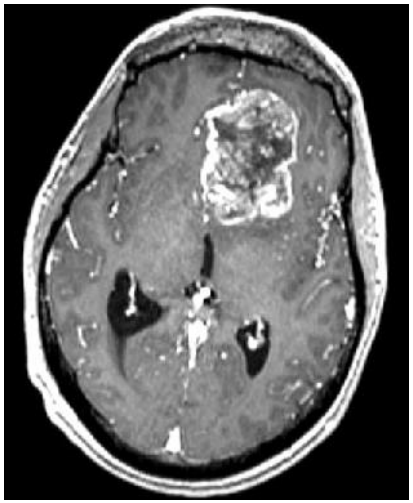
Robustifying Causal Inference Methods Via Model Averaging

Francesca Dominici, PhD
Professor of Biostatistics
Co-Director of the Harvard Data Science Initiative
International Society of Clinical Biostatistics
Vigo, Spain
July 12 2017

HARVARD T.H. CHAN
SCHOOL OF PUBLIC HEALTH
Powerful ideas for a healthier world

Glioblastoma (GBM)

- GBM is the most common malignant primary brain tumor
- Approximately 15,000 GBM cases/year in U.S.
- Median age at GBM diagnosis: 65 years
- Median survival range for GBM: 6 to 18 months
- Despite half of all GBM patients being elderly, older patients under-represented in RCTs
- Uncertainty about how to treat and how to improve quality of life



TMZ/RT vs. RT in Elderly Patients with Glioblastoma

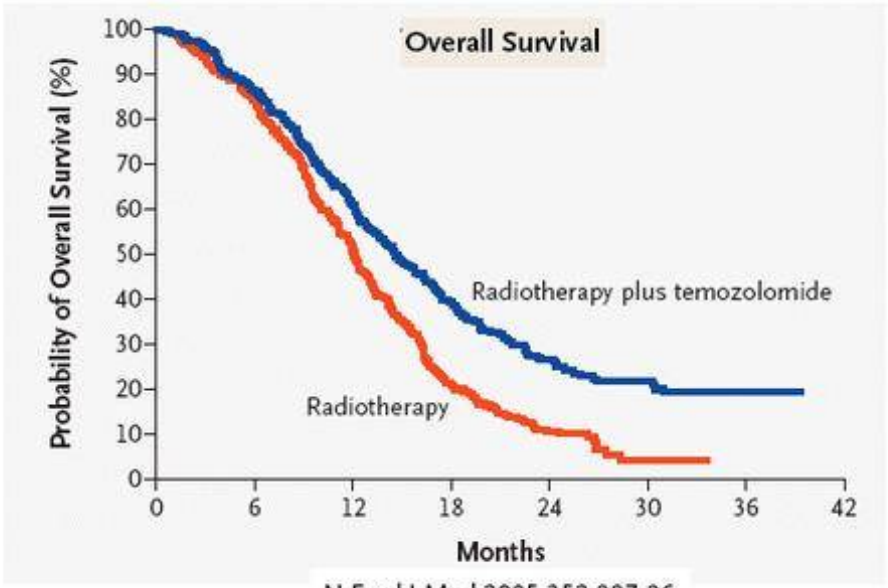
Purpose:

To examine overall survival among elderly GBM patients receiving TMZ/RT vs. RT alone

Rationale:

Concurrent TMZ/RT widely used/recommended for elderly GBM patients, but benefit of TMZ is unclear in this population

Radiotherapy plus Concomitant and Adjuvant Temozolomide for Glioblastoma



N Engl J Med 2005;352:987-96.

RCTs have strict eligibility criteria and therefore results are often not generalizable to the entire population

Observational data

1. Structured data elements from electronic medical records (EMR)
2. Cancer registries linked to claims data (SEER linked to Medicare)
3. Claims data (Medicare claims)

Challenges

Confounding adjustment

Confounder selection

Model uncertainty

Glioblastoma

Question

Is there a survival benefit of adding TMZ to radiotherapy for patients 65 and older with GBM?

