# Propensity scores for repeated treatments: A tutorial for the `iptw` function in the `twang` package

Lane Burgette, Beth Ann Griffin and Dan McCaffrey [*]
RAND Corporation

December 22, 2025

## 1 Introduction

While standard propensity score methods attempt to answer the question of how expected outcomes change if a group of individuals received one treatment instead of another, researchers are often interested in understanding how *sequences* of treatments impact outcomes of interest. In this case, time-varying confounders may be impacted by prior treatments. Consequently, simply controlling for the time-varying confounders in standard regression models can yield biased results. Instead, it is possible to perform weighted regressions that account for time-varying confounders via *marginal structural models* (MSMs; Robins et al., 2000). In this method, observations are weighted by the inverse of the estimated probability of receiving the observed sequence of treatments the individual actually received, referred to as an *inverse probability of treatment weight* (IPTW). It has been proposed to use nonparametric models to estimate IPTWs (Griffin et al., 2014). Accordingly, we refer to the function in `twang` that performs this weighting as `iptw`, for inverse probability of treatment weighting.

For binary treatments, the `iptw` methods and syntax build directly on the `ps` functionality; users are encouraged to study that tutorial before using `iptw`. For treatment regimes with more than two categories, the `iptw` methods build on the `mnps` methods and software. For more background on marginal structural models, see e.g., Robins et al. (2000) and Cole and Hernán (2008).

## 2 An IPTW example

For the sake of illustration, we simulated data to demonstrate the functionality of the `iptw` command. For time-varying treatment data, one can either imagine a "wide" dataset, with one row per subject, or a "long" dataset with one row for each subject/time point combination. Our artificial data include time-invariant characteristics `gender`, and `age` at time of study enrollment. Conceptually, we have a substance use index that is measured four times: at baseline, after the first treatment period, after the second treatment period, and after the third treatment period, which concludes the study and is the outcome of interest. In the "wide" version of the dataset called `iptwExWide`, we have the `outcome`, baseline and intermediate measures, `use0`, `use1`, and `use2`. The treatment indicators are, in chronological order, `tx1`, `tx2`, and `tx3`. Our goal is to

[*] The development of this software and tutorial was funded by National Institute of Drug Abuse grants number 1R01DA015697 (PI: McCaffrey) and 1R01DA034065 (PIs: Griffin/McCaffrey).

estimate the average effect of each additional dose of treatment on substance use measured at the end of the study (which is recorded in outcome).

The "long" format data have a somewhat different form, and are included in the data object iptwExLong. For the long format, the outcomes are split from the covariates, and are available as iptwExLong$outcome. Similarly, the covariates and treatment indicators are available in covariates, which includes data elements gender, age, use, and tx; these include the same information as the wide version. Additionally, the long version contains elements ID (an individual-level identifier) and time, which corresponds to the study period.

One of the benefits of GBM for applied researchers is the automatic handling of missing data. For MSMs, however, this does not extend to partially observed treatment histories. We assume throughout that missingness exists only in the covariates.

## 2.1 Fitting the model

To begin, we will work with the "wide" version of the data, which are available after loading the twang package:

```
> library(twang)
> data(iptwExWide)
```

Next, we can fit the model using the iptw function. Unlike for the standard ps function, we are only able to use a single stop.method at a time. The treatment assignment models are specified as a list of formulas, starting at the earliest time period. For coding parsimony, terms that should appear in all of the formulas can be specified once via a one-sided formula using the timeInvariant argument. Similarly, including treatment indicators from previous models is achieved by setting priorTreatment = TRUE. Finally, if all terms included at period $t$ should be included in the period $t + 1$ model (as is typically the case in MSM models), setting cumulative = TRUE automatically includes all elements on the right-hand side of previous models.

