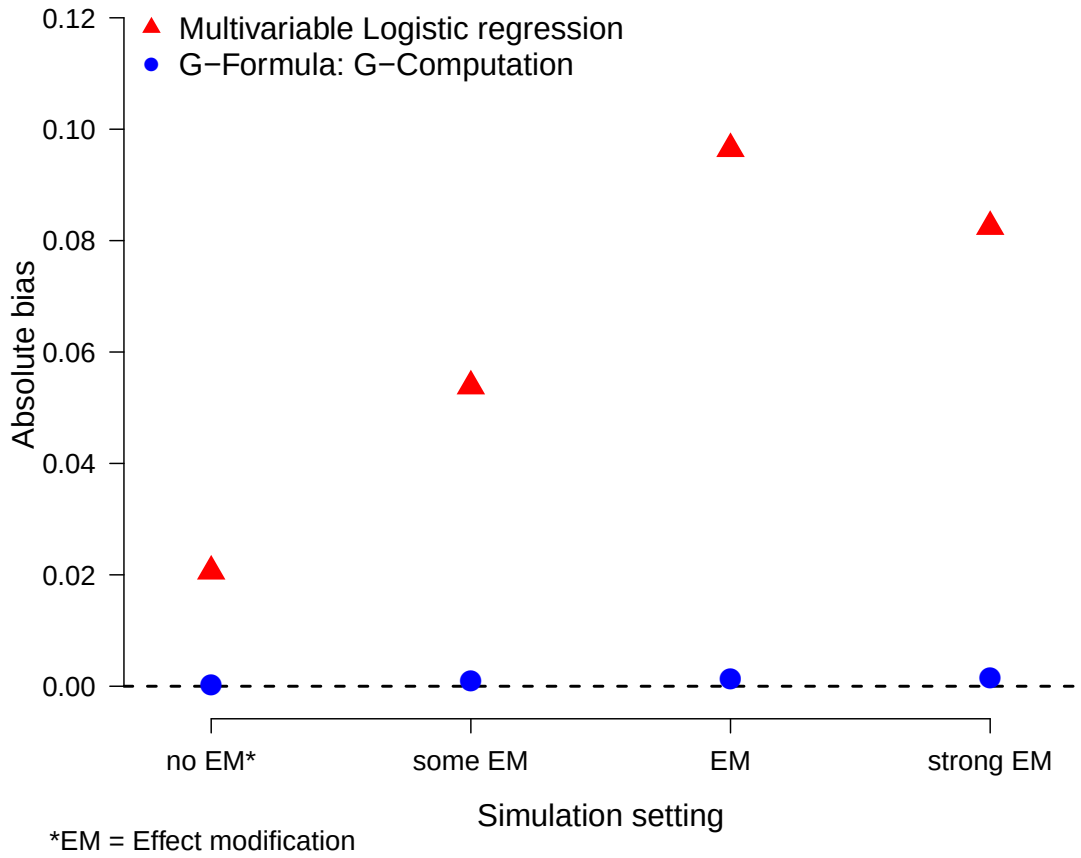
COR

$$COR = \frac{\frac{P(Y=1|A=1,W)}{(1-P(Y=1|A=1,W))}}{\frac{P(Y=1|A=0,W)}{(1-P(Y=1|A=0,W))}}$$

MOR

$$\text{MOR (G-Formula)} = \frac{\frac{\sum_w P(Y=1|A=1,W=w)\,\mathrm{P(W=w)}}{(1-\sum_w P(Y=1|A=1,W=w)\,\mathrm{P(W=w)})}}{\frac{\sum_w P(Y=1|A=0,W=w)\,\mathrm{P(W=w)}}{(1-\sum_w P(Y=1|A=0,W=w)\,\mathrm{P(W=w)})}}$$

Treatment heterogeneity: simulations



Simulation treatment heterogeneity and estimation using OR and MOR, n = 1,000.

Luque-Fernandez et al.(2019) Effect modification and collapsibility when estimating the effect of public health interventions: A Monte Carlo simulation comparing classical multivariable regression adjustment versus the G-Formula, based on a cancer epidemiology illustration. AJPH https://maluque.netlify.com/project/hetmor/

Ways to implement the G-Formula

ESTIMATION

- Non-parametric G-Formula.

Ways to implement the G-Formula

ESTIMATION

- Non-parametric G-Formula.
- **Modelling the Outcome (Y)**:

Ways to implement the G-Formula

ESTIMATION

- Non-parametric G-Formula.
- **Modelling the Outcome (Y)**:
  - Parametric (maximum likelihood substitution) G-computation.

Ways to implement the G-Formula

ESTIMATION

- Non-parametric G-Formula.
- **Modelling the Outcome (Y)**:
  - Parametric (maximum likelihood substitution) G-computation.
- **Modelling the treatment (A)**:

Ways to implement the G-Formula

ESTIMATION

- Non-parametric G-Formula.
- **Modelling the Outcome (Y)**:
  - Parametric (maximum likelihood substitution) G-computation.
- **Modelling the treatment (A)**:
  - Propensity score (adjusting for) or matching.