X=1 radiotherapy + TMZ

X=0 radiotherapy alone

Data source

SEER linked Medicare data (2005-2009)

- Inpatient and outpatient data
- 75 covariates

Notation

- C = Large set of potential confounders
- X = Treatment
- Y = Outcome
- Goal: estimate the causal effect of X on Y

The “ABC” of estimating a causal effect

- Outcome model
  - $E[Y|X,C]$
- Propensity score model
  - $P(X=1|C)$
- Confounding adjustment methods
  - Matching (exact, coarse, PS matching or stratification)
  - Inverse weighting (IPW)
  - Regression adjustment
  - Double robust estimation
  - Many many many others!

Confounder selection and model uncertainty

Regardless of your favorite method for estimating a causal effect, you need to choose some covariates for confounding adjustment.

Which ones and how?

In the Glioblastoma example:

Tumor information

- Location, number, size, …

Patient comorbidities

- Diabetes, hypertension, COPD, …

Patient characteristics

- Gender, age, race, …

$2^{75} \times 2^{75}$: possible subsets of the covariates that can be included into the propensity score model and/or into the outcome model

The challenges of confounder selection

- For p large, if we
  - Miss even a single strong confounder bias
  - Include all covariates high variance
  - Include instruments high variance

- Brookhart, Schwneeiweiss, Wang, De Luna, Vansteeelandt, Wilson, Zigler, and many others have pointed out that statistical methods to select confounders in causal inference pose unique challenges that are different from selecting covariates in prediction problems

GOALS

To provide a practical and generalizable strategy for

- robustifying a broad range of causal inference methodologies
- accounting for model uncertainty
- selecting the confounders

The Average Causal Effect

$$\Delta = E\left[Y(1) - Y(0)\right] = E\left[E(Y|X=1,C) - E(Y|X=0,C)\right]$$

$$(Y(0), Y(1)) \perp\!\!\!\perp X|C.$$

- *Strong Ignorable Treatment Assignment*
- *No unmeasured confounders*

$$\mathrm{P}(X = 1|C) = e(C)$$

Propensity score model

$$E(Y|X = 0, C) = m_0(C),$$

$$E(Y|X = 1, C) = m_1(C)$$

Outcome models

Model Averaging in Causal Inference

Propensity score model class, $\boldsymbol{M}^{PS}$

- Logistic regressions for the treatment that include all possible subsets of the covariates as linear predictors logitP(X|C)=e(C)

Outcome model class, $\boldsymbol{M}^{OM}$

- Models for E[Y | X , C]
- Normal linear regressions that include the treatment as main effect and all possible subsets of the covariates as linear predictors

Joint model space, $\boldsymbol{M}$

- All possible combinations of models in M^{PS} and M^{OM}

An example of M^{PS} and M^{OM}

$$\text{logit}P(X=1) = a_0 + a_1C_1$$
$$\text{logit}P(X=1) = a_0 + a_1C_1 + a_2C_2$$
$$\text{logit}P(X=1) = a_0 + a_1C_1 + a_2C_2 + a_3C_3$$

2^p possible models

$$E(Y \mid X, C) = b_0 + b_1X + g_1C_1$$
$$E(Y \mid X, C) = b_0 + b_1X + g_1C_1 + g_2C_2$$
$$E(Y \mid X, C) = b_0 + b_1X + g_1C_1 + g_2C_2 + g_3C_3$$

2^p possible models

General Framework for Model Averaging in Causal Inference

$$\Delta = E[Y(1) - Y(0)]$$

$$\hat{\Delta} = \sum_{ij} [\hat{\Delta}_{ij}(M_i^{PS}, M_j^{om}) \times P(M_i^{PS}, M_j^{om} \mid \text{data})]$$

You can robustify your favorite estimator by averaging over model choices

From Doubly Robust to "*Bayesianly Sturdy*"

Requires specification of:

1. A method for estimating the causal effect (e.g. IPW, DR, regression)
2. A collection of

    propensity score models

    outcome models

        based on which covariates are included into these models
3. Prior distribution on the model space

    prior probability of including a covariate into the propensity score model or outcome model

DR Estimation

$$\widehat{\Delta}^{DR} = \frac{1}{n}\sum_{h=1}^{n}\left\{\frac{Y_h X_h - (X_h - \widehat{e}_h)\widehat{m}_{1h}}{\widehat{e}_h} - \frac{Y_h(1-X_h) + (X_h - \widehat{e}_h)\widehat{m}_{0h}}{1-\widehat{e}_h}\right\}$$

- The propensity score model and the outcome model can be selected in any number of ways, depending which set of covariates we decide to include *(or their functional form*).

