INVERSE PROPENSITY SCORE-BASED WEIGHTING METHODS TO REMOVE BIAS IN OBSERVATIONAL STUDIES

SHANKAR S SRINIVASAN, PH.D.

DIRECTOR, MEDICAL STATISTICS, BAYER

RESTRICTED

2 DISCLOSURES

- This presentation does not present any Bayer data.
- The results presented are based on simulated data.
- Any interpretations, recommendations and opinions presented are those of the presenter not those of the Bayer corporation.
- A portion of the slides were presented at the BASS 2019 meeting by the presenter. This version extends and adds material

RESTRICTED

3 OUTLINE

- Observations on Observational Studies (OS) and Intentional Design
- Key Propensity Score Theorems (Two Group Case)
  - Balancing Scores - ATE and ATT Weights
- Propensity Scores for >2 Groups
  - Addressing analytical Issues in curated (non-random) sampling in OS
- The Sneetch Simulation Example – Treatment Bias and Outcome parameters
  - Results – Balance and Bias/Variance in Estimation
- Conclusions

RESTRICTED

4 OBSERVATIONAL DATA - ISSUES AND POSITIVES

- Inherent differences between patient cohorts as they are channeled to therapies based on the patient's profile and the patient/physician interaction.
- Without a pre-specified pathway through the data there can be publication bias due to unpublished results and inflation in false positive rates.
- Missing data requiring imputation methods especially in retrospective data
- Positives to real world data include
  - very little filtering out of patients though inclusion/exclusion criteria and
  - with retrospective data there are no biases driven by known hypotheses.

RESTRICTED

5 CLINICAL TRIALS VERSUS OBSERVATIONAL DATA

CLINICAL TRIALS

- Usually very controlled in order to reduce noise and detect differences effectively. Project management intensive.
- Specified Interventional Agents
- Limited contexts such as those at diagnosis, maintenance and after relapse.
- Somewhat strict regimen schedules, dosage, and treatment duration.
- Prospectively Collected with near mandatory collection and recording of items in a pre-designed Case Report Form. Very limited missing data.

OBSERVATIONAL DATA

- Data includes sources like electronic medical records and registries. Data handling and analysis intensive.
- Typically, non-interventional and physician is free to prescribe at will.
- Can record patient experience from diagnosis through progressions and death.
- Regimen schedules, dosage, treatment durations and combinations are highly variable.
- Transcription of available retrospective and prospective data. Missing data at baseline.

RESTRICTED

6 INTENTIONAL OBSERVATIONAL STUDY DESIGN

- Deliberate study design involving the following blinded to outcomes
  - Pre-specification of population carve-outs and Primary endpoints and hypotheses
  - The choice of baseline characteristics to reduce bias when comparing treated groups on outcomes
  - Pre-specified analysis methods to control for differences at baseline
- Collection/acquisition of all these baseline characteristics with leverage in the analysis:
  - Leading to a differential use of treatments and
  - Having effect on outcome.
- After design stipulation in an analysis plan, unblind to outcome data and conduct analysis.

RESTRICTED

KEY PROPENSITY SCORE THEOREMS - TWO GROUP CASE

- Rosenbaum and Rubin (1983):"A balancing score, *b(x)*, is a function of observed covariates x such that the conditional distribution of *x* given *b(x)* is the same for treated (*z=1*) and control (z=0) units."
- The coarsest such function is the propensity score e(***x***).
- Let $e(\boldsymbol{x}) = P(z = 1|\boldsymbol{x})$ for a vector ***x*** of *p* covariates.
- To show that this is a balancing score, let e(***x***) take some value *g* and χ_g be the set of all covariate values in R^p such that ***e***(.) = *g*. Then conditional on ***e***(***x***) taking a value *g*, ***x*** can vary over the set χ_g with the probabilities of treatment $P(z = 1|\boldsymbol{x} \epsilon \chi_g)$ invariantly equal to *g*, making treatment independent of covariate values conditional on ***e***(***x***). By definition, the set χ_g differs as g differs making e(***x***) coarsest.
- This allows for the use of the propensity score e(***x***) for matching, stratification by score intervals, covariate adjustment and weighting to help obtain unbiased estimates of treatment effect.

