Estimating Inverse Probability Weights using **Super Learner** when weight-model specification is unknown in a Marginal Structural Cox Model context

Mohammad Ehsanul Karim

CHÉOS
Centre for Health Evaluation & Outcome Sciences

Providence Health Care & University of British Columbia

38th Annual Conference of the International Society for Clinical Biostatistics

July 10th, 2017

Acknowledgement

Joint work with

1. Robert W. Platt, McGill University
2. The BeAMS Study (Long-term Benefits and Adverse Effects of Beta-interferon for Multiple Sclerosis)
    - Shirani, A.; Zhao Y.; Evans C.; Kingwell E.; van der Kop M.L.; Oger J.; Gustafson, P.; Petkau, J.; Tremlett, H.

CNODES
CANADIAN NETWORK FOR OBSERVATIONAL DRUG EFFECT STUDIES

Statistics in Medicine

DOI: 10.1002/sim.7266

Outline

1. Multiple sclerosis
   - Data description
   - Causal diagram for treatment effect
2. Marginal Structural Cox Model
   - Weight formula
   - Calculation using **Super learner**
3. Results
   - Simulation
   - Data analysis
4. Discuss

American Journal of Epidemiology

Vol. 180, No. 2
DOI: 10.1093/aje/kwu125
Advance Access publication:
June 17, 2014

Practice of Epidemiology

Marginal Structural Cox Models for Estimating the Association Between β-Interferon Exposure and Disease Progression in a Multiple Sclerosis Cohort

Mohammad Ehsanul Karim*, Paul Gustafson, John Petkau, Yinshan Zhao, Afsaneh Shirani, Elaine Kingwell, Charity Evans, Mia van der Kop, Joel Oger, and Helen Tremlett

* Correspondence to Mohammad Ehsanul Karim, Department of Statistics, Faculty of Science, University of British Columbia, 3182 Earth Sciences Building, Vancouver, British Columbia V6T 1Z4, Canada (e-mail: ehsan@stat.ubc.ca).

- Retrospective study (1995-2008), BC.
- 1,697 patients followed; 829 remained untreated.

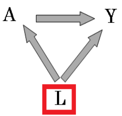
Variables under consideration:

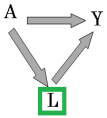
- Treatment: β-interferons (time-dependent exposure A_t)
- Survival outcome (Y): time from baseline to irreversible disability (sustained EDSS 6).
- Confounders (L_0) measured at baseline:
  1. Disability status (measured by EDSS score)
  2. Disease duration
  3. Age
  4. Sex
- Time-dependent confounder (L_t): Relapse

Cox PH model with baseline and time-dependent confounder:

$$\lambda(t|L_0, L_t) = \lambda_{0t} \exp\left(\psi_1 A_t + \psi_2 L_0 + \psi_3 L_t\right)$$

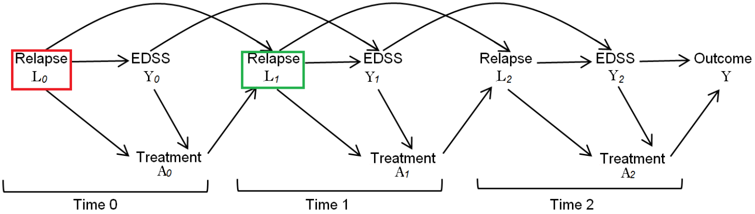
Multiple Sclerosis Data > Causal Diagram





Confounder
(L should be controlled)
$$\lambda_{0t} \exp\left(\psi_1 A_{it} + \psi_2 L_{i0} + \psi_3 L_{it}\right)$$

Intermediate variable
(L should not be controlled)
$$\lambda_{0t} \exp\left(\psi_1 A_{it} + \psi_2 L_{i0}\right)$$



Time-dependent confounder (relapse) is affected by prior treatment.