- Here we illustrate the approach with the DR estimator b/c is regarded as semi-parametric. The DR estimation procedure only depends on the specification of the conditional means and does not rely on a fully parametric specification of the models (Bang and Robins).

Prior distribution on the model space

We introduce a prior distribution on the model space that:

1. a priori links the outcome and the propensity score model
2. assigns higher prior probabilities to propensity score models that include necessary confounders (C that are associated with both X and Y); and
3. assigns low prior probability to propensity score models that include covariates that are only associated with the exposure

A Dependent Prior on the Model Space (Luna et al 2011)

1. Identify covariates that are important predictors of the outcome
2. A priori, assume that the set of potential confounders included in the propensity score model must be a subset of those that are included in the outcome model (to exclude instruments) (tau=0)

$$\frac{\mathrm{P}(\mathcal{M}_i^{ps}|\mathcal{M}_j^{om})}{\mathrm{P}(\mathcal{M}_1^{ps}|\mathcal{M}_j^{om})} = \begin{cases} 1, & \text{if } \mathcal{M}_i^{ps} \subset \mathcal{M}_j^{om} \\ \tau, & \text{otherwise} \end{cases}$$

An example of P($M^{PS} \mid M^{OM}$)

Let's assume that C1,C2 and C3 are strong predictors of the outcome, then we assume a priori that:

1. C1 or C2 or C3 or any of their combinations have 50-50 chance to be included into the propensity score model
2. Any other covariate, say C5, that is not a predictor of the outcome has 0 chance to be included into the propensity score model

$$M_j^{om} : E(Y \mid X, C) = b_0 + b_1 X + g_1 C_1 + g_2 C_2 + g_3 C_3$$

$$M_i^{ps} : \log itP(X = 1) = a_0 + a_1 C_1 \text{ Þ } M_i^{ps} \text{ Ì } M_j^{om} \text{ Þ } odds(M_i^{ps} \mid M_j^{om}) = 1$$

$$M_{i+2}^{ps} : \log itP(X = 1) = a_0 + a_1 C_5 \text{ Þ } M_{i+2}^{ps} \text{ Ë } M_j^{om} \text{ Þ } odds(M_{i+2}^{ps} \mid M_j^{om}) = t \text{ @} 0$$

Posterior weights

1. Compute the marginal posterior probability of each outcome model under an uniform prior on the model space (every covariate has equal chance to be included into the outcome model)

$$q_j = \mathrm{P}(\mathcal{M}_j^{om}|\mathcal{D})$$

2. Compute the conditional posterior probability of each propensity score model under the dependent prior ($\tau \approx 0$) (if a covariate is already in the outcome model, its prior probability to be included in the PS model is 0.5, otherwise is 0)

$$\mathrm{P}(\mathcal{M}_i^{ps}|\mathcal{M}_j^{om},\mathcal{D})$$

3. Compute the posterior weights

$$w_{ij}^* = q_j \mathrm{P}(\mathcal{M}_i^{ps}|\mathcal{M}_j^{om},\mathcal{D})$$

Model Average Doubly Robust estimator (MADR)

1. For each combination of the PS model (i^{th}) and outcome model (j^{th}), calculate the posterior weights

$$w_{ij} = \mathrm{P}(\mathcal{M}_i^{ps}, \mathcal{M}_j^{om} | \mathcal{D})$$

1. Independently from step 1 and for each (i,j) calculate $\widehat{\Delta}_{ij}^{DR}$
2. Model averaging

$$\widehat{\Delta}_{DR}^{MA} = \sum_{ij} w_{ij} \widehat{\Delta}_{ij}^{DR}.$$

Independent and uniform prior

$$\mathrm{P}(\mathcal{M}_i^{ps}, \mathcal{M}_j^{om}|\mathcal{D}) = \mathrm{P}(\mathcal{M}_i^{ps}|\mathcal{D})\mathrm{P}(\mathcal{M}_j^{om}|\mathcal{D}).$$

