Guidelines for Using State-of-the-Art Methods to Estimate Propensity Score and Inverse Probability of Treatment Weights When Drawing Causal Inferences

Session 7. Marginal Structural Models

Beth Ann Griffin
Dan McCaffrey

Marginal Structural Models

- For each of many treatment conditions $Z = j$, let $\mu_j = E(Y_j)$
- Model μ_j as a function of the treatment condition attributes
- For example, if conditions are no treatment (μ_{00}), treatment A only (μ_{10}), treatment B only (μ_{01}), and both treatments A and B (μ_{11}), i.e., a factorial design, then

$$\mu_{A,B} = \beta_0 + \beta_A A + \beta_B B + \beta_{AB} AB,$$

  where A and B are indicators for the conditioning including treatment A or B
  - β_A, β_B are main effects and β_{AB} is the interaction
- To estimate the model, we weight observation to estimate potential outcome means and fit the model to them
- Particularly useful with time varying treatments

Time Varying Treatment

- So far we considered only static treatment applied at one point in time and effects assessed sometime after treatment
- What if cases are followed longitudinally and treatment is time varying?
  - Students who do or do not participate in special reading intervention from year to year

Extending the Causal Model

- **Suppose we have K time points**
- **$T_k = 1$ if received treatment in period k, 0 otherwise**
- **Each case in the sample has a vector $\mathbf{T}_i = (T_{i1}, \ldots, T_{ik})'$ designating treatment**
- **For each of the 2^K possible values of $\mathbf{T}_i$ there is a potential outcome, $Y_{\mathbf{T}}$ at some point after last treatment interval**
- **Causal effects depend on relationships among these many potential outcomes**
  - $Y_{(1,1,\ldots,1)} - Y_{(0,0,\ldots,0)}$ **causal effect of treatment every period vs. no treatment**
  - $\{(Y_{(0,0,\ldots,0,1)} - Y_{(0,0,\ldots,0,0)}) + (Y_{(0,0,\ldots,1,1)} - Y_{(0,0,\ldots,0,1)}) + \ldots + (Y_{(0,1,\ldots,1,1)} - Y_{(0,0,\ldots,1,1)}) + (Y_{(1,1,\ldots,1,1)} - Y_{(0,1,\ldots,1,1)})\}/K$**, the average causal effect of one more period of treatment**

Extending the Causal Model (cont.)

- When K is even moderate, the possible number of causal effects becomes large and defining them may be difficult
- Reduce the dimensionality of the problem by developing a model for $Y_{\mathbf{T}}$ as a function of lower dimensional mapping of $\mathbf{T}$
  - E.g., the number of periods of treatment
- Use the parameters of this model to define causal effects of interest
- These are structural models because they describe the potential outcomes, not the observed outcomes
- Challenge is to estimate the structural model parameters with observed outcomes and treatment assignments

Heuristic on Estimating the Structural Model

- **Suppose the model is $E(Y_{\mathbf{T}}) = \alpha + \beta g(\mathbf{T})$, where $g(\mathbf{T})$ is a scalar function of T**
- **If we observed every case at all values of T we could model the relationship directly**
- **We can't make this observation, but suppose we knew how to weight cases with $\mathrm{T} = \mathrm{t}$ so the weighted average was an unbiased estimate of $E(Y_{\mathbf{t}})$ for all values of t**
- **We could then model these weighted averages or model all the weighted cases to obtain the structural model parameters**

Assumptions for Estimating Marginal Structural Models

- We need to extend the notion of strong ignorability
- Consistent estimation requires no unmeasured treatment confounders
  - There are no unmeasured risk factors for the potential outcomes that directly affect treatment assignment at *any* time point
- Observed risk factors can affect treatment assignment and they *can be affected by earlier treatment assignment*
- Unobserved variables can affect observed risk factors and they can be affected by prior treatment and unobserved risk factors but they *cannot affect treatment*

The Assumptions Are Less Stringent Than Those Required for Regression Modeling

- Regression modeling would include functions of treatment history and risk factor history as variables in the model
- If observed risk factors are affected by prior treatment assignments, then the model yields biased estimates of the parameters for treatment history
  - Putting variables on the causal pathway into the model
- For regression to work we would need to include only observed risk factors that were not affected by prior treatment

Weighting for Estimating Marginal Structural Models

- **Assume no unobserved confounding**
- **Weighting by the inverse probability of receiving observed treatment history given observed risk factors yields unbiased estimates**
- $p = \prod_{k=1}^{K} P(T_k = t_k | T_0, \ldots, T_{k-1}, x_1, \ldots, x_k)$
- **Weighting by** $1/p$ **would yield unbiased estimates**
- p **can be estimated by a series of models for dichotomous treatment indicators**
- **These weights can be unstable**

Stabilizing the Weighting for Estimating Marginal Structural Models

- **Stabilized weights can also yield unbiased estimates but not vary so much**
- **Let** $p_m = \prod_{k=1}^{K} P(T_k = t_k | T_0, \ldots, T_{k-1})$
- $w = p_m / p$
- **Notion is that we don't need to weight up rare treatment patterns (small** p_m**) because our goal is to compare across patterns, not combine them, but we do need to account for differential probabilities among people with each pattern**
- **Alternatively can weight by** $p'_m = P(g(\mathbf{T}) = g(\mathbf{t}))$

MSM and Time Varying Outcomes

- We can use MSM approach to model repeated outcomes
- For instance if modeling the effect of an additional quarter of a reading intervention
- We have repeated measures Y_{ik}, functions of treatment status up to time k, $g(\mathrm{T}_{ik})$, and weights w_{ik} for each individual i and time period $k = 1, \ldots, K$
- We can the model Y_{ik} by the appropriate function of $g(\mathrm{T}_{ik})$ and weight observations by w_{ik}
- Because we have repeated outcome measures on the same individuals, we need to adjust the standard error for weighting and clustering of observations within individual

Summary Marginal Structural Models

- Extends causal modeling paradigm to time-varying treatment effects
- Establishes structural models for potential outcomes on the basis of reductions of the multivariate treatment assignment vector
- Using stabilized weighting to estimate the structural model parameters
- Unbiased estimation requires assumptions that the structural models are correct and there are no unobserved confounders
- Can control for risk factors that can be affected by prior treatment without introducing bias whereas regression models cannot

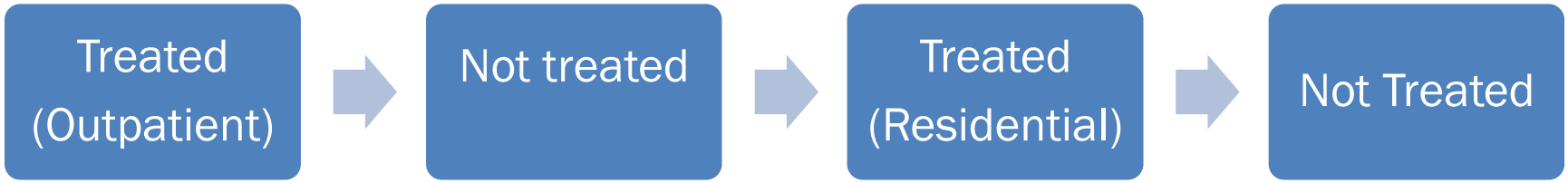
Motivating Case Study

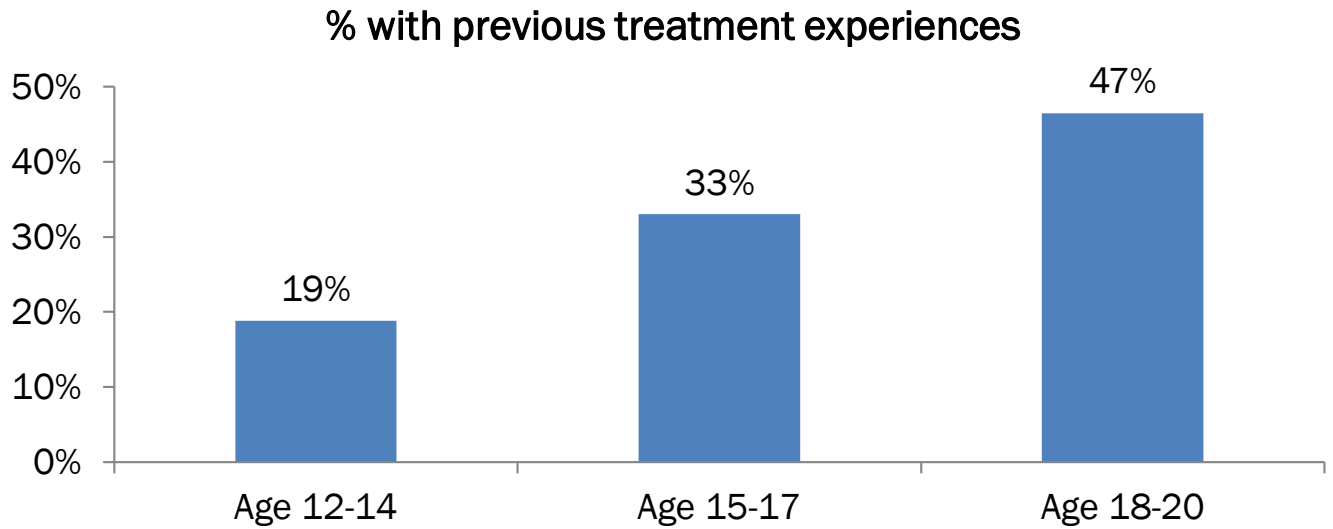
- Estimate the average effect of each additional treatment episode with:

