Approaches to inverse-probability-of-treatment—weighted estimation with concurrent treatments

Alan R. Ellis\textsuperscript{a}, M. Alan Brookhart\textsuperscript{a,b,*}

\textsuperscript{a}Cecil G. Sheps Center for Health Services Research, University of North Carolina at Chapel Hill, CB# 7590, Chapel Hill, NC 27599-7590, USA
\textsuperscript{b}Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, 2105F McGavran-Greenberg, CB# 7435, Chapel Hill, NC 27599-7435, USA

Accepted 13 March 2013

Abstract

Objectives: In a setting with two concurrent treatments, inverse-probability-of-treatment weights can be used to estimate the joint treatment effects or the marginal effect of one treatment while taking the other to be a confounder. We explore these two approaches in a study of intravenous iron use in hemodialysis patients treated concurrently with epoetin alfa (EPO).

Study Design and Setting: We linked US Renal Data System data with electronic health records (2004–2008) from a large dialysis provider. Using a retrospective cohort design with 776,203 records from 117,050 regular hemodialysis patients, we examined a composite outcome: mortality, myocardial infarction, or stroke.

Results: With EPO as a joint treatment, inverse-probability-of-treatment weights were unstable, confidence intervals for treatment effects were wide, covariate balance was unsatisfactory, and the treatment and outcome models were sensitive to omission of the baseline EPO covariate. By handling EPO exposure as a confounder instead of a joint treatment, we derived stable weights and balanced treatment groups on measured covariates.

Conclusions: In settings with concurrent treatments, if only one treatment is of interest, then including the other in the treatment model as a confounder may result in more stable treatment effect estimates. Otherwise, extreme weights may necessitate additional analysis steps.

Keywords: Propensity score; Statistics as topic; Models; Statistical; Epidemiologic methods; Estimation

1. Introduction

Inverse-probability-of-treatment—weighted (IPTW) estimation is now commonly used to control for confounding in nonexperimental studies of medical interventions [1–3]. IPTW estimation requires that the analyst specify a model of the treatment rather than the outcome. The fitted treatment model is then used to estimate inverse-probability-of-treatment weights that are applied to each subject. If the treatment model is correctly specified, the reweighting results in a population of patients in whom treatment assignment is unrelated to the baseline variables that are included in the treatment model. Under the assumptions of exchangeability, positivity, consistency, and a correctly specified treatment model [1,4,5], IPTW estimation results in estimates that can be interpreted as the average treatment effect (ATE) in the population being studied. (Informally, exchangeability refers to the absence of unmeasured confounders, positivity requires that each subject have a non-zero probability of receiving each of the treatments being compared, and consistency means that the observed outcome equals the counterfactual outcome of the treatment actually received.) The IPTW approach is attempting to mimic a situation in which treatment is randomly allocated to individuals.

IPTW estimation can be extended to settings with concurrently administered treatments. Herein, one is attempting to estimate the average joint effect of the treatments. Thus, IPTW estimation is attempting to mimic a situation in which each concurrent treatment is randomized individually. This requires that one construct a model of the probability that a patient would receive any particular combination of the treatments being studied.

Conflicts of interest: In the past 5 years, A.R.E. has received research funding from the University of North Carolina Center for Pharmacoepidemiology, which receives industry support, and from Merck. M.A.B. has received research support from Amgen, Rockwell Medical, and Pfizer. M.A.B. has sat on advisory boards for Amgen and Pfizer but has not accepted honoraria for these activities.

* Corresponding author. Tel.: +919-843-2639; fax: +919-966-2089.
E-mail address: abrookhart@unc.edu (M.A. Brookhart).

0895-4356 © 2013 Elsevier Inc. Open access under CC BY-NC-ND license.
http://dx.doi.org/10.1016/j.jclinepi.2013.03.020
2. Background

2.1. Anemia management in hemodialysis patients

Anemia affects about 10% of patients in the early stages of chronic kidney disease and more than 70% of patients with end-stage renal disease (ESRD) [10]. The anemia of ESRD is primarily caused by impaired production of renal erythropoietin. It is worsened by dialysis-related blood loss, which depletes iron reserves [11]. The anemia of ESRD involves treatment with ESAs to stimulate the production of red blood cells and administration of intravenous iron to address iron deficiency [12].