- All models are a priori independent and equally likely
- Prioritize Cs that are strongly associated with X only
- Prioritize Cs that are strongly associated with Y only

not a good idea…

Asymptotic features

We have showed that if:

1. either the true propensity score model is contained in the model class $\boldsymbol{M}^{PS}$ OR
2. the true outcome model is contained in $\boldsymbol{M}^{OM}$

THEN

the posterior model probabilities are consistent for selecting the true models AND

MADR is consistent for the average causal effect

This result implies that we have added another layer of robustness to the double robust estimator, as we only need the true models to be in the collection of models.

From Doubly Robust to "*Bayesianly Sturdy*"

Our goals are to:

- account for model uncertainty in the propensity score model and the outcome model
- assign high posterior weights to models that include the most likely confounders (by specifying a prior distribution that links the propensity score model and the outcome model class)
- separate the procedure of calculating the posterior model weights (BMA with informative prior) from the procedure of estimating the average causal effects (DR estimation)
- computational efficiency and scalability

Simulations

n=200, p=100

$$C_1, ..., C_p \overset{iid}{\sim} N(0,1)$$

$$X \sim \text{Bernoulli}(p = \text{expit}(C\alpha^{ps}))$$

$$Y \sim \text{N}(\beta X + C\alpha^{om}, 1)$$

MCMC required

- *Model space too large to enumerate (2^{200} possible models!)*
- *We search for these models stochastically using a (MC3) (Madigan et al., 1995)*

Table 1: Description of all estimators used in the simulations. Included is (1) the type of estimator; and (2) the choice of prior distribution for the model space. All Bayes factors are estimated using the BIC approximation.

Estimator	Description
$\widehat{\Delta}_{DR}^{MS}$	model selected double robust estimator that chooses the propensity score model and the outcome model separately based on the BIC
$\widehat{\Delta}_{DR}^{MA}$	MA-DR estimator assuming prior model dependence defined by (5)
$\widehat{\Delta}_{DR}^{MA-II}$	MA-DR estimator assuming prior model dependence defined by (5) and using the two-stage approach for calculating model weights
$\widehat{\Delta}^{C-TMLE}$	Collaborative double robust targeted maximum likelihood estimator using the super learner to select the outcome regression with prediction algorithms including the full model, stepwise selection, and ridge regression

Table 2: Summary of Parameters for Simulation Group 1. In each of the 4 scenarios considered, we generate data as follows: (1) $C_1, ..., C_5 \overset{iid}{\sim} N(0,1)$; (2) $X \sim$ Bernoulli$(p = \text{expit}(C\alpha^{ps}))$; and (3) $Y \sim \text{N}(\beta X + C\alpha^{om}, \sigma^2)$ with $\beta = 1$ and $\sigma^2 = 4$. All effects of confounders are linear on both the treatment and outcome.

Scenario	Description	α^{ps} (PS model)	α^{om} (Outcome model)
1	No confounding	(0.4,0.3,0.2,0.1,0)	(0,0,0,0,0)
2	Moderate confounding	(0.5,0.5,0.1,0,0)	(0.5,0,1,0.5,0)
3	Strong predictors of outcome, weak predictors of treatment	(0.1,0.1,1,1,1)	(2,2,0,0,0)
4	Strong confounding	(0.5,0.4,0.3,0.2,0.1)	(0.5,1,1.5,2,2.5)

All simulations set β=1

- Scenario 1: no confounding
- Scenario 2: moderate confounding
- Scenario 3: (C1,C2): strong predictors of outcome and weak predictors of treatment , (C3,C4,C5): strong predictor of X only (instrumental variables)
- Scenario 4: strong confounding

Simulations: Relative efficiency (to the true)
p=100, n=200

(C1, C2) Y (strongly)
(C1,C2) -X (weakly)
(C3,C4,C5) X (instruments)

