Overview of Some Key Methods for Observational Studies

Shankar Srinivasan, Ph.D (Presenter), Lihua Yue, PH.D, Lorraine Fang, MS and Everton Rowe, Ph.D.

BASS XXVI, Tuesday October 22nd

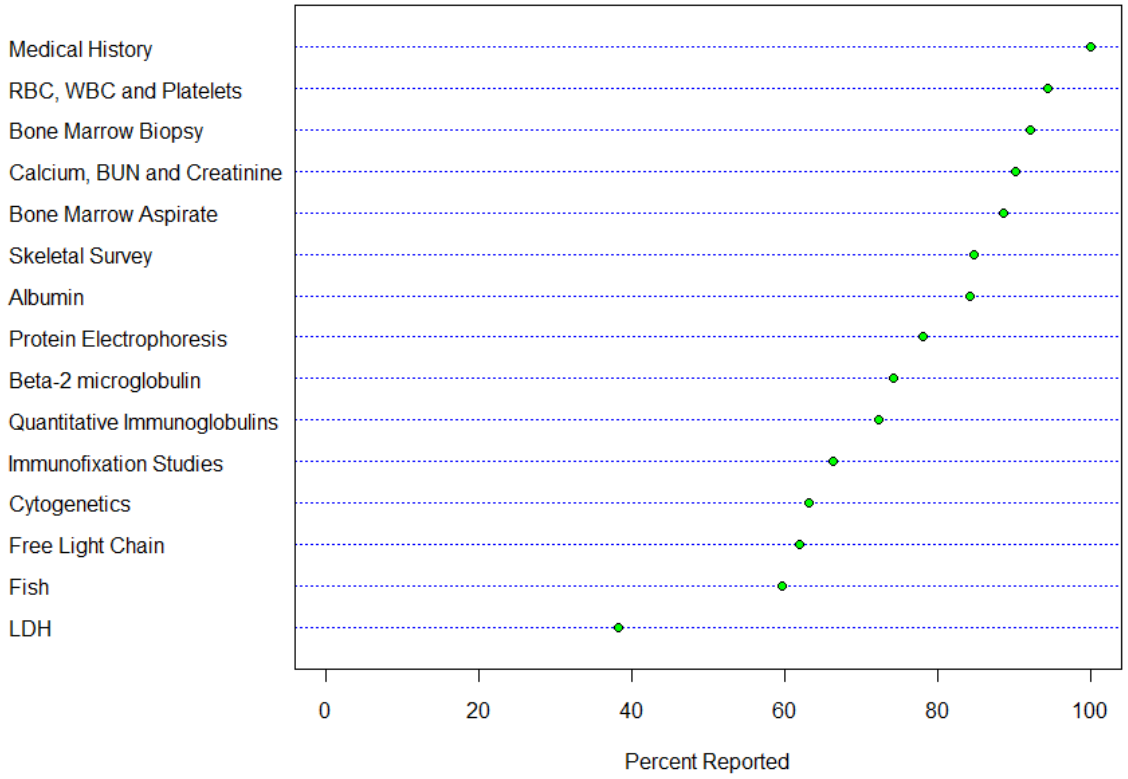
Disclosures

- Employment
  - Biostatistician at Celgene Corporation, with salary and stock compensation.
- Blog at resourcetepee.com. Some resources from that blog will be referenced during this tutorial.
- The material presented here represent the views of the presenter and not those of the Celgene Corporation
  - Slides have generic issues and details on Observational Data, some material from his website and a technical report he authored (to be published).

Outline

- About Observational Data
- General approaches for valid inferences not using Propensity scores
  - Notes on Missing Data, Covariate Adjustment, Exact Methods
- Propensity score based methods with emphasis on the setting with > 2 groups
  - Balancing scores
  - Weights conditional on treatment contrasts
  - IPTW weighting, Matching and Sample size
  - Simulation example
- Questions and Discussion

Real World Data Issues and Positives



- Missing data requiring imputation methods (see left from Rifkin et al.*).
- Significant 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- Without a pre-specified pathway through the data there can be publication bias due to unpublished results and inflation in false positive rates.
- 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.

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 especially “baselines” during the course of disease such as the start of maintenance or post-relapse therapy.

An Illustrative Example

- We will consider hypothetical real world oncology data with patients receiving an Immune drug or a Chemotherapy. For explanatory ease we will consider relative medians instead of the hazard ratios which are usually reported.
- Physicians channel older patients to Immune drug due to better tolerability. Those receiving the immune drug are on average 10 years older than those receiving the chemotherapy.
- Lets say, that for a 10 year older population, there is a drop in the median survival by 8 months in the aggregate curve.
- An unadjusted analysis might report medians of 60 months and 56 months for the chemo and the immunotherapy groups, whereas the adjusted analysis would have reported median differences of 4 months in the opposite direction of 56 versus 60.
- The adjusted analysis corrects the chemo group down by 4 months and the immune group up by 4 months to account for the 8 month artifact due to age differences.