Ways to implement the G-Formula

ESTIMATION

- Non-parametric G-Formula.
- **Modelling the Outcome (Y)**:
  - Parametric (maximum likelihood substitution) G-computation.
- **Modelling the treatment (A)**:
  - Propensity score (adjusting for) or matching.
  - Inverse probability of treatment weights (IPTWs).

Ways to implement the G-Formula

ESTIMATION

- Non-parametric G-Formula.
- **Modelling the Outcome (Y)**:
  - Parametric (maximum likelihood substitution) G-computation.
- **Modelling the treatment (A)**:
  - Propensity score (adjusting for) or matching.
  - Inverse probability of treatment weights (IPTWs).
- **Modelling Both (A) and (Y)**: Parametric and semi-parametric Double Robust Estimators.

Ways to implement the G-Formula

ESTIMATION

- Non-parametric G-Formula.
- **Modelling the Outcome (Y)**:
  - Parametric (maximum likelihood substitution) G-computation.
- **Modelling the treatment (A)**:
  - Propensity score (adjusting for) or matching.
  - Inverse probability of treatment weights (IPTWs).
- **Modelling Both (A) and (Y)**: Parametric and semi-parametric Double Robust Estimators.
  - RA-IPTW

Ways to implement the G-Formula

ESTIMATION

- Non-parametric G-Formula.
- **Modelling the Outcome (Y)**:
  - Parametric (maximum likelihood substitution) G-computation.
- **Modelling the treatment (A)**:
  - Propensity score (adjusting for) or matching.
  - Inverse probability of treatment weights (IPTWs).
- **Modelling Both (A) and (Y)**: Parametric and semi-parametric Double Robust Estimators.
  - RA-IPTW
  - AIPTW

Ways to implement the G-Formula

ESTIMATION

- Non-parametric G-Formula.
- **Modelling the Outcome (Y)**:
  - Parametric (maximum likelihood substitution) G-computation.
- **Modelling the treatment (A)**:
  - Propensity score (adjusting for) or matching.
  - Inverse probability of treatment weights (IPTWs).
- **Modelling Both (A) and (Y)**: Parametric and semi-parametric Double Robust Estimators.
  - RA-IPTW
  - AIPTW
- Ensemble Learning: Targeted Maximum Likelihood Estimation.

Ways to implement the G-Formula

ESTIMATION

- Non-parametric G-Formula.
- **Modelling the Outcome (Y)**:
  - Parametric (maximum likelihood substitution) G-computation.
- **Modelling the treatment (A)**:
  - Propensity score (adjusting for) or matching.
  - Inverse probability of treatment weights (IPTWs).
- **Modelling Both (A) and (Y)**: Parametric and semi-parametric Double Robust Estimators.
  - RA-IPTW
  - AIPTW
- Ensemble Learning: Targeted Maximum Likelihood Estimation.

G-Computation: Regression Adjustment

RA: Regression-adjustment

$$\widehat{ATE}_{RA} = N^{-1} \sum_{i=1}^{N} [E(Y_i \mid A = 1, W_i) - E(Y_i \mid A = 0, W_i)]$$

$$m_A(w_i) = E(Y_i \mid A_i = A, W_i)$$

$$\widehat{ATE}_{RA} = N^{-1} \sum_{i=1}^{N} [\hat{m}_1(w_i) - \hat{m}_0(w_i)]$$

IPTW (Inverse probability treatment weighting)

Survey theory (Horvitz-Thompson)

$$\hat{P}_i = E(A_i \mid W_i) \text{ ; So } , \frac{\mathbf{1}}{\hat{\mathbf{p}}_\mathbf{i}} \text{ , if A = 1 and } , \frac{\mathbf{1}}{(\mathbf{1} - \hat{\mathbf{p}}_\mathbf{i})} \text{ , if A = 0}$$

Average over the total number of individuals

$$\widehat{ATE}_{IPTW} = N^{-1} \sum_{i=1}^{N} \frac{A_i Y_i}{\hat{p}_i} - N^{-1} \sum_{i=1}^{N} \frac{(1 - A_i) Y_i}{(1 - \hat{p}_i)}$$

AIPTW

AIPTW (Augmented Inverse probability treatment weighting)

Solving Estimating Equations

$$\widehat{ATE}_{AIPTW} =$$

$$N^{-1} \sum_{i=1}^{N} [(Y(1) \mid A_i = 1, W_i) - (Y(0) \mid A_i = 0, W_i)] +$$

$$N^{-1} \sum_{i=1}^{N} \left(\frac{(A_i = 1)}{P(A_i = 1 \mid W_i)} - \frac{(A_i = 0)}{P(A_i = 0 \mid W_i)} \right) [Y_i - E(Y \mid A_i, W_i)]$$

ATE estimators: drawbacks

Nonparametric

- Course of dimensionality (sparsity: zero empty cell)

Parametric

- Parametric models are misspecified (all models are wrong but some are useful, Box, 1976), and break down for high-dimensional data.
- (RA) Issue: extrapolation and biased if misspecification, no information about treatment mechanism.
- (IPTW) Issue: sensitive to course of dimensionality, inefficient in case of extreme weights and biased if misspecification. Non information about the outcome.