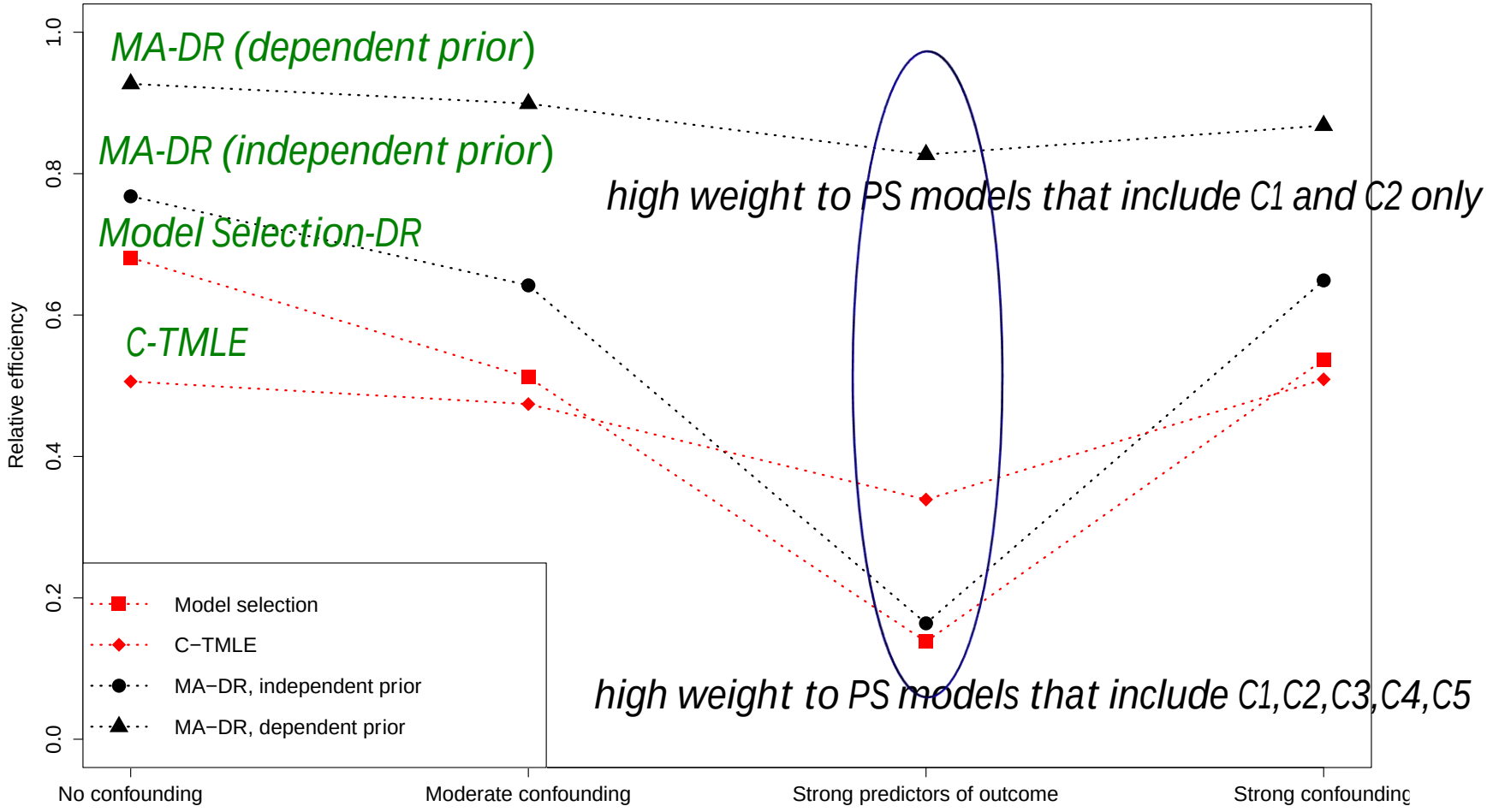


Table 4: The mean, standard error, relative efficiency (variance of the gold standard divided by the variance of the estimator), and 95% confidence interval coverage probability of various estimators when $n = 200$ and $p = 100$ for 500 replications of the data.