over 9-month period on outcomes at one-year post-intake

Treatment is Rarely A One-Time Event

One Hypothetical Treatment Trajectory





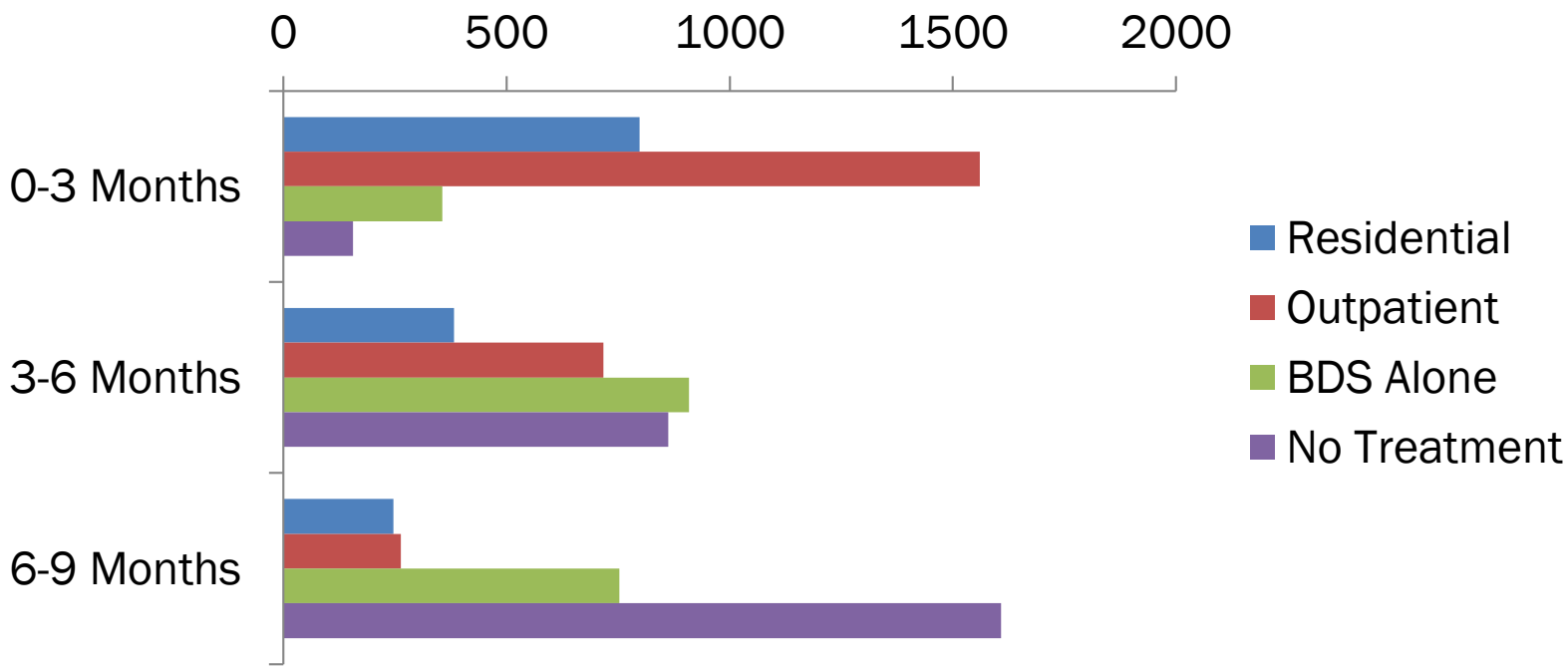
Based on 245,645 treatment admissions in 2010 (SAMHDA)

Case Study Data

- N = 2,870 adolescents who received treatment at CSAT-funded program sites
- Follow-ups at baseline, 3, 6, 9, and 12 months
- Patient characteristics and outcomes assessed using Global Appraisal of Individual Needs (GAIN)
- Multiply imputed data

Treatment assignment between follow-ups

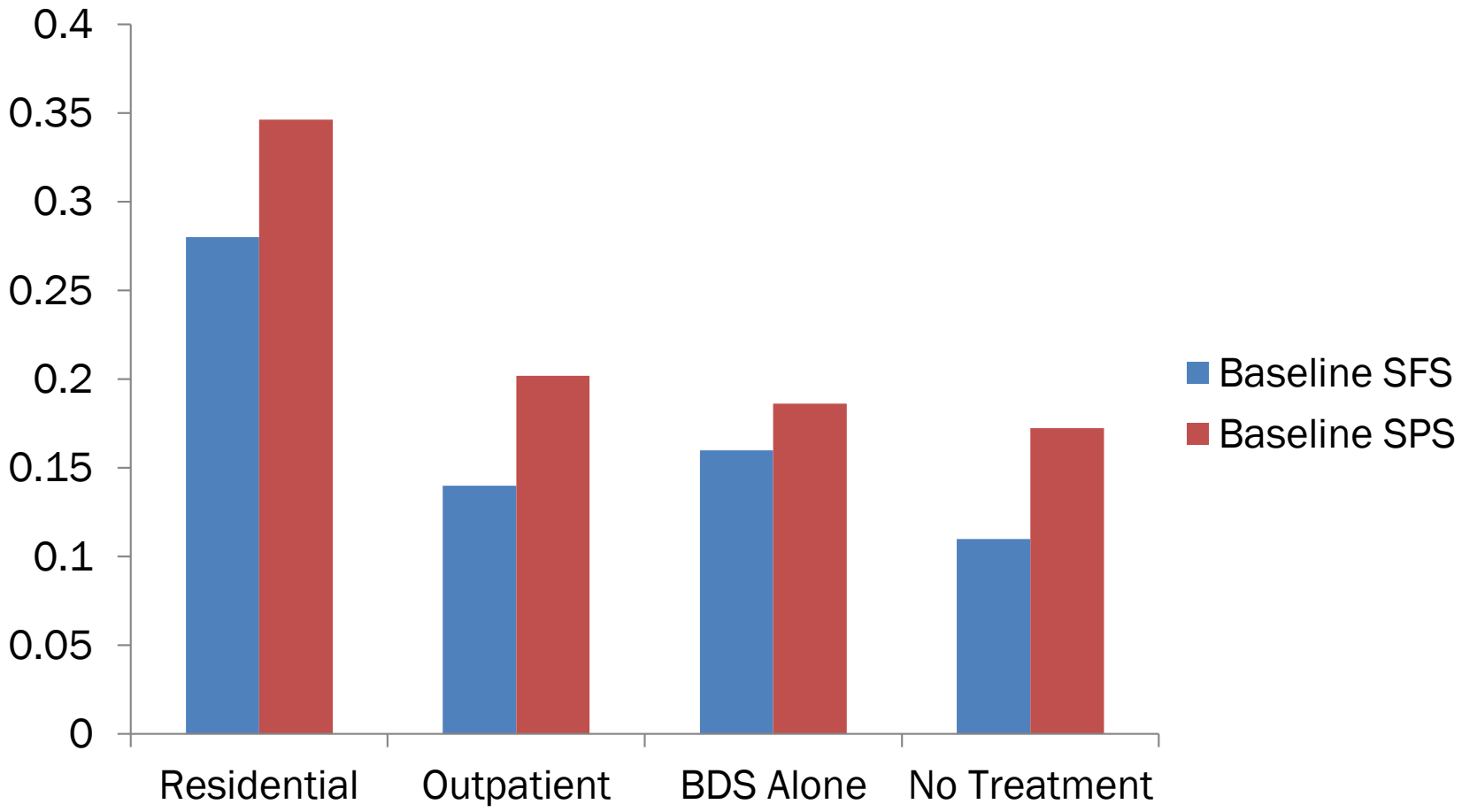
- Treatment = 4-level categorical variable
  - Time-varying and defined at the end of each 3 month interval