RESTRICTED

WEIGHTS FOR SUBJECTS DURING ANALYSIS - FOR ATE AND ATT EFFECTS

- IPTW: let Z_i be an indicator variable denoting whether or not the ith subject was treated; furthermore, let p_i denote the propensity score (PS) for the ith subject then the inverse probability of treatment weighting w_i is defined, in the context of obtaining the Average Treatment Effect (ATE), as
- $w_i = 1/p_i$ for a target therapy subject

  $w_i = 1/(1-p_i)$ for a control subjects

Average treatment effect in the treated (ATT) multiples both weights above by the subject PS p_i

Stabilized weights (reduce variability in estimating treatment effect due to extreme weights due to PS near 0 or 1)

- w_i = P(subject in Target)/p_i for a Target Therapy subject

  w_i = P(subject in Control)/(1-p_i) for a control subjects

  P(subject in Target) = (number of Target subjects) /(number of Target subjects+ number of Control subjects)

RESTRICTED

9 PROPENSITY SCORE (PS) METHODS - MORE THAN 2 GROUPS

- For a large number p of covariates $\boldsymbol{X}$, the results assume that the probabilities of receiving the k treatments (z=1, 2 …k) can be determined without bias.
- Consider vector score
  - $\mathbf{e_s}(\boldsymbol{x})^{\mathsf{T}} = [e_{1s},\ e_{2s}, .., e_{ks}]$, with $e_{is}(\boldsymbol{x}) = P(z = i|\boldsymbol{x}, \boldsymbol{s})$ for a given $\boldsymbol{x}$.
- Proof of the result
  - Let $\mathbf{e_s}(\boldsymbol{x})$ equal some vector value $\boldsymbol{g}$ and χ_g be the set of all covariate values in R^p such that $\mathbf{e_s}(.) = \boldsymbol{g}$.
  - Then conditional on $\mathbf{e_s}(\boldsymbol{x})$ taking a value $\boldsymbol{g}$, $\boldsymbol{x}$ can vary over the set χ_g (the coarsest set by definition) with the probabilities of treatment $P(z = i|\boldsymbol{x} \epsilon \chi_g)$ invariantly equal to g_i
  - This makes treatment independent of covariate values conditional on $\mathbf{e_s}(\boldsymbol{x})$, and thus, a propensity score.
- Results are extensions of Rosenbaum and Rubin (1983) to $k > 2$ in Imai and Van Dyke (2004). Outcome differences on balancing are interpretable as a difference in treatment effects with each effect being the aggregate had all subjects in the population received that treatment (ATE).
- We add a limiting conditioning $\boldsymbol{s}$ referring to a curated sampling process in observational data to be described in our next slide.

RESTRICTED

CURATED (NON-RANDOM) SAMPLING IN OS – NATURE OF DATA ACQUISITION

- Observational data sets, not very much unlike those in clinical trials, can often be just as curated though the nature of the curation differs.
- A non-random sampling process characterized by
  - Data acquired depends on availability of records in electronic form, data purchase costs, data quality, availability of certain diagnostic data and time-frames for data pre-processing such as anonymization and IRB approvals for use.
  - Separate observational data collection, prospective or retrospective, conducted to provide one or more quasi-controls to interventional single arm trials requiring similar contexts.
  - Cohort sizes likely unrelated to any past, current or future population proportions of subjects on the therapeutic options studied.
  - The number and identity of therapeutic groups of interest can be influenced by resources available for data agglomeration and the commercial and research interests of the investigators.

RESTRICTED

ADDRESSING ANALYTICAL ISSUES IN CURATION – USING CONTRASTS

ANALYTICAL ISSUES

- Many therapy groups to select from in observational settings.
- Many choices of comparisons of interest, usually done using treatment contrasts.
- Disparate group sizes.
- We develop the use of contrasts in observational settings to address these issues.