		Scenario 1				Scenario 2			
	τ	Mean	Std. Error	Rel. Eff.	Coverage	Mean	Std. Error	Rel. Eff.	Coverage
Gold standard	-	1.00	0.14	1.00	0.95	1.00	0.15	1.00	0.94
$\widehat{\Delta}_{DR}^{MS}$	-	1.00	0.17	0.66	0.93	1.00	0.19	0.62	0.95
$\widehat{\Delta}^{C-TMLE}$	-	1.01	0.20	0.45	0.93	0.99	0.26	0.34	0.98
$\widehat{\Delta}_{DR}^{MA}$	1	1.00	0.16	0.72	0.95	1.00	0.18	0.71	0.94
$\widehat{\Delta}_{DR}^{MA-II}$	0	1.00	0.14	0.91	0.94	1.00	0.16	0.94	0.93

		Scenario 3				Scenario 4			
	τ	Mean	Std. Error	Rel. Eff.	Coverage	Mean	Std. Error	Rel. Eff.	Coverage
Gold standard	-	1.00	0.14	1.00	0.95	1.00	0.16	1.00	0.95
$\widehat{\Delta}_{DR}^{MS}$	-	1.00	0.33	0.19	0.95	1.00	0.19	0.70	0.96
$\widehat{\Delta}^{C-TMLE}$	-	0.97	0.36	0.16	1.00	1.01	0.30	0.27	0.97
$\widehat{\Delta}_{DR}^{MA}$	1	1.00	0.35	0.17	0.96	1.00	0.17	0.80	0.96
$\widehat{\Delta}_{DR}^{MA-II}$	0	1.00	0.15	0.87	0.97	1.01	0.16	0.95	0.95

Simulation Study Results

- We leverage information from the data to reduce the set of potential confounders to a manageable size
- We have found that for a set of potential confounders that is half the size of the number of observations, the proposed estimator dramatically improves efficiency when compared with model selection
- It can be scaled up to consider larger sets of potential confounders (Antonelli et al submitted)

Glioblastoma

Question

Is there a survival benefit of adding TMZ to radiotherapy for patients 65 and older with GBM?

X=1 radiotherapy + TMZ

X=0 radiotherapy alone

Data source

SEER linked Medicare data

- 2005-2009
  - 1887 cases of Glioblastoma
- Inpatient and outpatient data
- 75 covariates

Table 6: Baseline characteristics (% experiencing) and 1-year mortality rate for patients treated with temozolomide plus radiotherapy and radiotherapy alone, along with estimated inclusion probabilities (assuming prior model dependence defined by Equation 5 and using the two-stage approach with $\tau = 0$) for the analyses of the Medicare data.

	Radiotherapy only ($n = 1111$)	Temozolomide ($n = 776$)	$Pr\left(\text{Inclusion in outcome model}\right)$	$Pr\left(\text{Inclusion in propensity score}\right)$
Age group:				
65-69	25.6	30.7	*Reference*	*Reference*
70-74	28.1	32.1	**1.00**	0.07
75-79	23.8	23.3	**1.00**	0.09
≥80	22.5	13.9	**1.00**	**0.95**
Gender (female/male)	47.1	48.5	0.09	0.08
Married (yes/no)	64.9	71.0	0.03	0.03
Race (white/other)	85.4	89.9	0.10	0.10
Income (high/low)	63.0	67.8	**1.00**	0.23
Region:				
Northeast	21.7	20.3	*Reference*	*Reference*
Midwest	14.5	13.1	0.06	0.06
South	17.6	21.5	0.01	0.01
West	46.2	45.1	0.03	0.03
Dual eligible (yes/no)	14.3	11.6	0.02	0.02
Resection (yes/no)	73.2	79.1	**1.00**	0.48
CT scan(yes/no)	79.6	76.8	**1.00**	0.08
MRI (yes/no)	76.5	83.4	0.13	0.13
Initial discharge (home/other)	45.4	55.3	**1.00**	**0.94**
Extent of resection (major/minimal)	28.5	29.9	**1.00**	0.14
Atherosclerosis (yes/no)	17.8	14.4	**0.72**	0.05