Outcomes – Measured at 12-months

Substance Frequency Scale (SFS)	• 8-items • average proportion of alcohol and other drugs using days in the past 90
Substance Problem Scale (SPS)	• 16 items • count possible symptoms in the past 30 days related to drug dependence and substance abuse

Biggest challenge in estimating our outcome models: Addressing confounding



Not only have baseline differences but also differences at each time point

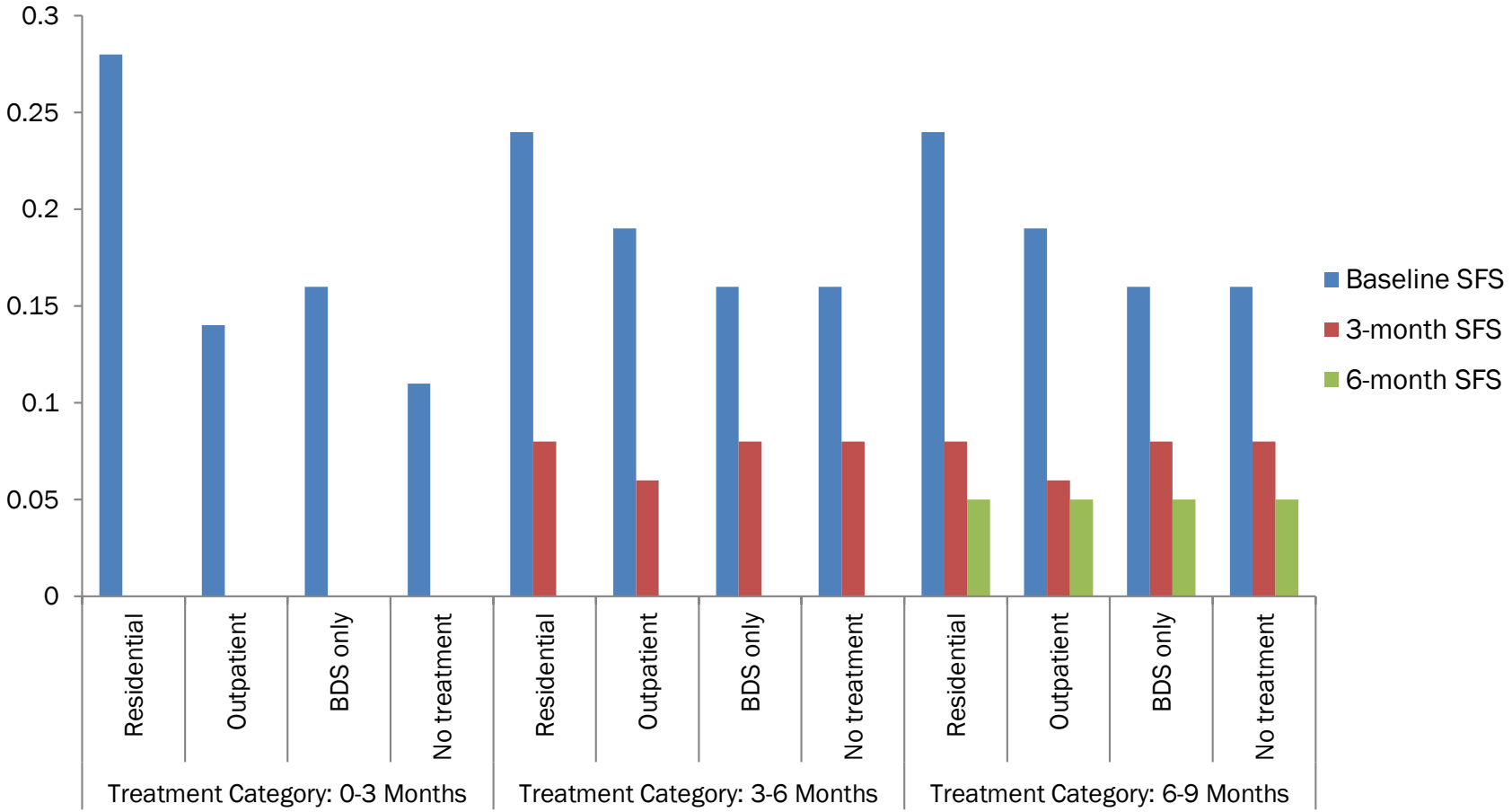
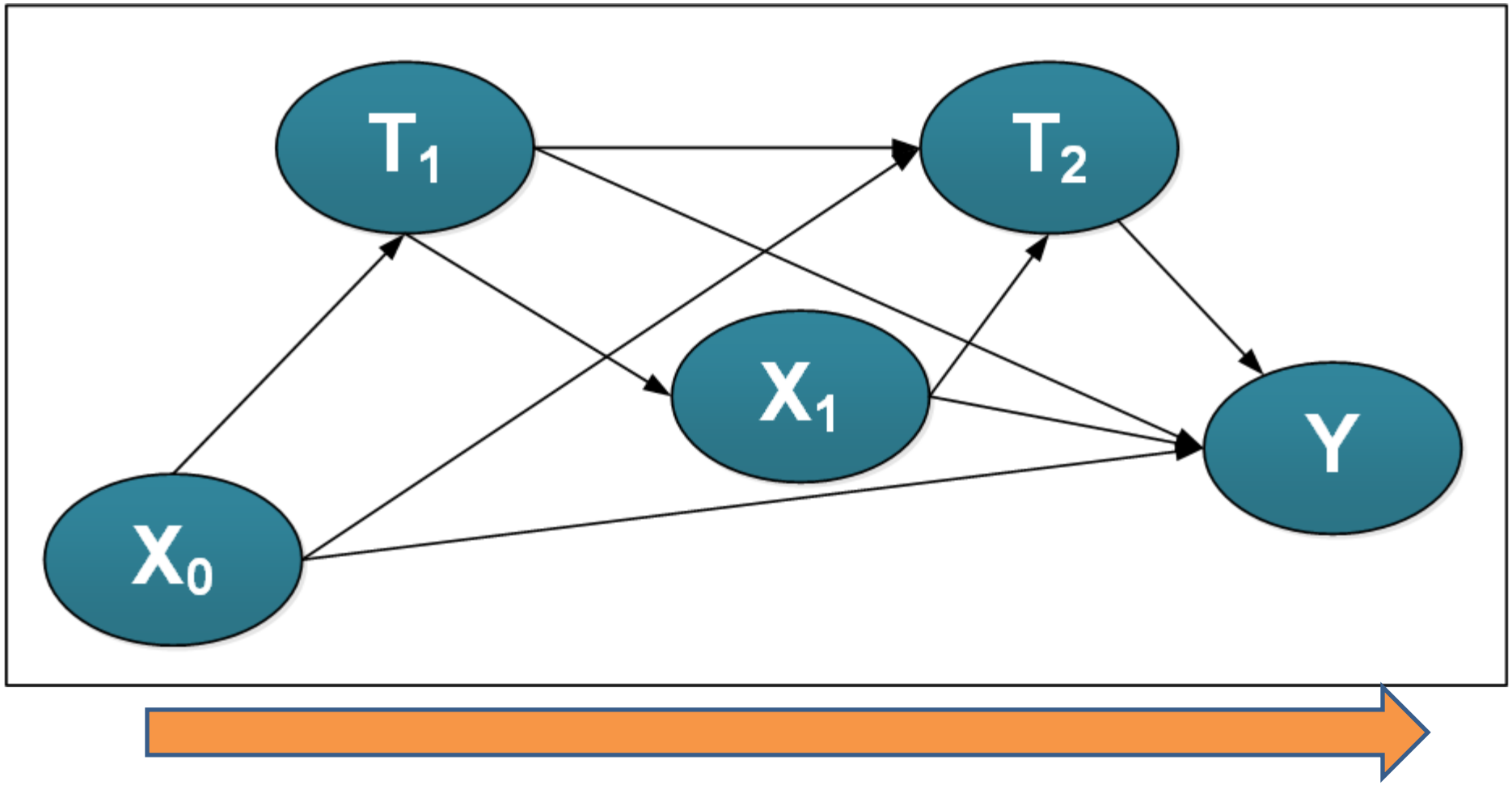
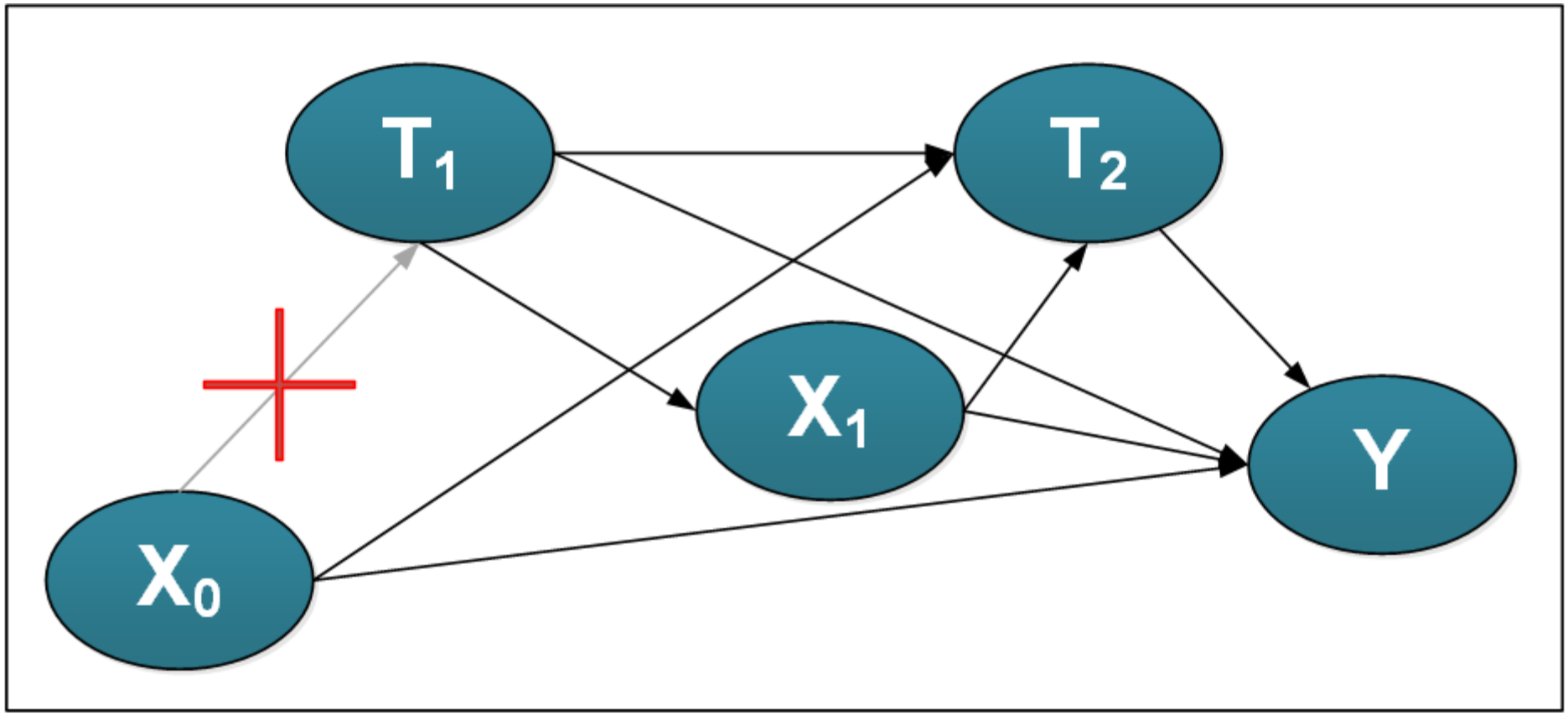