USING CONTRASTS

- The contrasts we use are a string of coefficients, with one for each treatment such that the sum of the contrasts is zero and the sum of the absolute values of the contrasts are equal to 1.0.
- Example [0.5, -0.5, 0] and [0.5, -0.25, -0.25] for three treatment groups.
- In controlled studies, one usually sees the contrast coefficients as 'proportions' in a comparison

RESTRICTED

12 CONDITIONING ON THE CONTRASTED INFERENCE

- Consider two vector propensity probabilities (for some two subject given their covariate profiles) of having treatments 1 to 3 of
  1. {0.1, 0.2, 0.7} and {0.2, 0.4, 0.4}
  2. Then the contrast [0.5, -0.5, 0] to compare treatments 1 and 2 ,
  3. would have had balancing propensity scores {0.33,0.66} for both subjects conditional on the chosen inference if our sampling process did not collect or consider treatment 3.
- Inverse propensity weighting in analysis of outcome, conditionally (# 3 above) requires equal weights while unconditionally (# 1 above) the weights are larger by a factor of 2 for the first subject.
- Rather Odd!

RESTRICTED

13 CURATED DATA CONTRAST EFFECT (CCE)

- For *k* groups, a contrast is defined through
  - a vector $\boldsymbol{c}^T = [c_1, c_2, .., c_k]$
  - where $\sum_{i=1}^{k} c_i = 0$ and $\sum_{i=1}^{k} |c_i| = 1$.
- The absolute value $|c_i|$ is interpretable as the proportion randomly selected in Group *i* independent of the sampling process.
- The probability $P(z = i|\boldsymbol{c}) = |c_i|$.
- For propensities {0.1, 0.2, 0.7} and contrast [0.5, -0.5, 0], conditionally we get

$$\frac{PS_{ij} * |c_i|}{\sum_{i=1}^{k} PS_{ij} * |c_i|} = \{0.33, 0.66\}$$

- We will refer to inferences drawn in this setting as Curated Contrast Effects.

RESTRICTED

PROPENSITY SCORES UNCONDITIONAL (LEFT) AND CONDITIONAL (RIGHT) ON CONTRASTED INFERENCE

- Balancing Propensity Score

$$\mathbf{e_s}(\boldsymbol{x})^{\mathsf{T}} = [e_{1s},\ e_{2s}, .., e_{ks}], \text{ with}$$

$$e_{is}(\boldsymbol{x}) = P(z = i|\boldsymbol{x}, \boldsymbol{s}) \text{ for a given } \boldsymbol{x}.$$

- Balancing propensity Score

$$\mathbf{e_{cs}}(\boldsymbol{x})^{\mathsf{T}} = [e_{1cs},\ e_{2cs}, .., e_{kcs}], \text{ with}$$

$$e_{ics}(\boldsymbol{x}) = P(z = i|\boldsymbol{x}, \boldsymbol{c}, \boldsymbol{s})$$

$$= \frac{P(z = i \mid \boldsymbol{x}, \boldsymbol{s}) * P(z = i \mid \boldsymbol{c})}{\sum_{i=1}^{k} P(z = i \mid \boldsymbol{x}, \boldsymbol{s}) * P(z = i \mid \boldsymbol{c})}$$

RESTRICTED

WEIGHTS UNCONDITIONAL (LEFT) AND CONDITIONAL (RIGHT) ON CONTRASTED INFERENCE

- Expression for the Sample Weights

$$\Rightarrow\ e_{is}(\boldsymbol{x}) = P\left(z = i\middle|\boldsymbol{x}, \epsilon\chi_{\boldsymbol{g}}, \boldsymbol{s}\right) = \frac{P(\boldsymbol{x}\,\epsilon\chi_{\boldsymbol{g}}|z=i,\boldsymbol{s})P(z=i|\,\boldsymbol{s})}{P(\boldsymbol{x}\,\epsilon\chi_{\boldsymbol{g}}|\,\boldsymbol{s})}$$

$$\Rightarrow P(\boldsymbol{x}\,\epsilon\chi_{\boldsymbol{g}}|\,\boldsymbol{s}) = \frac{P(\boldsymbol{x}\,\epsilon\chi_{\boldsymbol{g}}|z=i,\boldsymbol{s})P(z=i|\,\boldsymbol{s})}{e_{is}(\boldsymbol{x})}$$

- Then the sample weight below will weight a covariate profile in $\boldsymbol{x}\,\epsilon\chi_{\boldsymbol{g}}$ such that it is not predictive of treatment in the weighted sample.