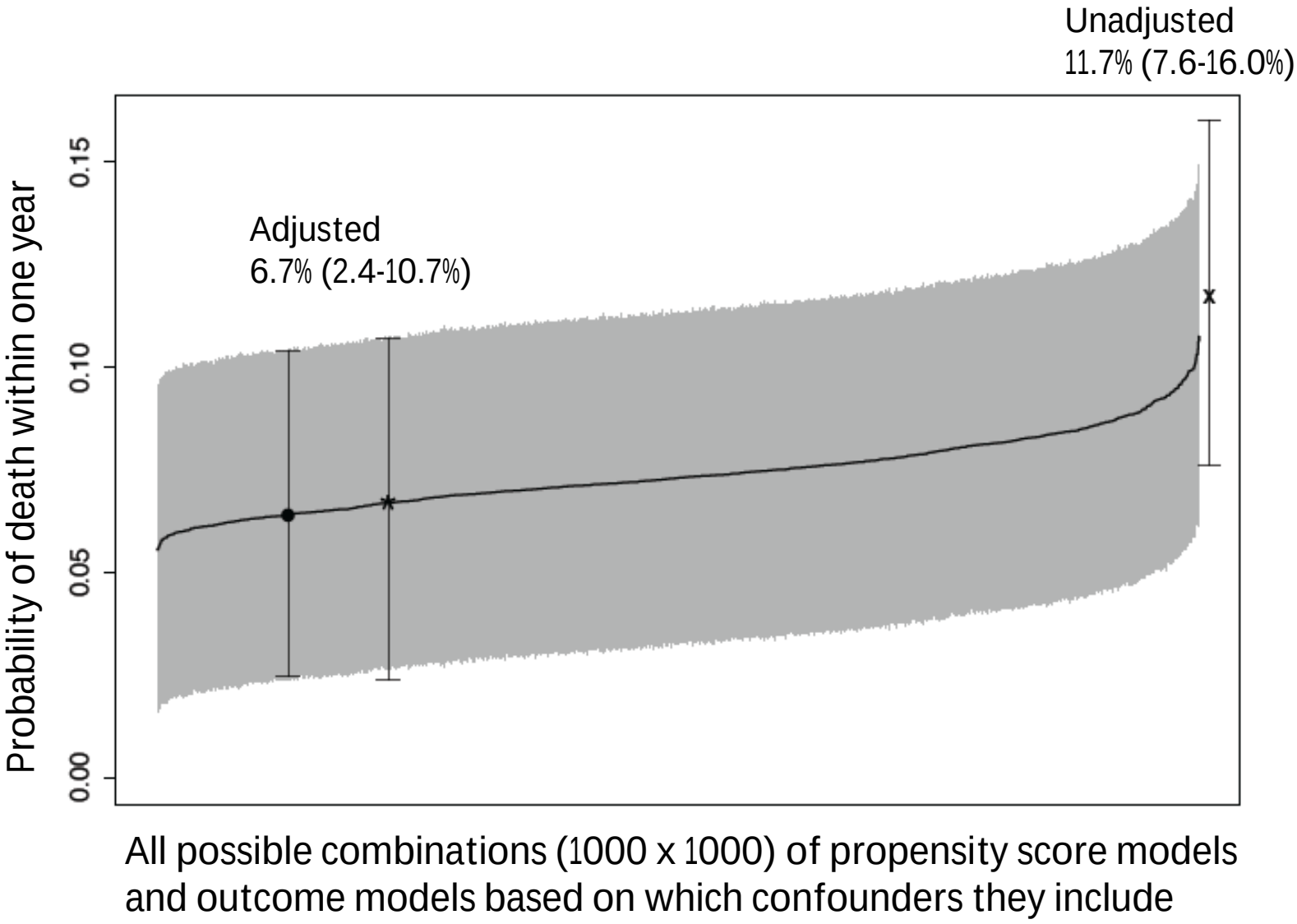


Figure 1: Model specific double robust estimator for the glioblastoma example and corresponding 95% confidence intervals for 1000 randomly chosen outcome models and 1000 randomly chosen propensity score models, sorted by point estimates. The unadjusted estimator (X), the model averaged double robust estimator (∗), and the double robust estimator that includes all covariates into both models (•) are included.