Several biological mechanisms suggest potential risks associated with the use of iron [13]. For example, frequent iron administration may lead to oversaturation of transferrin and the release of free, catalytically active iron into the plasma [14]. Free iron is also known to catalyze the formation of highly reactive oxygen species [15,16]. These could give rise to lipid radicals, which may damage the vasculature [17] and lead to atherogenesis [18], possibly increasing the long-term risk of cardiovascular events [13,19]. However, little is known about the relation between iron dose and cardiovascular outcomes.

We recently undertook a study in which we used data from the US Renal Data System linked with data from dialysis providers to assess the relative safety and effectiveness of different iron formulations and dosing strategies. We used multivariable models to examine multiple outcomes, including cardiovascular events and mortality. To minimize bias owing to confounding factors, we sought to implement IPTW estimation. Because joint treatment with ESAs and iron complicated the IPTW analysis, we conducted sensitivity analyses to identify the best way to model treatment. To demonstrate the sensitivity of treatment effect estimates to the treatment model, we used as an example a composite outcome: mortality, myocardial infarction, or stroke. In this article, we report the results of our sensitivity analyses and comment briefly on the handling of concurrent treatments in IPTW analyses.

3. Methods

3.1. Data, study design, and sample

Funded by the National Institute of Diabetes and Digestive and Kidney Diseases, the US Renal Data System collects, analyzes, and distributes information about ESRD treatment in the United States, including data from the Medical Evidence Report Form, the Medicare Enrollment Database, the ESRD Death Notification Form, and the standard analytic files, which contain final action claims data [20]. We linked US. Renal Data System data with 5 years of electronic health record data (2004–2008) from a large US dialysis provider that owns and manages more than 1,500 outpatient dialysis facilities throughout the United States.

IPTW estimates can be highly unstable in the presence of large weights because the estimates may be driven by outcomes occurring in a small number of patients [6–8]. This can happen if patients are treated contrary to indication, the population is poorly defined (i.e., includes patient subgroups that rarely receive treatment), or the treatment model is misspecified (i.e., predicted probabilities of treatment are incorrect). Large weights may be particularly likely in settings with concurrent treatments, where there may be many treatment categories, some treatment combinations may be uncommon, and correct specification of the treatment model may be more difficult.

In a setting with concurrent treatments, if interest focuses primarily on the effect of one treatment, IPTW estimation can be used to estimate the marginal effect of the treatment of interest, taking the concurrently used treatments as confounders. This simplifies the estimation of the treatment model. It also imposes fewer assumptions on the analysis because exchangeability, positivity, and consistency need to hold for only one treatment.

We explore these issues in a study of iron treatment outcomes in hemodialysis patients, a setting in which concurrent use of erythropoiesis-stimulating agents (ESAs) must be addressed and extreme weights are known to be problematic [9]. Specifically, our study examines the effect of iron treatment administered at a point in time, as measured during a 1-month exposure period, on a composite cardiovascular outcome.

What is new?

- Inverse-probability-of-treatment—weighted (IPTW) estimation can be applied to two or more concurrently administered treatments.
- In this setting, IPTW estimation may be more likely to result in large weights and estimates that are very sensitive to model specification.
- If only one treatment is of interest, then including the other treatment in the treatment model as a confounder may result in more stable estimates of the effect of interest.
- In studies of concurrent treatments, analysts using IPTW methods should be vigilant for problems arising from large weights. In some settings, it may be more feasible to estimate the marginal effect of a single treatment.
- If a joint treatment effect is of interest, extreme inverse probability of treatment weights may need to be addressed by excluding observations whose covariate values are rare in any treatment group, by limiting the analysis to common treatment regimes, or through other means.
States. Demographic characteristics, comorbidities, and hospital use were measured in the U.S. Renal Data System; epoetin alfa (EPO) and iron exposure, laboratory values, and intravenous antibiotic use were measured in the clinical database. EPO is a commonly used ESA.