Illustration of time-varying confounding

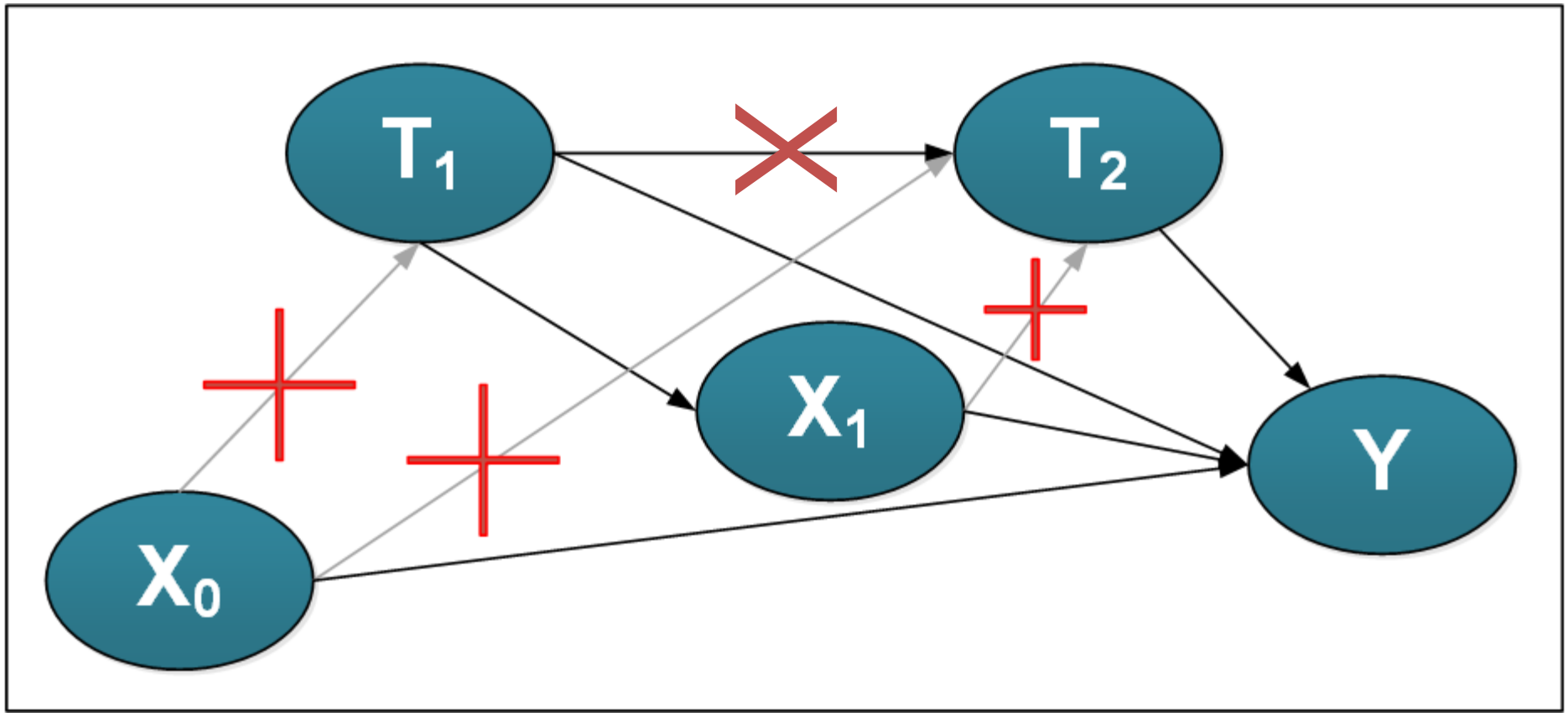


If we randomized individuals to levels of T_1



Time

If we were able to randomize individuals to levels of T_2



Time

Defining our MSM

- To estimate average effect of additional periods of each treatment on 12-month outcomes, we use

$$E[Y_{\bar{T}_{i,12}}] = a_0 + a_{OP} * cum(\overline{OP_i}) + a_{RES} * cum(\overline{RES_i}) + a_{BDS} * cum(\overline{BDS_i})$$

- MSM shows how we hypothesize population means of each treatment trajectory to vary as a function of the covariates in the model
  - We don't have the sample size to fit fully saturated model in our data
  - Reduced to a simpler function that assumes a constant effect of each additional episode

Four key steps

1) Define the primary treatment effect of interest via a MSM
2) Estimate IPTW
3) Evaluate the quality of the IPTW
4) Estimate the treatment effect

Step 1: Define the primary treatment effect of interest via a MSM

- As noted before, we focus on estimating the average effect of each additional period of a given treatment on 12-month outcomes
- To simplify presentation, only going to use binary treatment in next few slides
- So we will use the following MSM

$$E[Y_{\bar{T}_{i,12}}] = a_0 + a_{Tx} * cum(\overline{Tx}_i)$$

Step 2: Estimate the IPTW

- Recommend using iptw() in twang
- When designing data, can use either “long” or “wide” format
  - Wide format has a separate column for each time-varying characteristic at each time
  - Long format has a single column for each time-varying characteristic