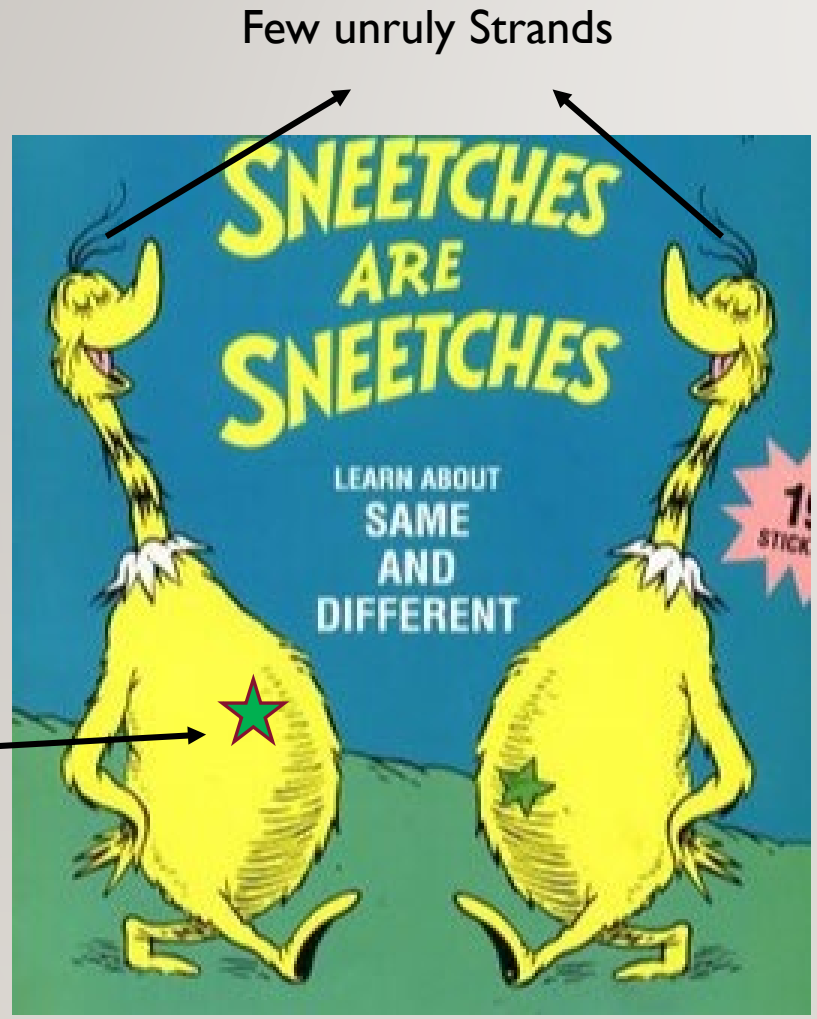
$$\frac{P(z=i|\,\boldsymbol{s})}{e_{is}(\boldsymbol{x})}$$

- Expression for the Sample Weights for covariate profiles to make un-predictive of treatment

$$\frac{P(z=i|\,\boldsymbol{c},\boldsymbol{s})}{e_{ics}(\boldsymbol{x})} \text{ for } e_{\boldsymbol{ics}}(\boldsymbol{x}) \text{ as above}$$

$$P(z=i|\,\boldsymbol{c},\boldsymbol{s}) = = \frac{P(z=i\,|\boldsymbol{s}) * |c_i|}{\sum_{i=1}^{k} P(z=i\,|\boldsymbol{s}) * |c_i|}$$

RESTRICTED



*Snapshot from goodreads.com

- As you can see, in the world of the Sneetches there is perhaps only one differentiating factor.
- Note that the differentiating factors should plausibly have effect on outcome.
- In addition to a visitation by a dubious entrepreneur selling star-on machines, what is a little less known, is perhaps a less dubious pitch of hair restoration therapies to the Sneetches.
- We will look at balancing scores in this context for 3 snake oil treatments for hair growth where there are biases based on just one covariate -- The Plain Belly/Green Star Belly