Thus, the model

```
> iptw.Ex <- iptw(list(tx1 ~ use0 + gender + age,
+                      tx2 ~ use1 + use0 + tx1 + gender + age,
+                      tx3 ~ use2 + use1 + use0 + tx2 + tx1 + gender + age),
+                 timeInvariant ~ gender + age,
+                 data = iptwExWide,
+                 cumulative = FALSE,
+                 priorTreatment = FALSE,
+                 verbose = FALSE,
+                 stop.method = "es.max",
+                 n.trees = 5000)
```

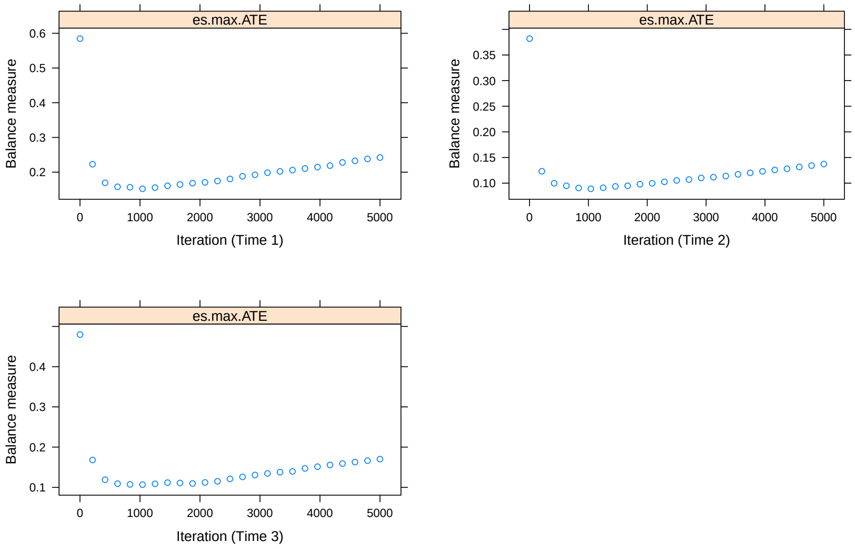
can be specified more simply as:

```
> iptw.Ex <- iptw(list(tx1 ~ use0, tx2 ~ use1, tx3 ~ use2),
+                 timeInvariant ~ gender + age,
+                 data = iptwExWide,
+                 cumulative = TRUE,
+                 priorTreatment = TRUE,
+                 verbose = FALSE,
+                 stop.method = "es.max",
+                 n.trees = 5000)
```

After having fit the `iptw` object, the diagnostic checks are similar to those specified for `ps` objects.

First, we check to make sure that the GBM models were allowed to run long enough (i.e., `n.trees` is sufficiently large).

```
>     plot(iptw.Ex, plots = 1)
```



Next, we can get a quick sense of the balance at each timepoint via

```
> summary(iptw.Ex)

Summary for time period  1 :
           n.treat n.ctrl ess.treat ess.ctrl    max.es   mean.es     max.ks
unw            706    294  706.0000 294.0000 0.5891037 0.4095762 0.29446339
es.max.ATE     706    294  655.4712 216.1356 0.1510383 0.1141335 0.08302253
           max.ks.p    mean.ks iter
unw              NA 0.22414613   NA
es.max.ATE       NA 0.06700952 1117

Summary for time period  2 :
           n.treat n.ctrl ess.treat ess.ctrl     max.es    mean.es     max.ks
unw            508    492    508.00 492.0000 0.38549444 0.23634295 0.19268933
es.max.ATE     508    492    475.83 450.1773 0.08850081 0.05137071 0.04422542
           max.ks.p    mean.ks iter
unw              NA 0.13255874   NA
es.max.ATE       NA 0.03667021  966

Summary for time period  3 :
           n.treat n.ctrl ess.treat ess.ctrl    max.es    mean.es     max.ks
unw            585    415  585.0000 415.0000 0.4843836 0.26101696 0.24228195
es.max.ATE     585    415  540.7177 352.3807 0.1057480 0.05608106 0.05289369
           max.ks.p    mean.ks iter
```

```
unw              NA 0.15423746   NA
es.max.ATE       NA 0.04086099  976
```