Calculate weights (stabilized treatment IPW):

Step 1: $$sw_t = \prod_{j=0}^{t} \frac{pr(A_j = a_j | \bar{A}_{j-1} = \bar{a}_{j-1}, L_0 = l_0)}{pr(A_j = a_j | \bar{A}_{j-1} = \bar{a}_{j-1}, L_0 = l_0, \bar{L}_j = \bar{l}_j)}$$

Outcome model (in the pseudo-population) with baseline confounder:
Marginal Structural Cox Model

Step 2: $$\lambda(t|L_0) = \lambda_0(t) \exp\left(\psi_1 A_t + \psi_2 L_0\right)$$

Numerator weight model:

Step 1a: $$logit\ Pr(A_j = 1 | \bar{A}_{j-1}, L_0, \boldsymbol{\alpha'}) = \alpha'_0(j) + \alpha'_1 A_{j-1} + \alpha'_2 L_0 \quad (1)$$

Denominator weight model:

Step 1b: $$logit\ Pr(A_j = 1 | \bar{A}_{j-1}, L_0, \bar{L}_j, \boldsymbol{\alpha}) = \alpha_0(j) + \alpha_1 A_{j-1} + \alpha_2 L_0 + \alpha_3 L_j \quad (2)$$

Weights play a key role in the Marginal Structural Cox Model approach:

1. In practical applications, researchers are often unaware of the **true form of the weight model**:
   - non-linearity (e.g., quadratic or higher-order effects)
   - non-additivity (e.g., interaction terms)
2. Marginal Structural Cox Model estimates are highly **sensitive to the weight-model mis-specification**.

IPW Model	Parametric Regression	Data-adaptive Methods
Example	Logistic regression	Classification and regression trees
Pros	**Efficient** MSCM estimates.	Data-adaptively **detects** data features.
Cons	Assumptions may be too **restrictive**.	Possibly **inefficient** MSCM estimates.

1. **Super Learner** utilizes a user-specified candidate library; may include
   - parametric regression models
   - semi-parametric regression models
   - data-adaptive statistical learning methods

Statistical Applications in Genetics and Molecular Biology

Volume 6, *Issue* 1 | 2007 | *Article* 25

Super Learner

Mark J. van der Laan* Eric C. Polley† Alan E. Hubbard‡

2. Using cross-validation, **Super learner** approach **optimally combines the predicted values** from each candidate learner.

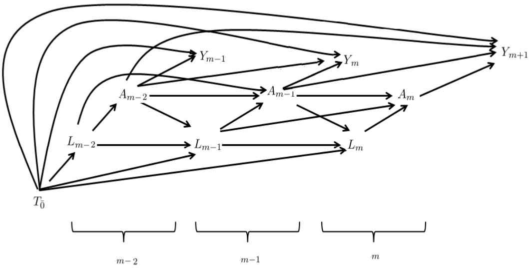
3. Prediction-wise, **Super learner** generally **asymptotically outperforms** each of the candidate estimators (if correct parametric specification absent).

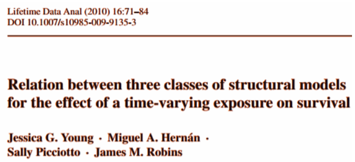
4. **Super learner** may offer a **better alternative** (when true parametric weight-model specification unknown).

Candidate learners in the **Super learner** library

Learner ($4+6=10$)	Description
Logistic regression	The main terms of the covariates
Stepwise logistic	Variables selected from quadratic terms and two-way interactions based on AIC criterion
Elastic net	Mixing parameter $= 0.5$
Bayesian logistic	Cauchy prior with scale $= 2.5$
CART	Complexity parameter $= 0.01$
Pruned CART	Complexity parameter chosen such that the cross-validated error rate is minimum
Bagged CART	Based on 100 replications
Boosted CART	Based on $5,000$ trees and interaction depth $= 3$
Random Forest	Based on $1,000$ trees
SVM	Polynomial kernel

Simulation specifications >




Lifetime Data Anal (2010) 16:71–84
DOI 10.1007/s10985-009-9135-3

Relation between three classes of structural models for the effect of a time-varying exposure on survival

Jessica G. Young · Miguel A. Hernán · Sally Picciotto · James M. Robins

Figure: Causal diagram of Marginal Structural Cox Model data generation algorithm. L_m is a continuous variable & a **time-dependent confounder**, where $L_m = \beta_0 + \beta_1\big(1/\log(T_0)\big) + \beta_2 A_{m-1} + \beta_3 L_{m-1}$.