Double-robust (DR) estimators

Prons: Semi-parametric Double-Robust Methods

- DR methods give **two chances at consistency** if any of two nuisance parameters is consistently estimated.
- DR methods are **less sensitive** to **course of dimensionality**.

Cons: Semi-parametric Double-Robust Methods

- DR methods are unstable and inefficient if the propensity score (PS) is small (**violation of positivity assumption**) (vand der Laan, 2007).
- AIPTW and IPTW-RA do not respect the **limits of the boundary space of Y**.
- **Poor performance** if **dual misspecification** (Benkeser, 2016).

Targeted Maximum Likelihood Estimation (TMLE)

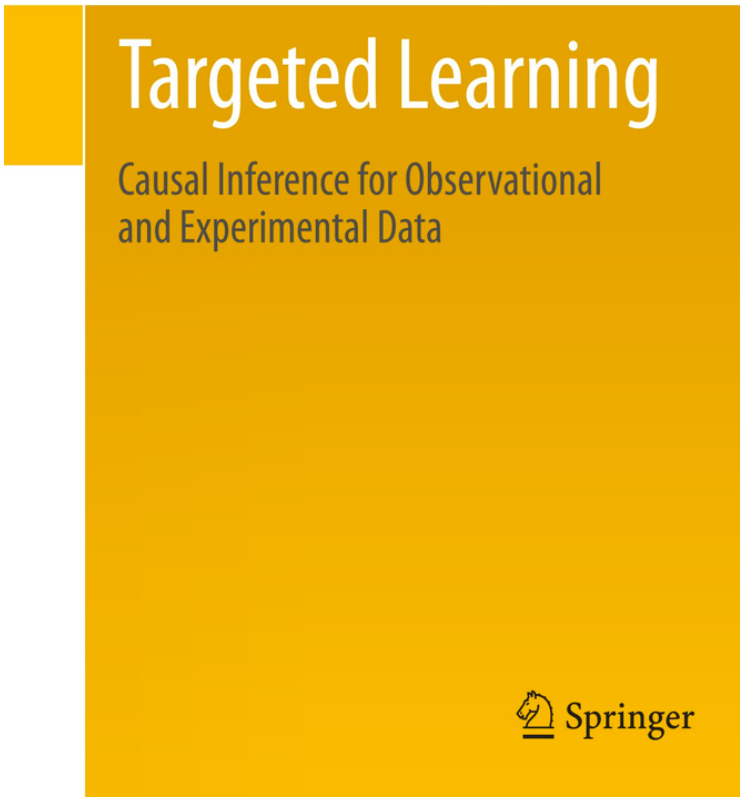
Pros: TMLE

- (TMLE) is a general algorithm for the construction of **double-robust**, **semiparametric** MLE, efficient **substitution** estimator (Van der Laan, 2011)
- Better performance than competitors has been largely documented (Porter, et. al.,2011).
- (TMLE) **Respect bounds on Y**, **less sensitive** to **misspecification** and to **near-positivity** violations (Benkeser, 2016).
- (TMLE) **Reduces bias** through **ensemble learning** if misspecification, even dual misspecification.
- For the ATE, **Inference** is based on the **Efficient Influence Curve**. Hence, the **CLT** applies, making inference easier.

Cons: TMLE

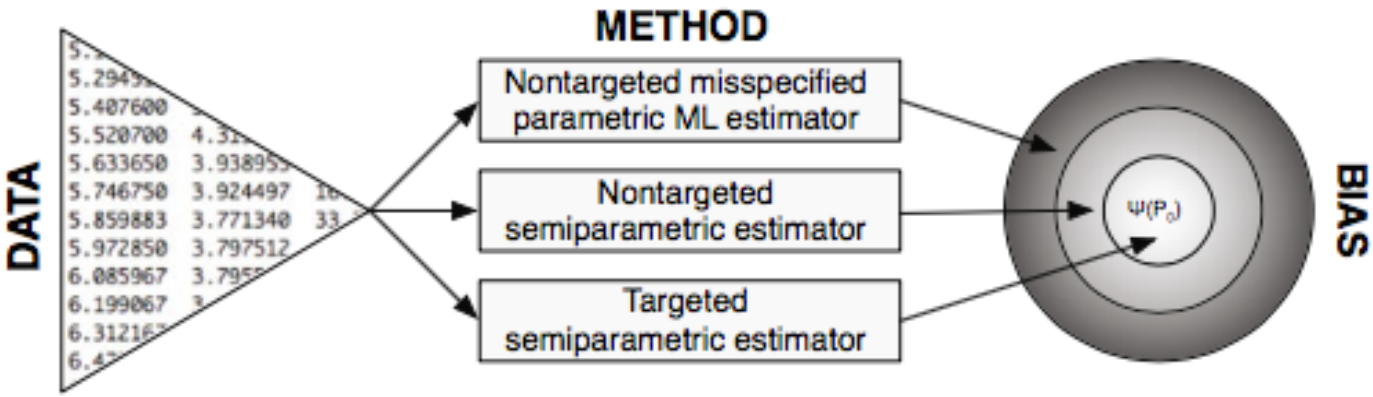
- The procedure is only available in R: **tmle** package (Gruber, 2011).