Example data – wide format

```
> head(wideDat)
  gender age        use0        use1       use2 tx1 tx2 tx3
1      0  43  1.13496509  0.467482544  0.3174825   1   1   1
2      0  50  1.11193185  0.455965923  0.4059659   1   0   1
3      1  36 -0.87077763 -0.535388817 -0.5853888   1   0   0
4      1  63  0.21073159  0.005365793 -0.1446342   1   1   1
5      0  24  0.06939565 -0.065302176 -0.1153022   1   0   1
6      1  20 -1.66264885 -0.931324426 -1.0813244   1   1   1
```

Example of iptw code: “wide” format

```
tv1 <- iptw(list(tx1 ~ use0,
                 tx2 ~ use1,
                 tx3 ~ use2),
            data = wideDat,
            timeInvariant = ~ gender + age,
            n.trees = 5000,
            cumulative = TRUE,
            priorTreatment = TRUE,
            stop.method = “es.max”)
```

Note: List of equations should be ordered from earliest to latest treatments

Example of iptw code: “wide” format

```
tv1 <- iptw(list(tx1 ~ use0,
                 tx2 ~ use1 + use0 + tx1,
                 tx3 ~ use2 + use1 + use0 + tx2 + tx1),
            data = wideTvEx,
            timeInvariant = ~ gender + age,
            n.trees = 5000,
            cumulative = TRUE,
            priorTreatment = TRUE),
            stop.method = “es.max”)
```

Note: List of equations should be ordered from earliest to latest treatments

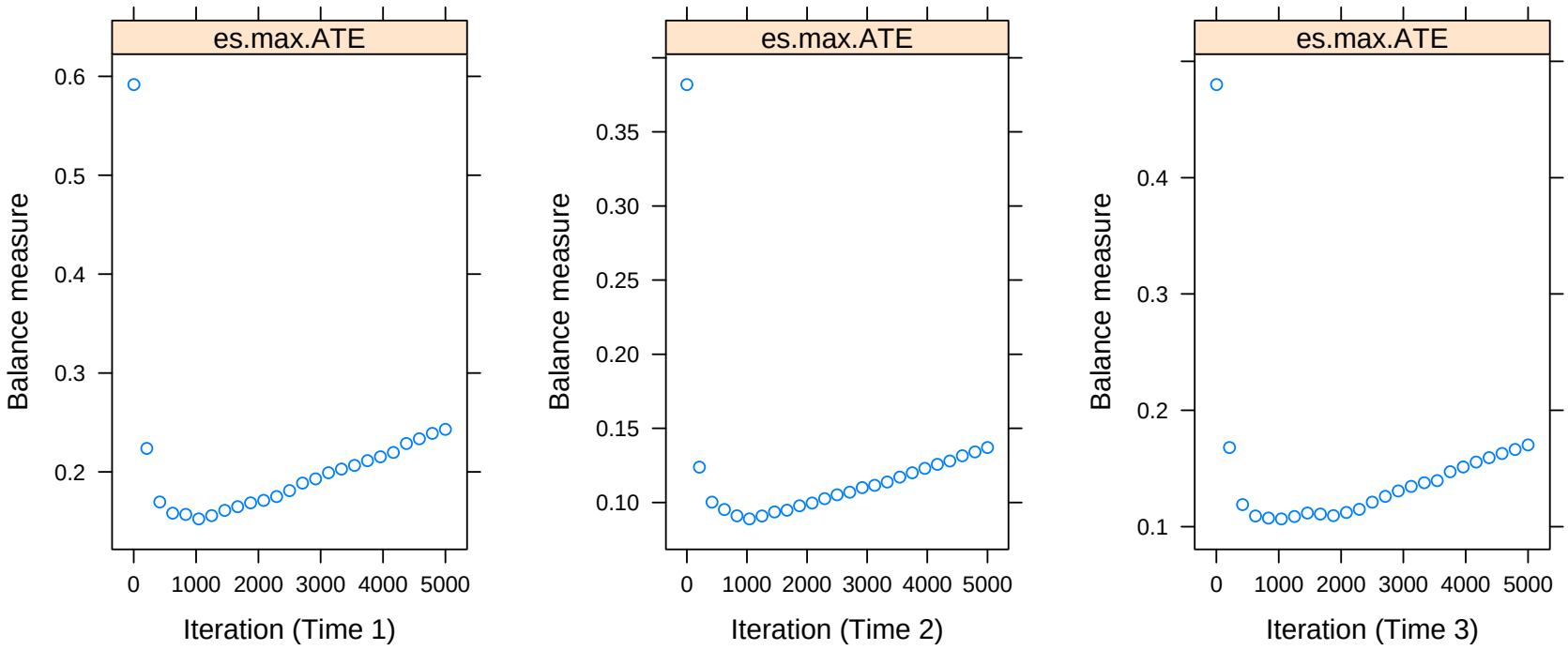
Step 3: Evaluate the quality of the IPTW

- Recommend assessing balance at each time point
- Use standard metric for assessing balance with propensity score weights
  - For more then two groups, we recommend examining all pairwise standardized mean difference (SMD) and ks statistics

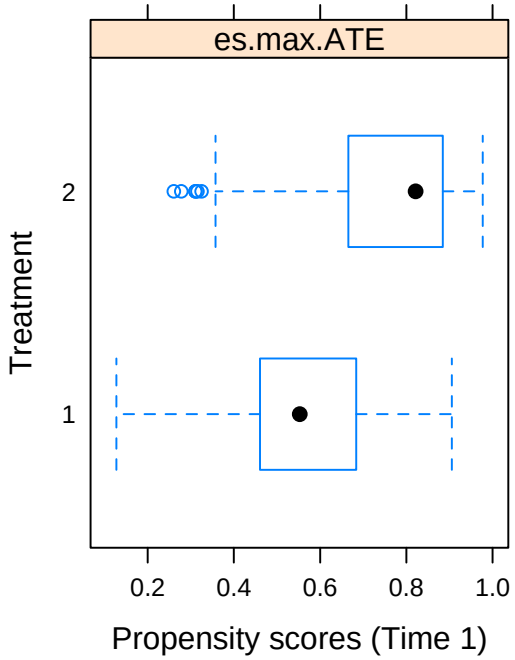
Summary (detail)

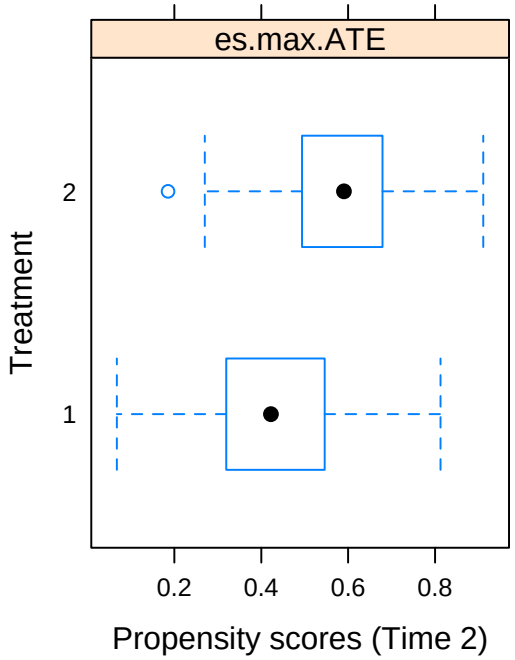
```
> summary(tv1)
Summary for time period  1 :
            n.treat n.ctrl ess.treat ess.ctrl    max.es
unw             706    294  706.0000 294.0000 0.5891037
es.max.ATE      706    294  656.0565 216.4759 0.1519663
Summary for time period  2 :
            n.treat n.ctrl ess.treat ess.ctrl     max.es
unw             508    492  508.0000  492.000 0.38549444
es.max.ATE      508    492  475.1469  449.064 0.08884127
Summary for time period  3 :
            n.treat n.ctrl ess.treat ess.ctrl    max.es
unw             585    415  585.0000 415.0000 0.4843836
es.max.ATE      585    415  541.1826 353.0498 0.1058122
```