Further detail regarding the model at, e.g., the third time period is available using

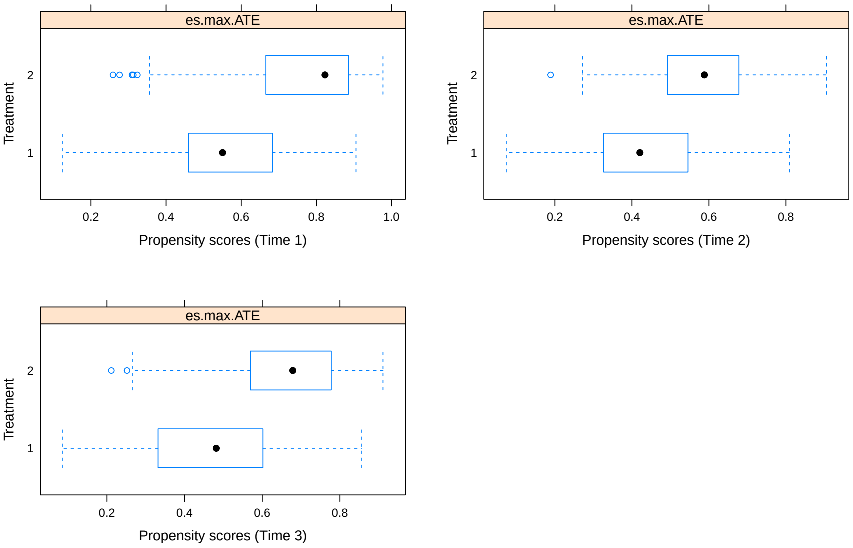
```
> bal.table(iptw.Ex, timePeriods = 3)

Balance at time  3 :
$unw
        tx.mn  tx.sd  ct.mn  ct.sd std.eff.sz  stat     p    ks ks.pval
use0    0.064  1.018 -0.129  1.062      0.186 2.888 0.004 0.135   0.000
gender  0.544  0.499  0.390  0.488      0.307 4.849 0.000 0.153   0.000
age    43.002 13.391 38.267 14.198      0.340 5.322 0.000 0.186   0.000
use1   -0.043  0.506 -0.129  0.528      0.166 2.582 0.010 0.140   0.000
tx1     0.750  0.433  0.643  0.480      0.235 3.621 0.000 0.107   0.008
use2   -0.141  0.506 -0.198  0.531      0.109 1.687 0.092 0.116   0.003
tx2     0.609  0.488  0.366  0.482      0.484 7.789 0.000 0.242   0.000

$es.max.ATE
        tx.mn  tx.sd  ct.mn  ct.sd std.eff.sz  stat     p    ks ks.pval
use0   -0.004  1.025 -0.041  1.021      0.036 0.539 0.590 0.041   0.864
gender  0.501  0.500  0.459  0.499      0.084 1.212 0.226 0.042   0.850
age    41.507 13.739 40.580 14.004      0.067 0.978 0.329 0.047   0.737
use1   -0.074  0.510 -0.090  0.509      0.031 0.473 0.637 0.039   0.895
tx1     0.723  0.448  0.700  0.459      0.051 0.764 0.445 0.023   1.000
use2   -0.163  0.510 -0.173  0.510      0.019 0.284 0.776 0.041   0.871
tx2     0.530  0.500  0.477  0.500      0.106 1.529 0.127 0.053   0.589
```