Targeted learning



[1] **Source**: Mark van der Laan and Sherri Rose. Targeted learning: causal inference for observational and experimental data. Springer Series in Statistics, 2011.

Why Targeted learning?



[2] **Source**: Mark van der Laan and Sherri Rose. Targeted learning: causal inference for observational and experimental data. Springer Series in Statistics, 2011.

TMLE ROAD MAP

MC simulations: Luque-Fernandez et al., 2017 (American Journal of Epidemiology)

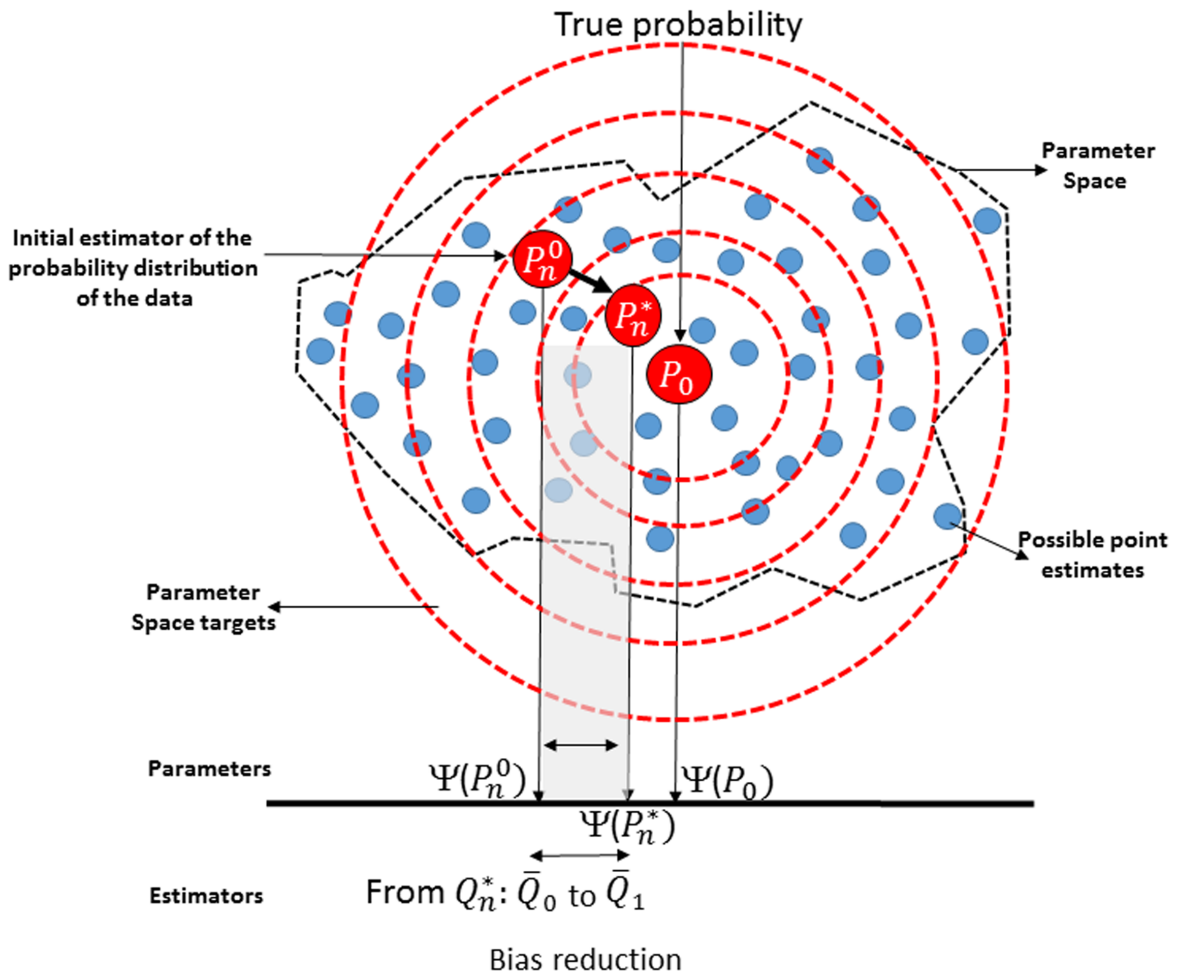
	ATE		BIAS (%)		RMSE		95%CI coverage (%)	
	N=1,000	N=10,000	N=1,000	N=10,000	N=1,000	N=10,000	N=1,000	N=10,000
First scenario* (correctly specified models)								
True ATE	**-0.1813**							
Naïve	-0.2234	-0.2218	23.2	22.3	0.0575	0.0423	77	89
AIPTW	-0.1843	-0.1848	1.6	1.9	0.0534	0.0180	93	94
IPTW-RA	-0.1831	-0.1838	1.0	1.4	0.0500	0.0174	91	95
TMLE	-0.1832	-0.1821	1.0	**0.4**	0.0482	0.0158	**95**	**95**
Second scenario ** (misspecified models)								
True ATE	**-0.1172**							
Naïve	-0.0127	-0.0121	89.2	89.7	0.1470	0.1100	0	0
BFit AIPTW	-0.1155	-0.0920	1.5	**11.7**	0.0928	0.0773	65	65
BFit IPTW-RA	-0.1268	-0.1192	8.2	**1.7**	0.0442	0.0305	52	73
TMLE	-0.1181	-0.1177	**0.8**	**0.4**	0.0281	0.0107	**93**	**95**