We used a retrospective cohort design. Because transferrin saturation is used to guide iron administration, we aligned the 1-month exposure period so that it began the day after a transferrin saturation laboratory test. We used a 6-month baseline period and followed patients for up to 3 months after exposure. The data structure was designed to allow us to estimate the effect of iron treatment during a single 1-month exposure period on short-term outcomes. We allowed multiple observations per patient, with a minimum of 90 days between index transferrin saturation laboratory tests (i.e., nonoverlapping follow-up periods); each observation included a 6-month baseline period that immediately preceded the exposure period.

Our sample included incident and prevalent hemodialysis patients who were covered by Medicare Part A and Part B, received regular center hemodialysis during the baseline and exposure periods, and received either ferric gluconate or iron sucrose, but not both, during the exposure period. We required at least 3 months of dialysis before baseline to allow for stability in claims processing [21], and excluded patients with polycystic kidney disease and observations with missing covariate information. The final sample included 776,203 records from 117,050 patients.

3.2. Measures

We used Medicare claims data from the US Renal Data System to measure the composite outcome: mortality, myocardial infarction, or stroke. One-month iron exposure, measured using the detailed clinical data, was classified as high (>200 mg), low (1–200 mg), or none. One-month EPO exposure was coded in tertiles. We measured 23 baseline covariates, including age, sex, race, comorbidities, years on dialysis, infection during baseline, and the use of a catheter as the primary vascular access.

3.3. Statistical analysis

We used a generalized logit model to predict EPO treatment category given the baseline covariates. We then used a second generalized logit model to predict iron treatment category given EPO treatment and covariates. Although we could have used ordinal logistic models, such models require the proportional odds assumption, meaning that a single coefficient describes the relation between the odds of being assigned to a given dose level and the odds of being assigned to the next higher dose level. We believed that this assumption was unrealistic for our data.

For each observation, the generalized logit models yielded one predicted probability per treatment category. That is, each observation received three predicted probabilities for EPO and three predicted probabilities for iron. We selected the estimates for the EPO and iron treatments actually received (i.e., one predicted probability for EPO and another for iron). For each observation, we estimated the probability of receiving the treatment actually received as Pr(actual EPO treatment category | covariates) × Pr(actual iron treatment category | actual EPO treatment category, covariates). The inverse-probability-of-treatment weights were calculated as the reciprocal of this value. Assuming that the treatment model is correct, these weights create a pseudopopulation in which confounders are unrelated to treatment assignment. Therefore, the weights can be used to make an unbiased estimate of the treatment effect in the population. We stabilized the weights by multiplying them by the marginal probability of receiving the treatment combination actually received. We suspected that the handling of baseline EPO treatment, like the handling of EPO exposure, might affect the stability of our results. Therefore, our sensitivity analysis included 12 different treatment models, representing all possible combinations of three factors:

1. Modeling of EPO exposure (as joint treatment vs. confounder)
2. Baseline EPO covariate: total number of units of EPO received during the last month of baseline (categorical variable with five levels, categorical variable with three levels, or excluded)
3. Patients receiving no EPO during the exposure period and no EPO during the last month of baseline (included vs. excluded)

To clarify the contribution of the EPO exposure variable to the variability of the inverse-probability-of-treatment weights, we also ran six models predicting only iron treatment and ignoring EPO treatment. In these models, we varied factors 2 and 3: the baseline EPO covariate and the handling of the no-EPO group. Weights were calculated and stabilized in the same way as before based on the probability of receiving the iron treatment actually received.