Monte Carlo study: $N = 1000$ datasets, $n = 250$ subjects, each followed for up to $m = 10$ subsequent monthly visits, $\lambda_0 = 0.01$ monthly event rate

Current treatment, one time-dependent confounder:

I. Additivity and linearity

$$logit(p_A) \quad = \quad \alpha_0 + \alpha_1 A_{m-1} + \alpha_2 L_m + \alpha_3 L_{m-1}.$$

II. Non-additivity (interaction)

$$logit(p_A) \quad = \quad \alpha_0 + \alpha_1 A_{m-1} + \alpha_2 L_m + \alpha_3 L_{m-1} + \alpha_4 (A_{m-1} \times L_m).$$

III. Non-linearity (polynomial)

$$logit(p_A) \quad = \quad \alpha_0 + \alpha_1 A_{m-1} + \alpha_2 (L_m)^2 + \alpha_3 (L_{m-1})^2.$$

IV. Non-linearity and non-additivity (both)

$$logit(p_A) \quad = \quad \alpha_0 + \alpha_1 A_{m-1} + \alpha_2 (L_m)^2 + \alpha_3 (L_{m-1})^2 + \alpha_4 (A_{m-1} \times L_m).$$

Current treatment, one time-dependent confounder, sw:

I. Additivity & linearity	MSCM Bias	MSCM MSE	SE	SD	Cov.Pr.
Super learner	-0.0719	0.0844	0.312	0.281	0.969
Elastic net	-0.1336	0.1031	0.308	0.292	0.934
Boosted CART	-0.1493	0.1039	0.314	0.286	0.951
Bayesian logistic	0.0195	0.1071	0.323	0.327	0.972
Logistic	0.0645	0.1218	0.329	0.343	0.972
Bagged CART	-0.2469	0.2749	0.386	0.463	0.837
Stepwise	0.1458	0.3750	0.346	0.595	0.950
CART	-0.4232	0.4221	0.397	0.493	0.722
Pruned CART	-0.6215	0.6246	0.342	0.488	0.507
SVM	0.3807	1.7024	0.502	1.248	0.601
Random Forest	-0.6002	2.4178	0.309	1.434	0.148

Ordered by MSE (ascending).

Current treatment, one time-dependent confounder, sw:

I. Additivity & linearity	II. Non-additivity	III. Non-linearity	IV. Both
Super learner	**Super learner**	**Super learner**	**Super learner**
Elastic net	Boosted CART	CART	CART
Boosted CART	Bagged CART	Bagged CART	Bagged CART
Bayesian logistic	Stepwise	Boosted CART	Boosted CART
Logistic	Random Forest	Pruned CART	Pruned CART
Bagged CART	CART	Stepwise	Bayesian logistic
Stepwise	Pruned CART	Elastic net	Elastic net
CART	Elastic net	Bayesian logistic	Stepwise
Pruned CART	Bayesian logistic	Logistic	Logistic
SVM	Logistic	Random Forest	Random Forest
Random Forest	SVM	SVM	SVM

Ordered by MSE (ascending).

Current treatment, one time-dependent confounder, sw, large n:

I. Additivity & linearity	II. Non-additivity	III. Non-linearity	IV. Both
Boosted CART	Boosted CART	**Super learner**	Boosted CART
Super learner	**Super learner**	Elastic net	**Super learner**
Elastic net	Pruned CART	Bayesian logistic	Elastic net
Stepwise	Stepwise	Logistic	Bayesian logistic
Bayesian logistic	CART	Boosted CART	Logistic
Logistic	Bagged CART	Bagged CART	Bagged CART
Pruned CART	Elastic net	Pruned CART	CART
CART	Bayesian logistic	CART	Pruned CART
Bagged CART	Logistic	Stepwise	SVM
Random Forest	Random Forest	Random Forest	Stepwise
SVM	SVM	SVM	Random Forest

Ordered by MSE (ascending).

Cumulative treatment, one time-dependent confounder, sw:

I. Additivity & linearity	II. Non-additivity	III. Non-linearity	IV. Both
Boosted CART	**Super learner**	**Super learner**	**Super learner**
Super learner	Boosted CART	Boosted CART	Boosted CART
Elastic net	Elastic net	Elastic net	Elastic net
Bagged CART	Bayesian logistic	Bayesian logistic	Bayesian logistic
Bayesian logistic	Logistic	Logistic	Logistic
Logistic	SVM	Stepwise	Stepwise
CART	Bagged CART	Bagged CART	Bagged CART
Pruned CART	CART	Pruned CART	Pruned CART
Stepwise	Pruned CART	CART	CART
SVM	Stepwise	SVM	SVM
Random Forest	Random Forest	Random Forest	Random Forest

Ordered by MSE (ascending).