*First scenario : correctly specified models and near-positivity violation

**Second scenario: misspecification, near-positivity violation and adaptive model selection

[3] Data-Adaptive Estimation for Double-Robust Methods in Population-Based Cancer Epidemiology: Risk differences for lung cancer mortality by emergency presentation (2017). AJE. https://academic.oup.com/aje/article/doi/10.1093/aje/kwx317/4110407

TMLE ROAD MAP



[4] Luque-Fernandez, MA. 2017. TMLE steps adapted from van der Laan MJ, 2011.

Ways to implement the G-Formula: RA

G-computation RA: LAB 1

- It is a maximum likelihood substitution estimator of the G-Formula.

Ways to implement the G-Formula: RA

G-computation RA: LAB 1

- It is a maximum likelihood substitution estimator of the G-Formula.
- It is equivalent to using the marginal distribution of the covariates as the standard in standardization.

Ways to implement the G-Formula: RA

G-computation RA: LAB 1

- It is a maximum likelihood substitution estimator of the G-Formula.
- It is equivalent to using the marginal distribution of the covariates as the standard in standardization.

Ways to implement the G-Formula: IPTW

IPTW: LAB 2

- It is a maximum likelihood substitution estimator of the G-Formula using logistic regression.

Ways to implement the G-Formula: IPTW

IPTW: LAB 2

- It is a maximum likelihood substitution estimator of the G-Formula using logistic regression.
- It is equivalent to using the inverse weighting by the probability of treatment (i.e., propensity score) to estimate the parameters defined by a marginal structural model.

Ways to implement the G-Formula: IPTW

IPTW: LAB 2

- It is a maximum likelihood substitution estimator of the G-Formula using logistic regression.
- It is equivalent to using the inverse weighting by the probability of treatment (i.e., propensity score) to estimate the parameters defined by a marginal structural model.

ESTIMATION

G-Computation and IPTW in STATA

Classical regression adjustment

```
use http://www.stata-press.com/data/r14/cattaneo2.dta
reg bweight i.mbsmoke mage
```

```
      Source |       SS           df       MS      Number of obs   =        60
-------------+----------------------------------   F(2, 57)        =     83.71
       Model |   1341718.7         2  670859.348   Prob > F        =    0.0000
    Residual |  456829.236        57  8014.54801   R-squared       =    0.7460
-------------+----------------------------------   Adj R-squared   =    0.7371
       Total |  1798547.93        59  30483.8633   Root MSE        =    89.524
------------------------------------------------------------------------------
     bweight |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     mbsmoke |
     smoker  |  -368.1796    29.8213   -12.35   0.000    -427.8957   -308.4635
       mage  |   15.9611   3.324991     4.80   0.000      9.30292    22.61928
       _cons |   3191.268   73.22399    43.58   0.000      3044.64    3337.897
------------------------------------------------------------------------------
```

Stata v.14 new module for Tx effects

```
teffects ra (bweight mage) (mbsmoke)

Treatment-effects estimation                    Number of obs     =         60
Estimator      : regression adjustment
Outcome model  : linear
Treatment model: none
------------------------------------------------------------------------------
                             |               Robust
                    bweight  |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----------------------------+------------------------------------------------
ATE                          |
                    mbsmoke  |
(smoker vs nonsmoker)        |  -375.5157   29.62863   -12.67   0.000    -433.5867   -317.4446
-----------------------------+------------------------------------------------
POmean                       |
                    mbsmoke  |
                 nonsmoker   |    3598.81   24.22443   148.56   0.000     3551.331    3646.289
------------------------------------------------------------------------------}
```

Computing the RA by ”hand”

```
reg bweight mage if mbsmoke==1
predict double y1hat

reg bweight mage if mbsmoke==0
predict double y0hat

mean y1hat y0hat

lincom _b[y1hat] - _b[y0hat]
------------------------------------------------------------------------------
        Mean |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |  -375.5157   6.601436   -56.88   0.000    -388.7251   -362.3062
------------------------------------------------------------------------------
```