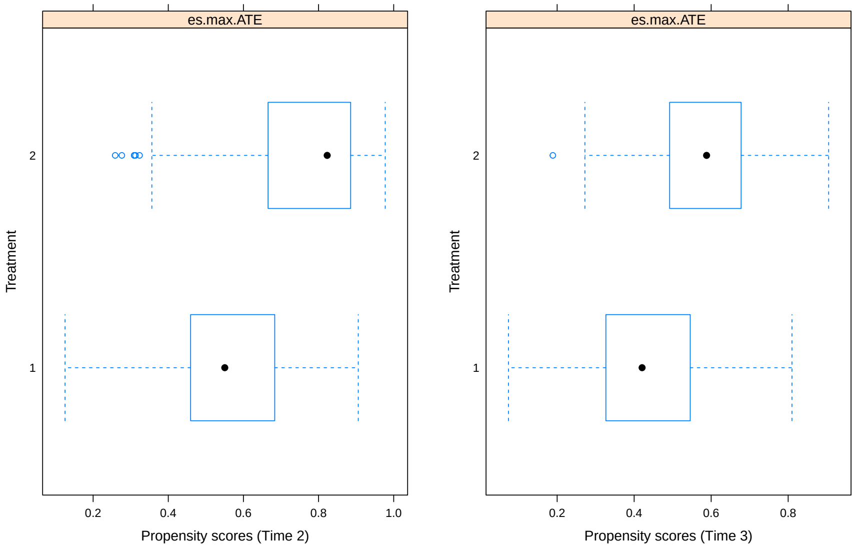
Next, we can examine propensity score overlap at each time point:

```
>      plot(iptw.Ex, plots = 2)
```



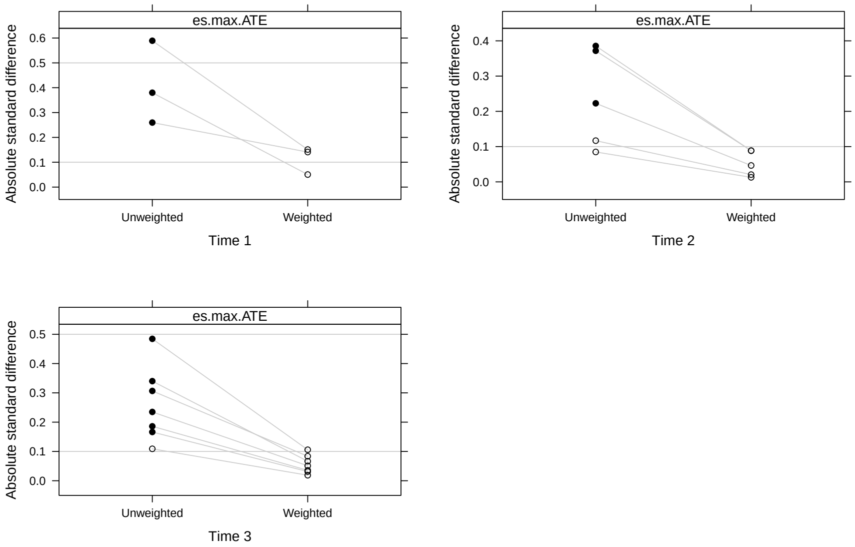
These figures can focus on the results from particular time periods using the `timePeriods` argument:

```
>     plot(iptw.Ex, plots = 2, timePeriods = 2:3)
```