We evaluated treatment models with regard to the distribution of the stabilized weights, covariate balance, and stability of treatment effect estimates. Because we were interested in contrasting treatment groups with each other, covariate balance was measured using standardized absolute mean difference on covariates between the high iron dose group and each of the other two iron dose groups. Considering all the pairwise standardized absolute mean differences together, we used the median and maximum standardized absolute mean difference to characterize the level of covariate balance. The treatment effect of high iron dose vs. low iron dose was estimated using Cox proportional hazard models of the time to death, first myocardial infarction, or first stroke, using the weights to adjust for confounding. Censoring occurred at transplant, loss to follow-up, or the end of available data. To account for the clustering of observations within patients, we used the robust sandwich estimator to derive standard errors.
In addition to assessing stability of treatment effect estimates for the comparison of primary interest (high vs. low iron dose), we held factors 2 and 3 constant (classifying baseline EPO into five levels and including patients who received no EPO) and compared the EPO-as-joint-treatment approach to the EPO-as-confounder approach with regard to main effects and iron—EPO interactions.

This study was approved by the University of North Carolina Institutional Review Board.

4. Results

4.1. Models predicting joint EPO and iron treatment

When the treatment model predicted both EPO and iron treatment, the stabilized weights were variable and ranged as high as 28,050 (Table 1). Balance generally improved compared with the pre-propensity—scoring results (median standardized absolute mean difference 0.06, maximum 0.55), but imbalance remained on some covariates. Omitting baseline EPO gave the weights a mean of 1.0 and made them less variable. Excluding nonrecipients of EPO had little effect on weight distribution and no clear effect on balance. Treatment effect estimates for high vs. low iron were sensitive to the handling of the no-EPO group and the baseline EPO covariate (Fig. 1). Confidence intervals were wider when the baseline EPO covariate was included.

4.2. Models predicting only iron treatment, with EPO treatment as a confounder

Including EPO treatment as a confounder in the treatment model, instead of as a joint treatment, resulted in stabilized weights with a mean of 1.0 and much less variation, regardless of how the baseline EPO covariate and EPO nonrecipients were handled (Table 2). This set of models also resulted in improved covariate balance (Table 2) and stable treatment effect estimates (Fig. 2). The models that ignored EPO treatment had nearly identical results (not shown).

4.3. Main effects and interactions when holding factors 2 and 3 constant

When we fixed the number of baseline EPO categories at 5 and included patients with no EPO exposure, the model handling EPO as a joint treatment yielded hazard ratios of 1.17 (95% confidence interval [CI]: 0.70—1.96) for high vs. low iron and 0.49 (95% CI: 0.21—1.10) for high vs. no iron. Hazard ratios for high vs. medium EPO and high vs. low EPO were 0.90 (95% CI: 0.70—1.17) and 0.39 (95% CI: 0.13—1.15), respectively.

In the model handling EPO as a confounder, the hazard ratio for high vs. low iron was 1.05 (95% CI: 1.00—1.09), and the hazard ratio for high vs. no iron was 1.00 (95% CI: 0.96—1.06). Confidence intervals were narrower for this model than for the model handling EPO as a joint treatment. This model did not allow estimation of causal effects for EPO.

Fig. 3 shows the estimated hazard ratios for high vs. low dose iron within EPO tertiles by treatment model. When EPO was handled as a joint treatment, confidence intervals were much wider, especially for the highest and lowest tertiles of EPO exposure. When EPO was handled as a confounder, hazard ratio estimates were more consistent. All confidence intervals included 1.0.

5. Discussion

We considered approaches to IPTW estimation in a setting in which two treatments are used concurrently but one is of primary interest. In our study of treatment of anemia in...
hemodialysis patients, the treatment of primary interest was intravenous iron and the concurrently used treatment was EPO. We conducted a series of analyses in which we varied the treatment model, the handling of the baseline EPO covariate, and the inclusion vs. exclusion of patients who received EPO during neither the baseline period nor the exposure period. When the stabilized weights were based on a model predicting joint treatment, they were poorly behaved. They did not have the usual mean of 1.0 and were highly variable. Although handling EPO as a joint treatment allowed estimation of causal effects for both iron and EPO as well as interaction effects, the variability in the weights led to wide confidence intervals for treatment effect estimates. Also, covariate balance in the reweighted population was poor, and the treatment and outcome models were sensitive to the omission of the baseline EPO covariate.