Non-parametric estimation: conditional means saturated model

```
reg bweight ibn.mbsmoke ibn.mbsmoke#c.mage, noconstant vce(robust) vsquish

//Potential outcomes

margins mbsmoke, vce(unconditional)

Predictive margins                                 Number of obs     =         60
Expression   : Linear prediction, predict()
------------------------------------------------------------------------------
             |            Unconditional
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     mbsmoke |
  nonsmoker  |    3598.81   25.07467   143.52   0.000     3548.579     3649.04
     smoker  |   3223.294   21.43679   150.36   0.000     3180.351    3266.237
------------------------------------------------------------------------------
```

Non-parametric estimation: conditional means saturated model

```
reg bweight ibn.mbsmoke ibn.mbsmoke#c.mage, noconstant vce(robust) vsquish
//ATE
margins r.mbsmoke, contrast(nowald)

Contrasts of predictive margins
Model VCE    : Robust
Expression   : Linear prediction, predict()
------------------------------------------------------------------------------
                        |            Delta-method
                        |   Contrast   Std. Err.     [95% Conf. Interval]
------------------------+-----------------------------------------------------
                mbsmoke |
(smoker vs nonsmoker)   |  -375.5157   30.87175     -437.3592   -313.6721
------------------------------------------------------------------------------
```

Estimation of the CRR

```
teffects ra (bweight mage) (mbsmoke), coeflegend

nlcom 100*_b[ATE:r1vs0.mbsmoke]/_b[POmean:r0.mbsmoke]

------------------------------------------------------------------------------
     bweight |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _nl_1 |  -10.43444   .7741387   -13.48   0.000    -11.95172   -8.917157
------------------------------------------------------------------------------
```

Estimation of the IPTW

```
. teffects ipw (bweight) (mbsmoke mage, logit), nolog vsquish

Treatment-effects estimation                    Number of obs     =      4,642
Estimator      : inverse-probability weights
Outcome model  : weighted mean
Treatment model: logit
------------------------------------------------------------------------------
                             |               Robust
                     bweight |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----------------------------+------------------------------------------------
ATE                          |
                     mbsmoke |
(smoker vs nonsmoker)        |  -275.5595   22.63192   -12.18   0.000    -319.9173   -231.2018
-----------------------------+------------------------------------------------
POmean                       |
                     mbsmoke |
                  nonsmoker  |   3408.938   9.308235   366.23   0.000     3390.695    3427.182
------------------------------------------------------------------------------
```

Estimation of the IPTW by hand

```
. clear
. use http://www.stata-press.com/data/r14/cattaneo2.dta
. global Y bweight // Outcome
. global A mbsmoke // Exposure or treatment
. global W mmarried prenatal1 fbaby medu mage foreign mrace // Confounders.
. teffects ipw ($Y) ($A $W, logit), nolog vsquish

Treatment-effects estimation                    Number of obs     =      4,642
Estimator      : inverse-probability weights
Outcome model  : weighted mean
Treatment model: logit
------------------------------------------------------------------------------
                             |               Robust
                     bweight |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----------------------------+------------------------------------------------
ATE                          |
                     mbsmoke |
(smoker vs nonsmoker)        |  -240.2212   24.82773    -9.68   0.000    -288.8826   -191.5597
-----------------------------+------------------------------------------------
POmean                       |
                     mbsmoke |
                  nonsmoker  |   3408.235   9.440857   361.01   0.000     3389.732    3426.739
------------------------------------------------------------------------------
```

Estimation of the IPTW by hand

```
. logit $A $W, vce(robust) nolog
. predict double ps
. generate double ipw1 = ($A==1)/ps
. regress $Y [pw=ipw1]

Linear regression                               Number of obs     =        864
------------------------------------------------------------------------------
             |               Robust
     bweight |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   3168.014   23.35919   135.62   0.000     3122.167    3213.862
------------------------------------------------------------------------------

. generate double ipw0 = ($A==0)/(1-ps)
. regress $Y [pw=ipw0]

Linear regression                               Number of obs     =      3,778
------------------------------------------------------------------------------
             |               Robust
     bweight |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   3408.235   9.523029   357.89   0.000     3389.565    3426.906
------------------------------------------------------------------------------
```

Estimation of the IPTW by hand: bootstrapping for SE

```
. program drop ATE
. program define ATE, rclass
  . capture drop y1
  . capture drop y0
  . regress $Y [pw=ipw1]
  . matrix y1 = e(b)
  . gen double y1 = y1[1,1]
  . regress $Y [pw=ipw0]
  . matrix y0 = e(b)
  . gen double y0 = y0[1,1]
  . mean y1 y0
  . lincom _b[y1]-_b[y0]
  . return scalar ace = `r(estimate)'
. end
```