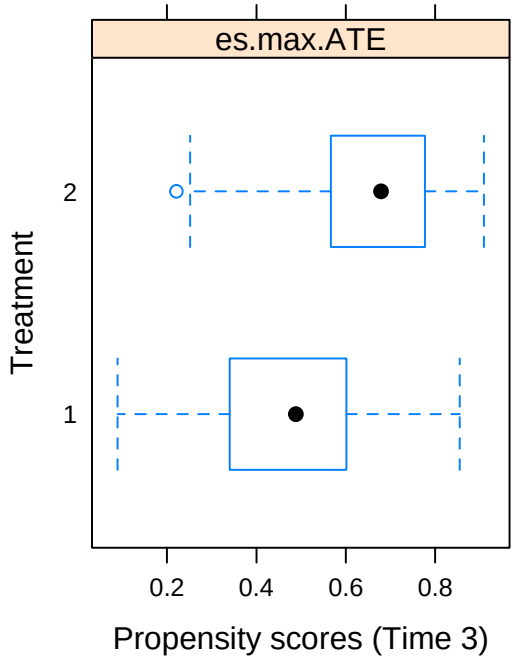
plots = 1


es.max.ATE
Balance measure
0.6
0.5
0.4
0.3
0.2
0
1000
2000
3000
4000
5000
Iteration (Time 1)
es.max.ATE
Balance measure
0.35
0.30
0.25
0.20
0.15
0.10
0
1000
2000
3000
4000
5000
Iteration (Time 2)
es.max.ATE
Balance measure
0.4
0.3
0.2
0.1
0
1000
2000
3000
4000
5000
Iteration (Time 3)

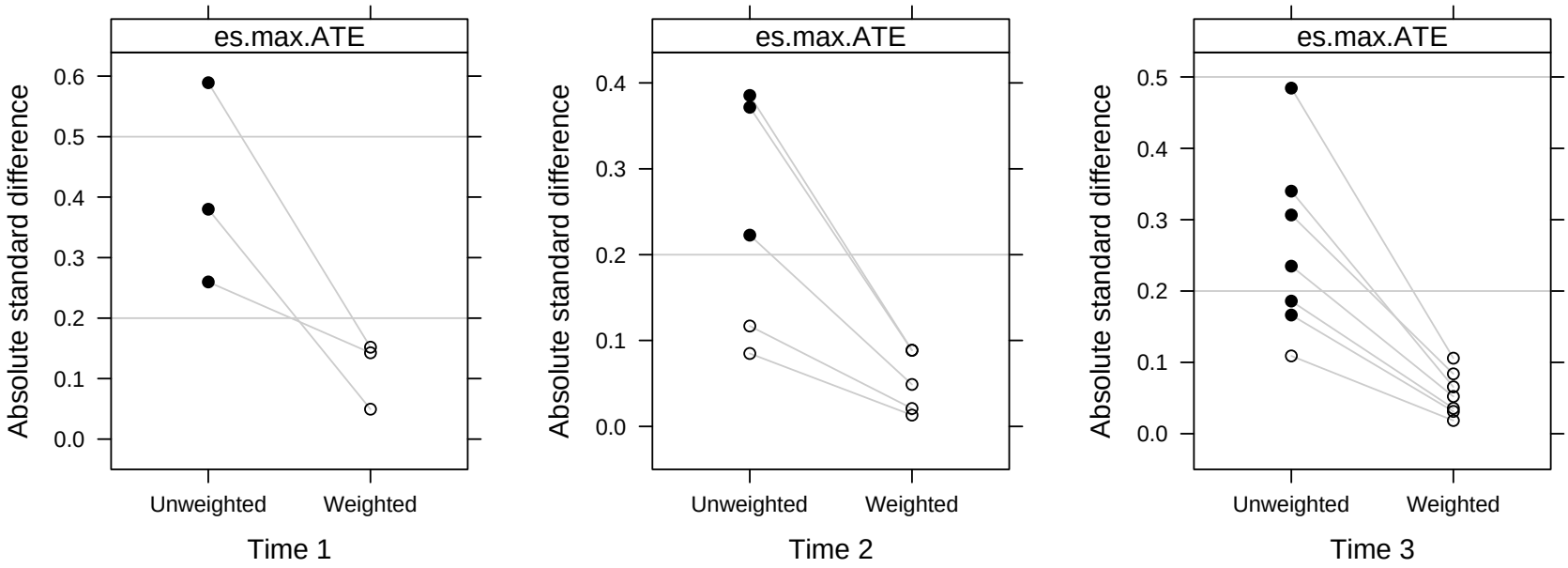
plots = 2


es.max.ATE
Treatment
2
1
0.2
0.4
0.6
0.8
1.0
Propensity scores (Time 1)


es.max.ATE
Treatment
2
1
0.2
0.4
0.6
0.8
Propensity scores (Time 2)


es.max.ATE
Treatment
2
1
0.2
0.4
0.6
0.8
Propensity scores (Time 3)

plots = 3, color = FALSE



bal.table() detail

```
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   0.007
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    -0.005  1.026 -0.041  1.021      0.036 0.531 0.596 0.041   0.840
gender   0.501  0.500  0.459  0.499      0.084 1.217 0.224 0.042   0.826
age     41.504 13.738 40.594 14.006      0.066 0.961 0.337 0.047   0.717
use1    -0.075  0.510 -0.090  0.509      0.031 0.463 0.643 0.040   0.873
tx1      0.723  0.448  0.700  0.459      0.052 0.778 0.437 0.024   0.999
use2    -0.164  0.510 -0.173  0.511      0.018 0.274 0.784 0.041   0.847
tx2      0.530  0.500  0.477  0.500      0.106 1.531 0.126 0.053   0.564
```

Note: We haven't even begun to talk about the outcome yet

- Steps 1 to 3 do not involve any outcomes
- We first focus on dealing with selection (pre-treatment and time-varying) group differences
- Then, if we do a good job, we will move to outcome analyses

Step 4: Estimate the treatment effect

- Estimate the regression model defined in the MSM using commands from your favorite survey package
- Fitting a weighted regression model
  - Since we are using weights, we need to adjust our standard errors for the weighting
  - Analogous to fitting regression models with survey data with survey weights

Getting weights

- For iptw, twang produces unstabilized weights
  - Use get.weights.unstab()
- Stabilization depends on specification of the marginal outcome model
  - Can use get.weights.num() to obtain stabilization factor for iptw (e.g, numerator of $SW_i(t)$)

$$SW_i(t) = \frac{\prod_{j=1}^{t} P_j(T_{ij} = k_{ij} | \bar{T}_{i,j-1})}{\prod_{j=1}^{t} P_j(T_{ij} = k_{ij} | \bar{T}_{i,j-1}, \bar{L}_{ij})}$$

Getting weights (cont.)

Code with stabilizing weights (using glm)

```
ftList <- list(glm(tx1 ~ 1, family = binomial),
                        glm(tx2 ~ tx1, family = binomial),
                  glm(tx3 ~ tx1 + tx2, family = binomial))
numWt1 <- get.weights.num(tv1, ftList)
unstabWt1 <- get.weights.unstab(tv1)
stabWt1 <- numWt * unstabWt1
outDt <- data.frame(outcome, nTx, wt = stabWt1$es.max.ATE)
sv1stab <- svydesign(~1, weights = ~wt, data = outDt)
```

Getting weights (cont.)

Code with stabilizing weights (using glm)

```
ftList <- list(glm(tx1 ~ 1, family = binomial),
               glm(tx2 ~ tx1, family = binomial),
               glm(tx3 ~ tx1 + tx2, family = binomial))
numWt1 <- get.weights.num(tv1, ftList)
unstabWt1 <- get.weights.unstab(tv1)
stabWt1 <- numWt * unstabWt1
outDt <- data.frame(outcome, nTx, wt = stabWt1$es.max.ATE)
sv1stab <- svydesign(~1, weights = ~wt, data = outDt)
```

Fits logistic regression models for estimating $\boldsymbol{P_j(T_{ij} = k_{ij}|\overline{T}_{i,j-1})}$

Getting weights (cont.)

Code with stabilizing weights (using glm)

```
ftList <- list(glm(tx1 ~ 1, family = binomial),
                        glm(tx2 ~ tx1, family = binomial),
                  glm(tx3 ~ tx1 + tx2, family = binomial))
numWt1 <- get.weights.num(tv1, ftList)
unstabWt1 <- get.weights.unstab(tv1)
stabWt1 <- numWt * unstabWt1
outDt <- data.frame(outcome, nTx, wt = stabWt1$es.max.ATE)
sv1stab <- svydesign(~1, weights = ~wt, data = outDt)
```


Gets numerator value for each weight

Getting weights (cont.)