RESTRICTED

17 THOUGHT EXERCISE – UNCONDITIONAL TO LEFT - CONDITIONAL ON CONTRAST ON RIGHT

INPUTS				
TRT	Contrast	N	# Green (G)	# Plain (P)
1	0.500	80	30	50
2	-0.250	40	10	30
3	-0.250	70	50	20

Probability of TRT (Propensity) Given Covariate Value and Inverse				
TRT	Given G	Given P	Inverse G	Inverse P
1	0.333	0.500	3.000	2.000
2	0.111	0.300	9.000	3.333
3	0.556	0.200	1.800	5.000

Scaled Ordinary Weight and Sum of Weights				
Scaled G	Scaled P	Sum G	Sum P	Sum All
1.263	0.842	37.895	42.105	80.000
1.895	0.702	18.947	21.053	40.000
0.663	1.842	33.158	36.842	70.000

INPUTS				
TRT	Contrast	N	# Green (G)	# Plain (P)
1	0.500	80	30	50
2	-0.250	40	10	30
3	-0.250	70	50	20

Probability and Inverse Given Covariate and Contrast				
TRT	Given G	Given P	Inverse G	Inverse P
1	0.500	0.667	2.000	1.500
2	0.083	0.200	12.000	5.000
3	0.417	0.133	2.400	7.500

Derived Weight and Sum Given Covariate and Contrast				
Derived Wt G	Derived Wt P	Sum G	Sum P	Sum All
1.185	0.889	35.556	44.444	80.000
1.778	0.741	17.778	22.222	40.000
0.622	1.944	31.111	38.889	70.000

RESTRICTED

SAS CODE - RAW P-SCORE TO LEFT AND - WEIGHTS TO RIGHT

```
*Dataset has Subject ID, TRT (3 groups) and Covariates X1,
X2 and X3 and sample proportions P_T1, P1_T2 and P1_T3;

 Proc Logistic data = sample1_sort noprint;
     CLASS trt X1;
     Model trt = X1 X2 X3;      by SIM;
     Output out = pred1 pred = S1_PP;
  run;

  Data pred1_T1; set pred1;
    if _Level_ = 1;
    PS_T1 = S1_PP;     Drop S1_PP;
  run;

  proc sort data = pred1_T1 out = pred1_T1s;
    by SIM ID;
   Run;

   Data pred1_T2; set pred1;
    if _Level_ = 2;
    PS_T12 = S1_PP;     Drop S1_PP;
  run;
```

```
data PS1;
      Merge pred1_T1s pred1_T2s;
      by SIM ID;
      PS_T2 = PS_T12 - PS_T1 ;
      PS_T3 = 1 - PS_T1  - PS_T2;
      IF TRT = 1 then do;
        *sample treatment proportions P(Z=i/s);
         P_UC = P_T1;
         *Eis(x) →;          PS_UC = PS_T1;
        *treatment proportion conditional on contrast P(Z=i/s,C);
        P1_C2 =  0.5*P_T1/(0.5*P_T1 +0.25*P_T2+ 0.25*P_T3);
        *Eisc(x) →;         PS_C2 = 0.5*PS_T1/(0.5*PS_T1
+0.25*PS_T2+ 0.25*PS_T3);
       *compute unconditional weight and weights conditional on
C2;
        WT_UC = P_UC/PS_UC;
        WT_C2 = P_C2/PS_C2;
       end;
      *similar code for treatment 2 and 3;
   Else if TRT = 2 then do;
     *.....;
     *.....;
   end;
  run;
```