By including EPO treatment as a confounder instead of a joint treatment, we were able to derive stable weights and to balance the treatment groups on measured covariates. Although some residual imbalance remained, it was negligible for most covariates (median standardized absolute mean difference 0.01), small (0.21 standard deviation units) even for the covariate with the most residual imbalance, and acceptable for our purposes. Residual imbalance can be decreased by making the treatment model more flexible (e.g., adding interaction or spline terms or using nonparametric estimation). When we ran treatment models that ignored EPO treatment entirely (not recommended in practice), the resulting covariate balance and weight distribution were nearly identical, suggesting that instability owing to the EPO treatment variable disappeared entirely when EPO was included as a confounder instead of a joint treatment.

If both of the concurrent treatments had been of substantive interest, extreme weights would have required us to take additional steps to have an identifiable treatment effect, such as excluding observations whose covariate values were rare in any treatment group [22], restricting the set of treatments to be examined [23], or truncating weights to reduce variability [24]. Weight truncation simply reduces extreme weights directly to decrease reliance on a few highly influential observations that were treated contrary to expectation, decreasing standard errors at the expense of some residual bias. The other two steps, by removing rare covariate combinations or rare treatment assignments, eliminate

Table 2. Weight distribution and covariate balance by treatment model, for models predicting only iron treatment with EPO treatment as a confounder

<table>
<thead>
<tr>
<th>No-EPO group</th>
<th>Baseline EPO variable</th>
<th>Weight mean (SD)</th>
<th>Weight maximum</th>
<th>SAMD median</th>
<th>SAMD maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Included</td>
<td>Five categories</td>
<td>1.0 (1.6)</td>
<td>85.8</td>
<td>0.01</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>Three categories</td>
<td>1.0 (1.6)</td>
<td>82.5</td>
<td>0.01</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>Excluded</td>
<td>1.0 (1.6)</td>
<td>84.6</td>
<td>0.01</td>
<td>0.21</td>
</tr>
<tr>
<td>Excluded</td>
<td>Five categories</td>
<td>1.0 (1.6)</td>
<td>87.4</td>
<td>0.01</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>Three categories</td>
<td>1.0 (1.6)</td>
<td>85.0</td>
<td>0.01</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>Excluded</td>
<td>1.0 (1.6)</td>
<td>86.4</td>
<td>0.01</td>
<td>0.21</td>
</tr>
</tbody>
</table>

Abbreviations: EPO, epoetin alfa; SD, standard deviation; SAMD, standardized absolute mean difference.
situations that lead to extreme weights. In the current setting, only iron treatment was of interest, and including EPO treatment as a confounder was an effective way to achieve stable inverse-probability-of-treatment weights and still prevent confounding by balancing the treatment groups on measured covariates.

One subtle drawback to handling concurrent treatments as confounders is that they become characteristics of the population for whom the ATE is being estimated. Therefore, if the use of concurrent treatments changes, the ATE of the primary treatment may also change. This may limit the generalizability of the estimate to other patient populations or time periods where use of the concurrent treatments may be different. This can be addressed by stratifying the analysis over levels of the concurrent treatment. Subgroup effects can then be standardized to a population with a different pattern of concurrent treatment.

Our analysis suggests that IPTW estimation of the joint effects of concurrent treatments may be challenging in some situations because it complicates estimation of the treatment model, imposes more assumptions on the analysis, and may be more likely to involve extreme weights. If one treatment is of primary interest, it may be preferable to take the concurrently administered treatment as a confounder. This may result in estimates that are more stable and valid.

Acknowledgments

This work was supported by a contract from the Agency for Healthcare Research and Quality (AHRQ), US Department of Health and Human Services (DHHS), to The University of North Carolina Developing Evidence to Inform Decisions about Effectiveness Center (contract no. HHSA29020050040).

Some of the data reported here have been supplied by the United States Renal Data System. The authors of this report are responsible for its content. No statement in the report should be construed as endorsement by AHRQ or DHHS, and in no way should the report be seen as an official policy or interpretation of the US Government. The authors thank DaVita Clinical Research for providing data for this study. DaVita Clinical Research had no role in the design or implementation of this study or in the decision to publish. Finally, the authors would like to acknowledge helpful comments from Iván L. Díaz.

References