Features of MA-DR

- It can be applied to different types of causal estimators (IPW, DR, standard regression)
- It separates estimation from model selection, thus being less dependent on model assumptions and more computationally efficient
- The "dependent prior" allows to exclude from the outcome model the covariates that are only associated with the treatment and to assign higher weight to models that include the likely confounders
- It outperform the competitors under several scenarios of degree of confounding and in high dimensional case

The MA-DR estimators extend the desirable double robustness property by achieving consistency under the much weaker assumption that either the true propensity score model or the true outcome model be within a specified, possibly large, class of models

Other approaches: similar philosophy

Wang et al 2012, 2015 (Biometrics)

- Outcome model: GLM
- X: continuous or binary
- Confounding adjustment: regression

Zigler et al 2014 (JASA)

- Outcome model: GLM
- X: binary
- Confounding adjustment: PS as a linear predictor

Wilson et al 2017 (Biometrics, under revision)

- Outcome model: Linear Regression
- X: continuous but multivariate, multiple treatments and their interactions
- Confounding adjustment: regression

Limitations

- We have focused our simulations on situations where the model classes include models that vary only by the confounders that are selected, and the data generating model is included in the class
- More work is needed to determine the relative merits of the model averaged double robust estimators when these assumptions are relaxed
- Does not deal with unmeasured confounding or mediation
- Studying the properties of alternate weights and the use of model averaging on other estimators in causal inference is an exciting line of research.

Cefalu M, Dominici F, Arvold N, Parmigiani G (2016) A Model averaged double robust estimator, Biometrics, DOI:10.1111/biom.12622

"madr" package in CRAN

Nils Arvold, MD
Attending Physician, Radiation Oncology Associates, St. Luke's Cancer Center and University of Minnesota Duluth

Matthew Cefalu, PhD
Associate Statistician at the RAND Corporation

Former Postdoctoral Fellow, Department of Biostatistics, Harvard T.H. Chan School of Public Health

Giovanni Parmigiani, PhD
Professor of Biostatistics DFCI and Harvard TH Chan School of Public Health

R package (madr)

Package 'madr'

September 5, 2016

Type Package

Title Model Averaged Double Robust Estimation

Version 1.0.0

Author Matthew Cefalu

Maintainer Matthew Cefalu <Matthew_Cefalu@rand.org>

Description Estimates average treatment effects using model average double robust (MA-DR) estimation. The MA-DR estimator is defined as weighted average of double robust estimators, where each double robust estimator corresponds to a specific choice of the outcome model and the propensity score model. The MA-DR estimator extend the desirable double robustness property by achieving consistency under the much weaker assumption that either the true propensity score model or the true outcome model be within a specified, possibly large, class of models.

License GPL-3

LazyData TRUE

RoxygenNote 5.0.1

NeedsCompilation no

References

Wang, C., G. Parmigiani, and F. Dominici (2012). Bayesian effect estimation accounting for adjustment uncertainty. Biometrics 68(3), 661–671.

Wang C, Dominici F, Parmigiani G, Zigler CM (2015) Accounting for uncertainty in confounder and effect modifier selection when estimating average causal effects in generalized linear models. *Biometrics,* 71(3):654-65. PMCID: PMC4575246.

Zigler CM, Watts K, Yun Wang, Coull B, Dominici F (2013) Model Feedback in Bayesian propensity score estimation. *Biometrics,* 69(1):263-73. PMCID: PMC3622139.

Zigler CM, Dominici F (2014) Uncertainty in propensity score estimation: Bayesian methods for variable selection and model averaged causal effects. *Journal of the American Statistical Association,* 109(505):95-107. PMCID: PMC3703764

Wilson A, Zigler CM, Patel CJ, Dominici F (2017) A. Model-Averaged Confounder Adjustment for Estimating Multivariate Exposure Effects. *Biometrics* (under revision)

Antonelli JL, Zigler CM, Dominici F. (2017). Guided Bayesian imputation to adjust for confounding when combining heterogeneous data sources in comparative effectiveness research. *Biostatistics.* PMID: 28334230 DOI: 10.1093/biostatistics/kxx003

Antonelli JL, Parmigiani G, Dominici F. (2017) High dimensional confounding adjustment using continuous spike and slab priors. *Biometrics (submitted)*