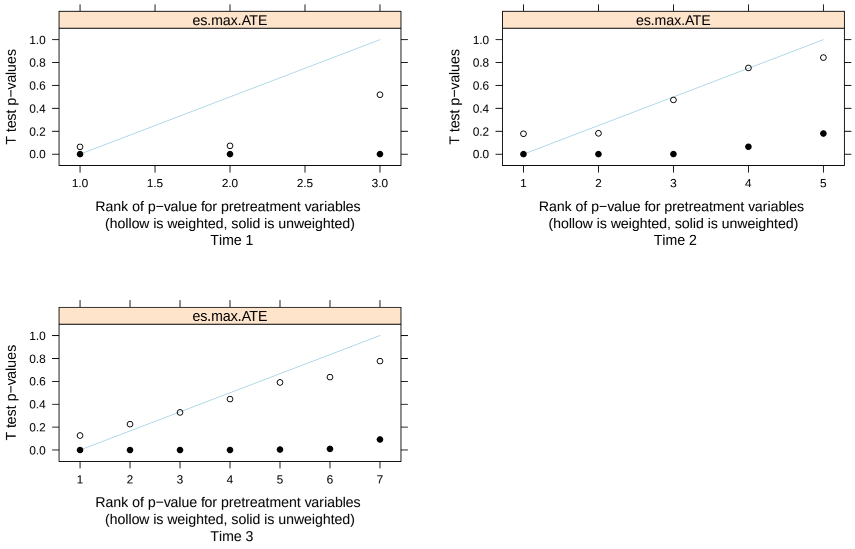
Next, we can check balance as measured by standardized mean differences between the treated and control samples at each of the time points by specifying `plots = 3`. As with other TWANG figures, we can specify `color = FALSE` to produce black and white figures.

```
>     plot(iptw.Ex, plots = 3, color = FALSE)
```

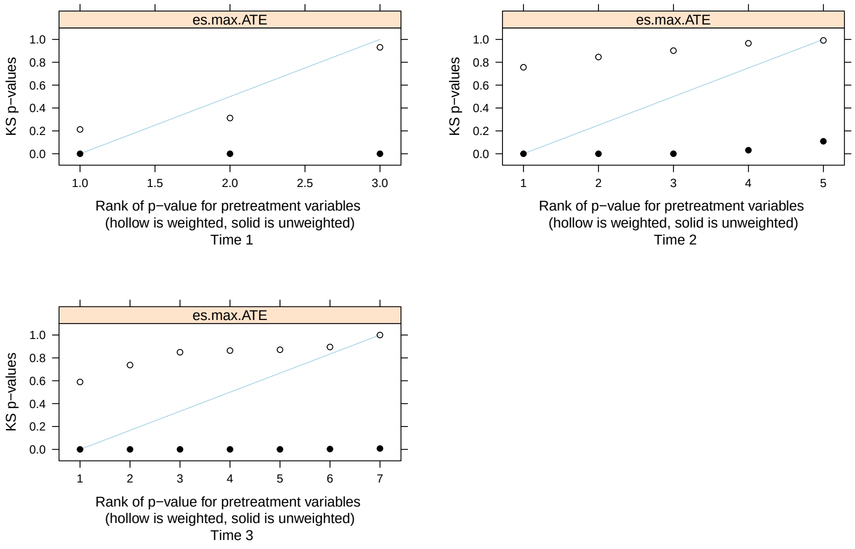


Finally, we can compare the differences between the treated and control samples using *t*-test and KS *p*-values by specifying `plots = 4` and `plots = 5`, respectively.

```
>     plot(iptw.Ex, plots = 4)
```



```
>     plot(iptw.Ex, plots = 5)
```



Further, `iptw` can accommodate treatments with more than two levels (McCaffrey et al., 2013). An example can be explored in the following call, though we do not discuss it further in this vignette. See the `mnps` vignette for more information on the diagnostic plots.

```
> data(mnIptwExWide)
> mniptw.Ex <- iptw(list(tx1 ~ use0, tx2 ~ use1, tx3 ~ use2),
+                   timeInvariant ~ gender + age,
+                   data = mnIptwExWide,
+                  cumulative = TRUE,
+                  priorTreatment = TRUE,
+                  verbose = FALSE,
+                  stop.method = "es.max",
+                  n.trees = 5000)
```

## 3 Estimating treatment effects

After having estimated the relevant propensity scores, the final step is translating them into analytic weights and estimating treatment effects. Twang provides several functions to facilitate this process. For this analysis, we assume an additive treatment model, where the mean change in outcomes depends on the number of periods of treatment. Because the weights often have substantial variation, the weights are commonly *stabilized* where the standard inverse probability of treatment weights are multiplied by the estimated probability of receiving the treatment that each individual received, conditioning only on previous periods' treatment indicators.