RESTRICTED

19 BALANCE DIAGNOSTICS

- Side by Side Raw Means (SD) juxtaposed with Side-by-Side Weighted Means (SD) for continuous covariates by Treatment
- Side by Side Raw Counts and Proportions juxtaposed with Side by Side Weighted Proportions for discrete covariates by Treatment
- Standardized Mean Differences (SMD) for Raw versus Weighted covariates
  - Two Group Case => {M1 - M2}/SQRT{(V1+V2)/2} for sample means M and Variances V
  - For our Contrasts C1 = [0.5 -0.5, 0] => 2(0.5*M1 -0.5*M2)/Sqrt(0.5*V1+0.5*V2) and
  - For C2 = [0.5 -0.25, -0.25] => 2(0.5*M1 -0.25*M2-0.25*M3)/Sqrt(0.5*V1+0.25*V2+0.25V3)

RESTRICTED

SAS CODE – WEIGHTED MEANS AND SD TO LEFT - STANDARDIZED MEAN DIFFERENCE RIGHT

```
*PS1_sorted data contains Subject ID, TRT (3 groups) and
Covariate X2
and the Weights WT_C2 WT_CI  and WT_UC;
Proc means data = PS1new noprint;
      Var X2;
      by sim trt;       Weight WT_C1;
      Output out = base mean = M STD = SD;
 run;
Data base1 (drop = trt _type_ _Freq_) ; set base;
  if trt = 1;   M_1 = M;   SD_1 =  SD;
run;
Proc sort data = base1;   by sim; run;

Data base2  (drop = trt _type_ _Freq_) ; set base;
  if trt = 2;   M_2 = M;   SD_2 =  SD;
run;
Proc sort data = base2;    by sim; run;

Data base3  (drop = trt _type_ _Freq_) ; set base;
  if trt = 3;   M_3 = M;   SD_3 = SD;
run;
Proc sort data = base3;    by sim; run;
```

```
Data base_all; merge base1 base2 base3;
   by sim;
run;

*For contrast C2 = [0.5,-0.25,-0.25];
Data SMD; Set base_all;
     V1 = (SD_1)**2;
     V2 = (SD_2)**2;
     V3 = (SD_3)**2;
     DENOM_C2 = SQRT(0.5*V1+0.25*V2 +0.25*V3);
     DIFF2 = 2*(0.5*M_1 - 0.25*M_2 - 0.25*M_3);
     SMD_C2 = DIFF2/DENOM_C2;
run;
```

RESTRICTED

21 WEIGHTED TIME TO EVENT ANALYSIS

SAS SETS TRT3 COEFFICIENT TO 0 AND USES THE MODEL:

LOG $H(T)$ = LOG $H_0(T)$ + $B1$*TRT1 + $B2$*TRT2

	Design Matrix	
TRT	*B1*	*B2*
1	1	0
2	0	1
3	0	0
Contrast C2	Contrast by subtraction	
1	1	0
0.5*(2 and 3)	0	0.5
Trt 1 vs Trt2 &3	1	-0.5

WEIGHTED COX REGRESSION

*PS1_sorted2 data contains Subject ID, TRT (3 groups), TTE duration and Censor and the Weights;
*SAS contrast for C2 given reference group of trt 3;

```
Proc phreg data = PS1s;
      class trt (ref = '3')/ param=ref
order=internal;
      Model TTE*censor(0) = trt;
      Weight WT_C1;
      by SIM;
      contrast 'trt1 vs trt2' TRT 1 -0.5 /
estimate =parm;
      ods output ContrastEstimate = WT_C2;
run;
```