Code with stabilizing weights (using glm)

```
ftList <- list(glm(tx1 ~ 1, family = binomial),
                        glm(tx2 ~ tx1, family = binomial),
                  glm(tx3 ~ tx1 + tx2, family = binomial))
numWt1 <- get.weights.num(tv1, ftList)
unstabWt1 <- get.weights.unstab(tv1)
stabWt1 <- numWt * unstabWt1
outDt <- data.frame(outcome, nTx, wt = stabWt1$es.m
sv1stab <- svydesign(~1, weights = ~wt, data = outDt)
```

Gets denominator value for each weight

Getting weights (cont.)

Code with stabilizing weights (using glm)

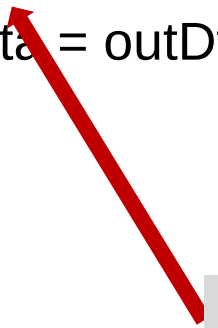
```
ftList <- list(glm(tx1 ~ 1, family = binomial),
                        glm(tx2 ~ tx1, family = binomial),
                  glm(tx3 ~ tx1 + tx2, family = binomial))
numWt1 <- get.weights.num(tv1, ftList)
unstabWt1 <- get.weights.unstab(tv1)
stabWt1 <- numWt * unstabWt1
outDt <- data.frame(outcome, nTx, wt = stabWt1$es.max.ATE)
sv1stab <- svydesign(~1, weights = ~wt, data = outDt)
```

Combines the numerator and denominator

Getting weights (cont.)

Code with stabilizing weights (using glm)

```
ftList <- list(glm(tx1 ~ 1, family = binomial),
                        glm(tx2 ~ tx1, family = binomial),
                  glm(tx3 ~ tx1 + tx2, family = binomial))
numWt1 <- get.weights.num(tv1, ftList)
unstabWt1 <- get.weights.unstab(tv1)
stabWt1 <- numWt * unstabWt1
outDt <- data.frame(outcome, nTx, wt = stabWt1$es.max.ATE)
sv1stab <- svydesign(~1, weights = ~wt, data = outDt)
```

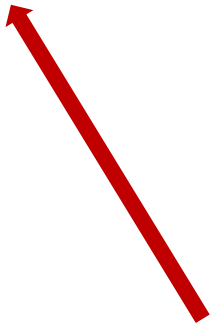


Attaches weights to dataset

Getting weights (cont.)

Code with stabilizing weights (using glm)

```
ftList <- list(glm(tx1 ~ 1, family = binomial),
                        glm(tx2 ~ tx1, family = binomial),
                  glm(tx3 ~ tx1 + tx2, family = binomial))
numWt1 <- get.weights.num(tv1, ftList)
unstabWt1 <- get.weights.unstab(tv1)
stabWt1 <- numWt * unstabWt1
outDt <- data.frame(outcome, nTx, wt = stabWt1$es.max.ATE)
sv1stab <- svydesign(~1, weights = ~wt, data = outDt)
```



Creates weighted survey design for outcome models

Results (True treatment effect = -0.1)

Unadjusted Model Code and Output

```
> confint(lm(outcome~nTx))
                  2.5 %       97.5 %
(Intercept) -0.12892466  0.058683047
nTx         -0.09333637 -0.001694947
```

Model with stabilized weights Code and Output

```
> confint(svyglm(outcome ~ nTx, sv1stab))
                 2.5 %      97.5 %
(Intercept) -0.0195951  0.21011847
nTx         -0.1864824 -0.07504334
```

Results (True treatment effect = -0.1)

Unadjusted Model Code and Output

```
> confint(lm(outcome~nTx))
                  2.5 %       97.5 %
(Intercept) -0.12892466  0.058683047
nTx         -0.09333637 -0.001694947
```

Model with stabilized weights Code and Output

```
> confint(svyglm(outcome ~ nTx, sv1stab))
                 2.5 %      97.5 %
(Intercept) -0.0195951  0.21011847
nTx         -0.1864824 -0.07504334
```

95% CI for unweighted model doesn't contain true treatment effect of -0.1 because there is confounding from pretreatment and time-varying confounders

Results (True treatment effect = -0.1)

Unadjusted Model Code and Output

```
> confint(lm(outcome~nTx))
                  2.5 %       97.5 %
(Intercept) -0.12892466  0.058683047
nTx         -0.09333637 -0.001694947
```

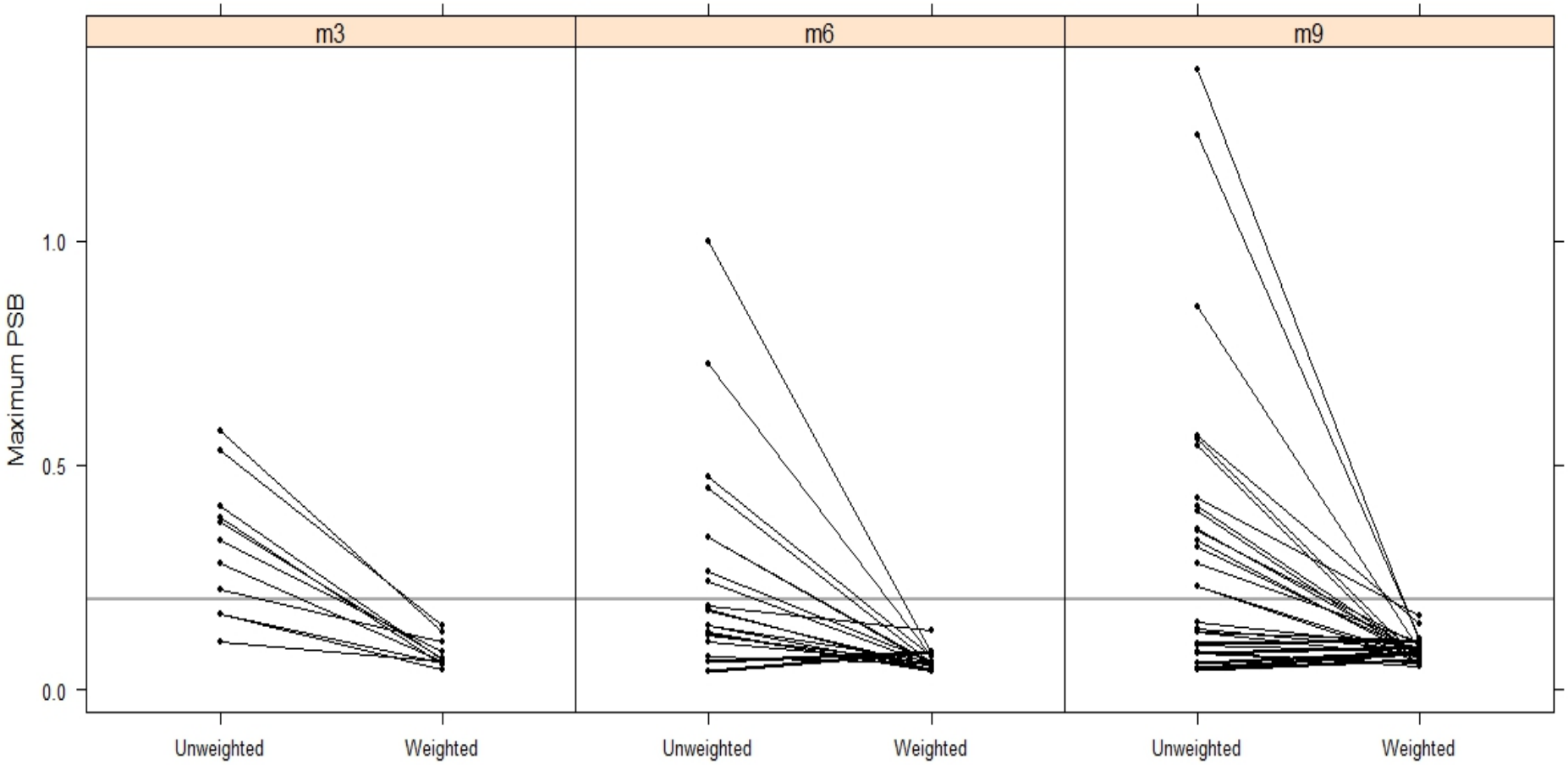
Model with stabilized weights Code and Output

```
> confint(svyglm(outcome ~ nTx, sv1stab))
                 2.5 %      97.5 %
(Intercept) -0.0195951  0.21011847
nTx         -0.1864824 -0.07504334
```