Estimation of the IPTW by hand

```
. qui bootstrap r(ace), reps(1000): ATE
. estat boot, all

Bootstrap results                                 Number of obs     =      4,642
                                                  Replications      =       1000

      command:  ATE
        _bs_1:  r(ace)

------------------------------------------------------------------------------
             |    Observed                Bootstrap
             |       Coef.       Bias    Std. Err.  [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _bs_1 |  -240.22116   .2588013   24.989583   -289.1998  -191.2425   (N)
             |                                      -287.8047  -192.4459   (P)
             |                                      -287.6467  -192.4318  (BC)
------------------------------------------------------------------------------
(N)    normal confidence interval
(P)    percentile confidence interval
(BC)   bias-corrected confidence interval
```

Estimation of the IPTW by hand

```
. // Test balance after adjustment
. qui teffects ipw ($Y) ($A $W)
. tebalance summarize

  Covariate balance summary
                                                  Raw     Weighted
                          ----------------------------------------
                          Number of obs =         4,642      4,642.0
                          Treated obs   =           864      2,282.2
                          Control obs   =         3,778      2,359.8
                          ----------------------------------------

  ------------------------------------------------------------------
                  |Standardized differences          Variance ratio
                  |        Raw    Weighted             Raw   Weighted
  ----------------+-------------------------------------------------
         mmarried |  -.5953009    -.020219      1.335944   1.017057
       prenatal1  |  -.3242695   -.0186635      1.496155     1.0278
           fbaby  |  -.1663271    .0217106      .9430944   1.005001
            medu  |  -.5474357   -.1156001      .7315846   .5112843
            mage  |   -.300179   -.0740731      .8818025   .7979432
         foreign  |  -.1706164    -.034976      .4416089   .8643349
           mrace  |  -.1029446    -.028183      1.198452   1.052116
  ------------------------------------------------------------------
```

Survival-time treatment effects example

The time to a second heart attack among women aged 45 to 55 years. The treatment, smoking, is stored in the 0, 1 indicator smoke. These data also contain each woman's age at the time of her first heart attack (age), and indices of her exercise level (exercise), diet quality (diet), and education attainment (education) prior to her first heart attack.

Survival-time treatment effects

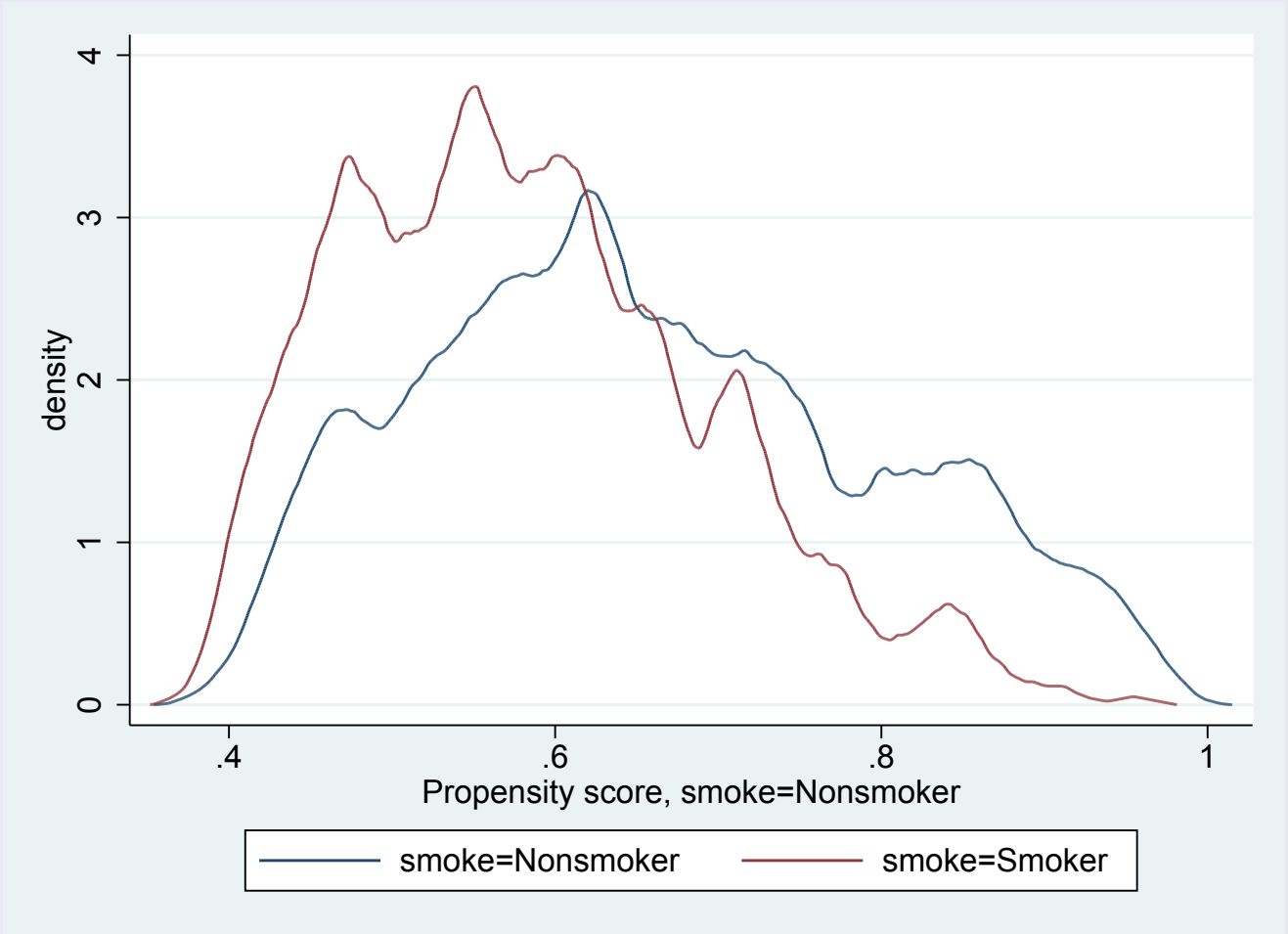
Does smoking decrease the time to a second heart attack in the population of women aged 45 to 55 who have had one heart attack?

```
use http://www.stata-press.com/data/r14/sheart
stset atime, failure(fail)
     failure event:  fail != 0 & fail < .
obs. time interval:  (0, atime]
 exit on or before:  failure
------------------------------------------------------------------------------
       2000  total observations
          0  exclusions
------------------------------------------------------------------------------
       2000  observations remaining, representing
       1208  failures in single-record/single-failure data
   3795.226  total analysis time at risk and under observation
                                                at risk from t =         0
                                     earliest observed entry t =         0
                                          last observed exit t =  34.17743
```