Current treatment, two time-dependent confounders, sw:

I. Additivity & linearity	II. Non-additivity	III. Non-linearity	IV. Both
Boosted CART	Boosted CART	**Super learner**	**Super learner**
Bagged CART	Bagged CART	Boosted CART	Boosted CART
Super learner	**Super learner**	Bagged CART	Bagged CART
Elastic net	Elastic net	CART	CART
CART	Bayesian logistic	Elastic net	Elastic net
Bayesian logistic	Logistic	Bayesian logistic	Bayesian logistic
Logistic	CART	Logistic	Logistic
Pruned CART	Pruned CART	Pruned CART	Pruned CART
Stepwise	SVM	Logistic	Stepwise
Random Forest	Random Forest	SVM	SVM
SVM	Stepwise	Random Forest	Random Forest

Ordered by MSE (ascending).

Table: The estimated treatment effect of β-IFN on reaching sustained EDSS 6 for BC MS patients (1995-2008) from the Marginal Structural Cox Model.

Estimated weights sw generated via **Super learner**		Treatment effect estimate		
Mean (log-SD)	Min-Max	HR	SE	95% CI
1.056 (-0.771)	0.392 - 2.379	1.349	0.316	0.853 - 2.134

- In our **Multiple Sclerosis application**, the hazard ratio estimates from the super learning approach is 1.349, and this effect estimate was not significant (95% CI $0.853 - 2.134$).
- This conclusion is consistent with those of the previous studies.

Summary >

- This study shows the **utility of using Super learner** approach with rich set of candidate learners in practical scenarios when
  1. the form of the **treatment decision model is unknown** and
  2. may deviate from linearity, additivity or both.

- Marginal Structural Cox Model estimates computed from **Super learner** (sw) **generally performed better** in terms of
  1. **MSE** and
  2. 95% **coverage probability**

  compared to individual candidate learners.

- However, these tools are not meant to replace **subject-matter knowledge and expert-opinion**:
  1. bias amplification (controlling IV),
  2. under-adjustment (omitting potential confounder).

Thank You!

ekarim @ cheos.ubc.ca

Current treatment, one time-dependent confounder, $sw^{(n)}$:

I. Additivity & linearity	II. Non-additivity	III. Non-linearity	IV. Both
Super learner	**Super learner**	**Super learner**	**Super learner**
Elastic net	Boosted CART	CART	CART
Boosted CART	Bagged CART	Bagged CART	Bagged CART
Bayesian logistic	Stepwise	Boosted CART	Boosted CART
Logistic	Random Forest	Pruned CART	Pruned CART
Bagged CART	CART	Stepwise	Bayesian logistic
Stepwise	Pruned CART	Elastic net	Elastic net
CART	Elastic net	Bayesian logistic	Stepwise
Pruned CART	Bayesian logistic	Logistic	Logistic
SVM	Logistic	Random Forest	Random Forest
Random Forest	SVM	SVM	SVM

Ordered by MSE (ascending).

Current treatment, one time-dependent confounder, $sw - 1\%$ truncation:

I. Additivity & linearity	II. Non-additivity	III. Non-linearity	IV. Both
Logistic	**Super learner**	Bagged CART	Bagged CART
Stepwise	Boosted CART	Logistic	Logistic
Bayesian logistic	Bagged CART	Elastic net	Elastic net
Super learner	Stepwise	Boosted CART	Bayesian logistic
Boosted CART	CART	**Super learner**	Boosted CART
Elastic net	SVM	Bayesian logistic	**Super learner**
Bagged CART	Elastic net	CART	CART
CART	Random Forest	Stepwise	Stepwise
SVM	Bayesian logistic	SVM	SVM
Pruned CART	Logistic	Pruned CART	Pruned CART
Random Forest	Pruned CART	Random Forest	Random Forest

Ordered by MSE (ascending).