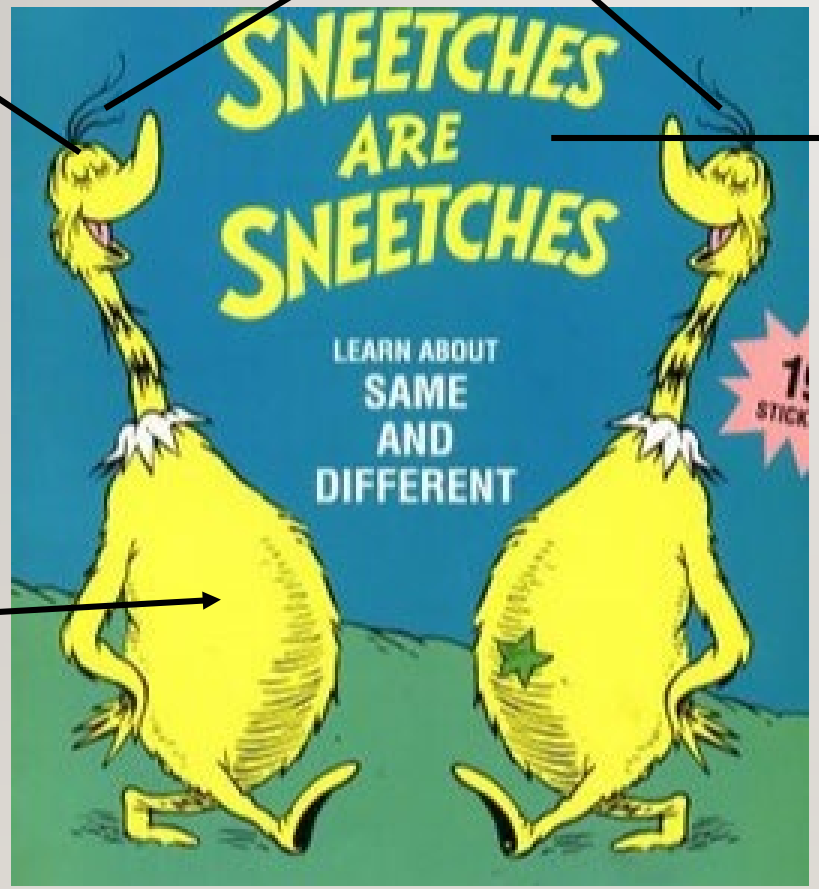
RESTRICTED

Simulation Exercise

24

Three Hair Growth Treatments

Bald Spot Luminosity ~ Exponential with a mean of 0.5 watts



Height ~ Exponential With Mean of 50 inches.

Plain or Star Belly ~ Bernoulli with p =0.4

1. Simulate covariate values on 50K subjects in the population over 50 simulations
2. Generate treatments using coefficients predicting treatment
3. Generate time-to event outcome (time to complete baldness)
4. Pick curated sample selecting 100, 220 and 310 simulation subjects on treatments 1, 2 and 3 respectively
5. Find conditional and unconditional weights
6. Use weights in outcome models and compare estimated treatment effects versus simulation parameters (#3 above)

RESTRICTED

SIMULATION PARAMETERS GENERATING TREATMENTS: TREATMENT BIAS MODEL

Treatment Bias Model	X1 (Belly Star)	X2 (Height)	X3 (Bald Patch Luminosity)
LOGIT (Trt 1 vs Trt 2)			
Odds Ratios	1.65	1.45	1.2
Coefficients Corresponding to Odds Ratio Above	0.5008	0.3716	0.1823
LOGIT (Trt 1 vs Trt 3)			
Odds Ratios	1.25	1.10	1.5
Coefficients Corresponding to Odds Ratio Above	0.2231	0.0953	0.4055

RESTRICTED

SIMULATION PARAMETERS GENERATING OUTCOME: OUTCOME MODEL

Outcome Model	Trt 1 vs Trt 2 Contrast: [0.5, -0.5, 0]	Trt 1 vs Trt 3 Contrast: [0.5, -0.5, 0]	Trt 1 vs {Trt2 and Trt 3} Contrast: [0.5, -0.25, -0.25]	X1 (Belly Star)	X2 (Belly Volume)	X3 (Bald Patch Luminosity)
Hazard Ratio for Time to Complete Baldness (X3)	0.6	0.75	0.6708	0.95	1.35	1.60
Coefficients for Cox Regression for X3 above	-0.5108	-0.2877	-0.3993	-0.0513	0.3001	0.4700

A 5-year follow-up (20 Sneetch years) was used, 1.5 years uniform enrollment, exponential distributions and a baseline hazard (X1 =0 and X2 and X3 at mean values) corresponding to a median of 3 years. HR for X2 and X3 correspond to SD increments over mean.