Survival-time treatment effects

```
*Best model for the treatment model
bfit logit smoke age exercise diet education
display "`r(bvlist)'"
*I am using a double robust method to fit a IPW RA treatment effect model
with weighted adjustment for censoring
stteffects ipwra (age exercise diet education) (smoke age exercise diet education)  \\\
(age exercise diet)
         failure _d:  fail
   analysis time _t:  atime
Iteration 0:   EE criterion =  1.632e-16
Iteration 1:   EE criterion =  9.694e-31
Survival treatment-effects estimation           Number of obs     =      2,000
Estimator      : IPW regression adjustment
Outcome model  : Weibull
Treatment model: logit
Censoring model: Weibull
------------------------------------------------------------------------------
                     |               Robust
                  _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------------+--------------------------------------------------------------
ATE                  |
              smoke  |
(Smoker vs Nonsmoker)|  -2.037944   .6032549    -3.38   0.001    -3.220302    -.855586
---------------------+--------------------------------------------------------------
POmean               |
              smoke  |
          Nonsmoker  |    4.14284   .4811052     8.61   0.000     3.199891    5.085789
------------------------------------------------------------------------------
```

Overlap: positivity assumption and SUTVA



Survival-time treatment effects

```
*IMAI AND RATKOVIC (2014): Over-identification or Test for Balance
(Testing and checking conditional independence or unconfoundeness)
tebalance summ

  Covariate balance summary
                                                 Raw     Weighted
                          ----------------------------------------
                          Number of obs =        2,000      2,000.0
                          Treated obs   =          738        994.1
                          Control obs   =        1,262      1,005.9
                          ----------------------------------------
  ------------------------------------------------------------------
                  |Standardized differences          Variance ratio
                  |        Raw    Weighted            Raw   Weighted
  ----------------+-------------------------------------------------
              age |  -.3122094   -.0184574      .8547308   .9370065
         exercise |  -.4975269   -.0458412      .4966778   .8342339
             diet |  -.2479756    .0021802      .7937645   1.095347
        education |  -.4801442   -.0216366      .6015139    .978078
  ------------------------------------------------------------------
. tebalance over, nolog
Overidentification test for covariate balance
         H0: Covariates are balanced:
         chi2(5)      =  3.28142
         Prob > chi2  =   0.6567
```

Tutorials Causal Inference

TMLE R-markdown 2019

https://migariane.github.io/TMLE.nb.html

TMLE Statistics in Medicine 2018

https://www.ncbi.nlm.nih.gov/pubmed/29687470

SIM-2018. CODE in R and STATA

https://github.com/migariane/SIM-TMLE-tutorial

LABs 1-2: G-Comp. and IPTW

Practicals

G-Comp. and IPTW using R (RStudio cloud)

LABs 1-2: G-Comp. and IPTW

Practicals

G-Comp. and IPTW using R (RStudio cloud)

RStudio Cloud link

https://rstudio.cloud/spaces/19488/project/434105

LABs 1-2: G-Comp. and IPTW

Practicals

G-Comp. and IPTW using R (RStudio cloud)

RStudio Cloud link

https://rstudio.cloud/spaces/19488/project/434105

Thank you!

Miguel Angel Luque-Fernandez

miguel.luque.easp@juntadeandalucia.es @watzilei

Carlos III Institute of Health, Grant/Award Number: **CP17/00206**