To begin, we calculate *unstablilized* weights. These are computed as the inverse probability of treatment weight, and are available as

```
> unstabWt1 <- get.weights.unstab(iptw.Ex)
```

We can estimate the treatment effect using these unstabilized weights as follows. the number of periods of treatment for each individual

```
> library(survey)
> nTx <- with(iptwExWide, tx1 + tx2 + tx3)
> outDatUnstab <- data.frame(outcome = iptwExWide$outcome,
+                        nTx,
+                        wt = unstabWt1$es.max.ATE)
> sv1unstab <- svydesign(~1, weights = ~wt, data = outDatUnstab)
```

We can then calculate the point estimate and 95% confidence interval using the unstabilized weights as

```
> fitUnstab <- svyglm(outcome ~ nTx, sv1unstab)
> coef(fitUnstab)

(Intercept)         nTx
 0.08322708 -0.12797134

> confint(fitUnstab)

                  2.5 %      97.5 %
(Intercept) -0.03424796  0.20070211
nTx         -0.18666525 -0.06927742
```

To calculate the stabilized weights, we additionally calculate a stabilizing factor that depends on on the marginal probabilities of treatment. This can be done via

```
> fitList <- list(glm(tx1 ~ 1, family = binomial, data = iptwExWide),
+                 glm(tx2 ~ tx1, family = binomial, data = iptwExWide),
+                 glm(tx3 ~ tx1 + tx2, family = binomial, data = iptwExWide))
> numWt <- get.weights.num(iptw.Ex, fitList)
> stabWt1 <- unstabWt1 * numWt
> outDatStab <- data.frame(outcome = iptwExWide$outcome,
+                          nTx,
+                          wt = stabWt1$es.max.ATE)
> sv1stab <- svydesign(~1, weights = ~wt, data = outDatStab)
```

As before, we can then estimate the treatment effect and associated confidence interval

```
> fitStab <- svyglm(outcome ~ nTx, sv1stab)
> coef(fitStab)

(Intercept)         nTx
 0.09501726 -0.13104433

> confint(fitStab)

                  2.5 %      97.5 %
(Intercept) -0.01962316  0.20965767
nTx         -0.18668719 -0.07540147
```

Since these are simulated data, we know that the true treatment effect is -0.1. We can see that both of the propensity score-weighted estimates cover the true treatment effect.

For comparison, we examine the unadjusted effect estimate, which we see does not include the true value:

```
> confint(lm(iptwExWide$outcome ~ nTx))

                  2.5 %       97.5 %
(Intercept) -0.12892466  0.058683047
nTx         -0.09333637 -0.001694947
```

## 4 Conclusion

Frequently, researchers are interested in treatments that may vary period-by-period. Twang's `iptw` function provides a nonparametric method for calculating inverse probability of treatment weights for marginal structural models. The function can accommodate treatments with two or more levels. The diagnostic figures and tables build on of the `mnps` and `ps` commands, with additional features to help manage the numerous possible comparisons.

## 5 References

Cole, S.R. and Hernán (2008). "Constructing inverse probability weights for marginal structural models." *American Journal of Epidemiology*, 168(6), 656-664.

Griffin, B.A., R. Ramchand, D. Almirall, M.E. Slaughter, L.F. Burgette, and D.F. McCaffrey (2014). "Estimating the causal effects of cumulative treatment episodes for adolescents using marginal structural models and inverse probability of treatment weighting. *Drug and Alcohol Dependence*, 136(1), 69–78.

McCaffrey, D.F., B.A. Griffin, D. Almirall, M.E. Slaughter, R. Ramchand, and L.F. Burgette (2013). "A tutorial on propensity score estimation for multiple treatments using generalized boosted models." *Statistics in Medicine*, 32(19): 3388-3414.

Robins, J.M., M.Á. Hernán, and B. Brumback (2000). "Marginal structural models and causal inference in epidemiology." *Epidemiology*, 11(5), 550–560.