RESTRICTED

STANDARDIZED MEAN DIFFERENCE OVER 50 SIMULATIONS - TRT 1 VS {TRT 2 AND TRT3} CONTRAST

Standardized Mean Difference (SMD)	X1 (Belly Star)	X2 (Height)	X3 (Bald Patch Luminosity)
Raw Means			
Average SMD	0.204	0.188	0.260
STD of SMD	0.089	0.123	0.117
Weighted Means -Unconditional			
Average SMD	0.204	0.256	0.003
STD of SMD	0.077	0.089	0.065
Weighted Means - Conditional			
Average SMD	0.203	0.239	0.068
STD of SMD	0.072	0.092	0.074

RESTRICTED

SIMULATION RESULTS OVER 50 SIMULATIONS - TRT 1 VS {TRT 2 AND TRT3} CONTRAST

	Parameter	Non-weighted using covariates	Unconditional on Contrast	Conditional on Contrast
LN(HR)	-0.3993	-	-	-
Hazard Ratio (HR)	0.6708	-	-	-
Simulation Results	Curated Sample selecting 100, 220 and 310 simulated subjects in Treatments 1, 2 and 3 respectively			
AVG Estimate of LN(HR)	-	-0.561	-0.434	-0.414
Bias in LN(HR)	-	-0.162	-0.034	-0.015
Associated HR	-	0.570	0.648	0.661
STD of Estimates	-	0.170	0.168	0.165
Average SE over Simulations	-	0.149	0.148	0.147

RESTRICTED

CONCLUSIONS/QUESTIONS

- The unconditional weights are
  - Consistent with stabilized weights for >2 groups and consistent with the contrast conditional weights only for 2 groups and the contrast [0.5, -0.5].
  - For the three-group pair-wise case results were similar.
- Cox regression inferential analysis bias and variance using weights were much improved compared to using covariates without weighting, with improvements on conditioning on contrasts to be further explored.
- Standardized mean differences were only selectively reduced using both weighting methods. The Coarseness of PS scores, while making treatment independent of the overall covariate profile, may leave differences on individual covariates.

RESTRICTED

32 REFERENCES

- Rosenbaum, P.R., and Rubin, D. B. (1983): The central Role of the Propensity Score in Observational Studies for Causal Effects, Biometrika, 70, 41-55.
- Rosenbaum, P.R., and Rubin, D. B. (1984): Reducing Bias in observational studies using subclassification on propensity score, Journal of American Statistical Association 79, N387, 516-524.
- McCaffrey Daniel F et al (2013). Atutorial on Propensity Score Estimation for Multiple Treatments Using Generalized Boosted Models. Statistics in Medicine. 32(19): 3338-3414.
- Paul T Savarese and Michael J. Paletta, (2010). An Overview of the CLASS, CONTRAST, and HAZARDRATIO Statements in the SAS® 9.2 PHREG Procedure. Paper 253-2010. SAS Institute Inc., Cary, NC
- Peter C Austin (2009): Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. Statistics in Medicine 2009; 28: 3083-3107
- Peter C. Austin and Elizabeth A. Stuart (2015): Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. Statistics in Medicine 2015; 34: 3661-3679
- Rosenbaum PR. Observational Studies. Springer, New York, 2nd edition, 2002
- Yue LQ. Statistical and regulatory issues with the application of propensity score analysis to nonrandomized medical device clinical studies. J Biopharm Stat.2007;17(1):1-13.
- Levenson MS, Yue LQ. Regulatory issues of propensity score methodology application to drug and device safety
- studies.J Biopharm Stat. 2013;23(1):110-21.
- Matao Sun. A Review of propensity score use in drug post-marketing safety surveillance. EliScholar, Yale University, January 2016.
- Matthew J Smith et al (2022): Introduction to computational causal inference using reproducible Stata, R, Python code: A tutorial
- Imai K, Ratkovic M. Covariate balancing propensity score. J. R. Statist. Soc. B (2014) 76, Part 1, pp. 243–263
- Imai Kosuke and David A.. Van Dyk (2004): Causal inference with general treatment regimes: Generalizing the Propensity Score. , Journal of American Statistical Association 99, N467, 854-866
- The Sneetches and Other Stories. Dr Suess. Random House, 1961.

RESTRICTED