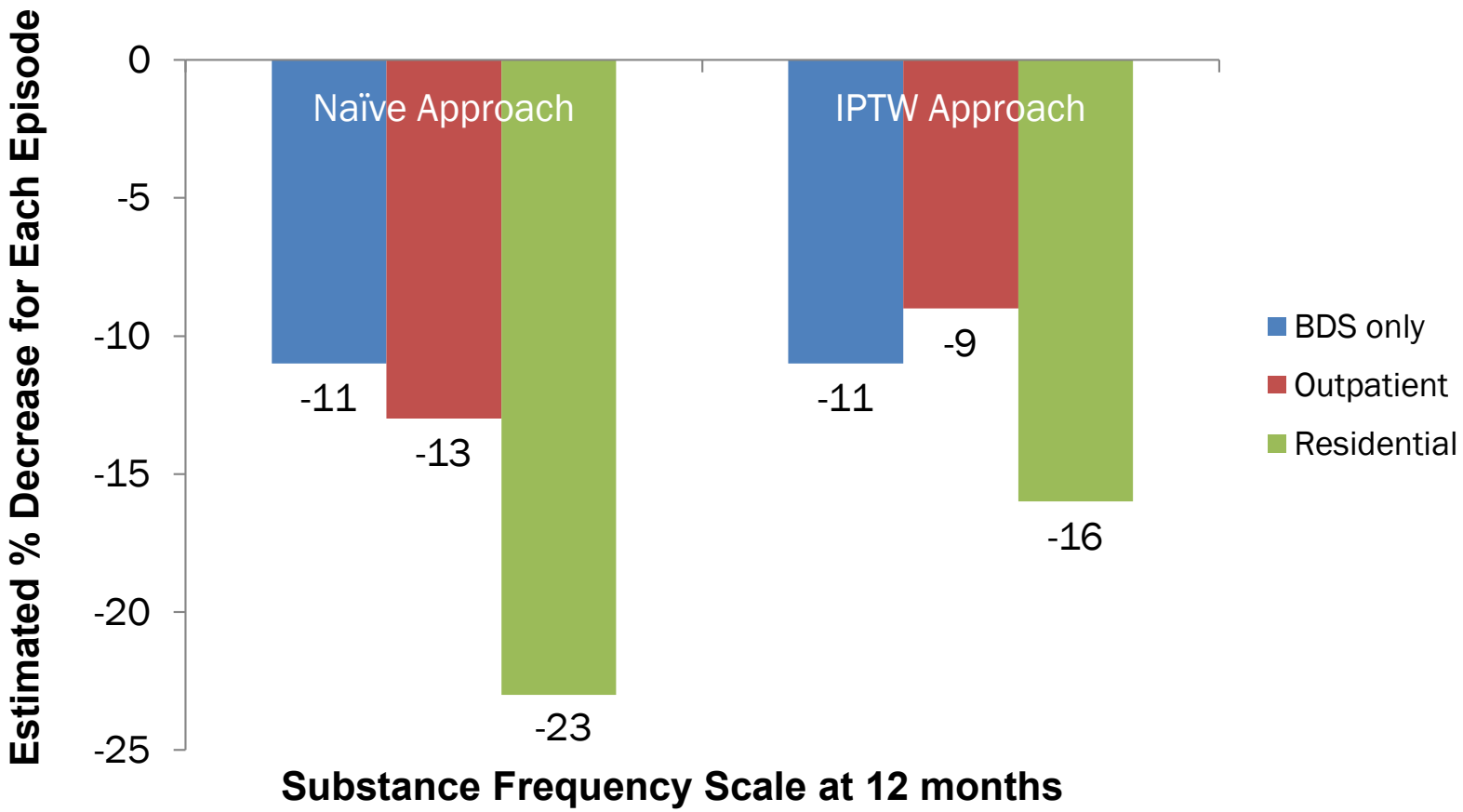
95% CI for weighted model does contain true treatment effect of -0.1; weights have removed confounding effect from pretreatment and time-varying confounders

RESULTS FROM CASE STUDY

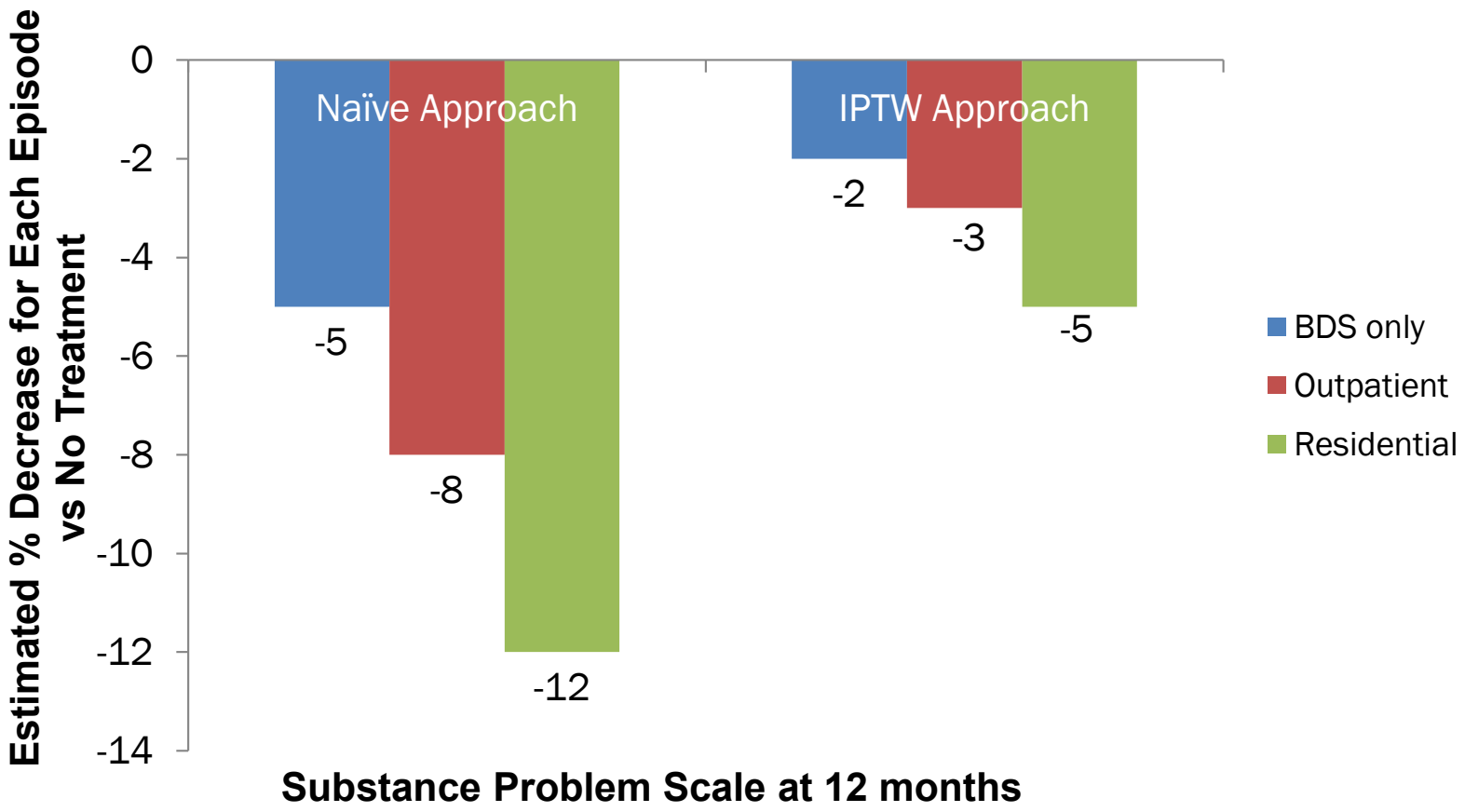
Assessing quality of our IPTW



Results: Significant reductions in use for each additional period of treatment



Results: No impact on substance problems



Conclusions from motivating study (Griffin et al 2014)

- Residential treatment, outpatient treatment, and biological drug screening each have effects that cumulate over time
  - On average, each additional period reduces use by about 10%
- Support adopting "life-course perspective" and re-thinking how we define "effective" programs
- Limitations: unobserved confounding; coarse measures of treatment; MSM too simple; generalizability

Final remarks

- Use of MSM+IPTW for time-varying treatments is a robust and meaningful way to draw causal inference
- Recommend estimating IPTW using machine learning or other state of the art methods rather than relying on parametric approaches
- Assessing balance is a critical step in understanding how robust inferences will be
- Ideally also do a sensitivity analysis to unobserved confounding