Framework for Inferential Analyses With Real World (RW) Data

- Inferential comparisons between cohorts are not valid unless we can argue that all relevant patient characteristics differentiating the groups and possibly effecting outcome have been collected, allowing us to adjust for these differences when comparing cohorts on outcome.
- A fire-wall* between determining characteristics differentiating groups and the eventual outcome analysis. This and population carve-outs, endpoints, hypotheses and analyses, which are primary, should ideally be pre-specified.
- Start with the screening of factors differentiating cohorts using univariate tests and multivariate variable selection on patient characteristics predicting cohort membership.
- Force in variables predictive of outcome per expert opinion to selected variables.
- Bring in the firewalled outcome data and conduct inferential analyses adjusting for the factors above or analyze differences between culled matched sets.

see Levenson and Yue, J Biopharm Stat. 2013;23(1):110-21

Data Imputation in Analyses: Imputation Theory

- Obtains the predictive distribution of the missing data given the remaining patient data.
- Imputes missing data repeatedly using such predictive distributions and creates Multiply Imputed (MI) datasets. Imputes a random smear.
- Inferences from each imputed dataset are combined and reported.
- Analysis under multiple imputation is robust under less restrictive assumptions of Missing at Random (MAR)
- Methods under MAR avoid understating data variability unlike an imputation method which plugs in a specific value instead of a smear.
- 'ANCOVA' like conventional inferential survival analysis in RW setting follows.

Direct Adjusted Survival Curves using Average Survival Curve Over Covariate Values Over Entire Dataset

Unstratified - Forced Proportional Hazard and Identical Baseline Hazard

$$\hat{S}_0(t) = \frac{1}{n}\sum_{l=1}^{n} \exp\{-\hat{\Lambda}_0(t)e^{\hat{\beta}_2 Z_{2l}+\hat{\beta}_3 Z_{3l}}\}$$

$$\hat{S}_1(t) = \frac{1}{n}\sum_{l=1}^{n} \exp\left\{-\hat{\Lambda}_0(t)e^{\hat{\beta}_1+\hat{\beta}_2 Z_{2l}+\hat{\beta}_3 Z_{3l}}\right\}$$

```
proc phreg data=spm;
  class trt;
  model spmm*spm_invas(0)=  trt diag
       pinvas  ;
  baseline out=covbase  survival =
       _all_ / diradj group=trt;
run;
```

Stratified – Differing Baseline Hazards Across Groups

$$\hat{S}_i(t) = \frac{1}{n}\sum_{l=1}^{n} \exp\{-\hat{\Lambda}_{0i}(t)e^{\hat{\beta}_2 Z_{2l}+\hat{\beta}_3 Z_{3l}}\}$$

Here i=0 and i=1 are the strata defined by the treatment variable. Both analyses use averaging of survival curves over for all patients (entire covariate profile set) instead of the the survival curve for average covariate values.

```
proc phreg data=spm;
  strata trt;
  model spmm*spm_invas(0)=
       diag pinvas  ;
  baseline out=stratbase  survival =
       _all_ / diradj ;
run;
```

Zhang Xu et. al. (2007). A SAS macro for estimation of Direct Adjusted Survival Curves based on a stratified Cox Regression Model. Computer methods and programs in biomedicine. 88: 95-101.

Code for Covariate Adjusted Inferential Analysis: Generating Adjusted Survival Curves

```
/&nimpute: number of imputations;
&imputvar: list of variables used in the
imputation (include all the variables
selected from the variable selection process,
but add more common variables);
&classvar: list of class variables in
&imputvar;
&strata: variable representing groups in the
direct adjusted survival plot;
&adjcov: list of covariates to be adjusted in
Cox model;
&classAdjcov: list of class covariates in
&adjcov;*/
```

```
proc mi data=&indata nimpute=&nimpute
        seed=&seed out=_MIdata;
      class &classvar;
      fcs nbiter=20 logistic(&classvar /
        likelihood=augment);
      var &imputvar;
run;
```

- get direct adjusted survival probabilities by _imputation_;

```
proc phreg data=_MIdata;
     by _Imputation_;
     class &strata &classAdjcov /
        param=ref;
     model &time.*&censor.(0) =
        &adjcov/ ties=Efron;
     strata &strata;
     baseline covariates= MIdata
        out=_est survival=_all_ / diradj;
run;
```

```
proc sort data=_est(where=(&time>0))
        out=_estD;
        by &strata &time _Imputation_; run;
```

```
proc mianalyze data=_estD;
     by &strata &time;
     modeleffects Survival;
     stderr StdErrSurvival;
     ods output ParameterEstimates=_est1;
run;
```

Code for Covariate Adjusted Inferential Analysis: Generating Inferential Statistics

```
ods output LSMEstimates=LSMEsts;
```

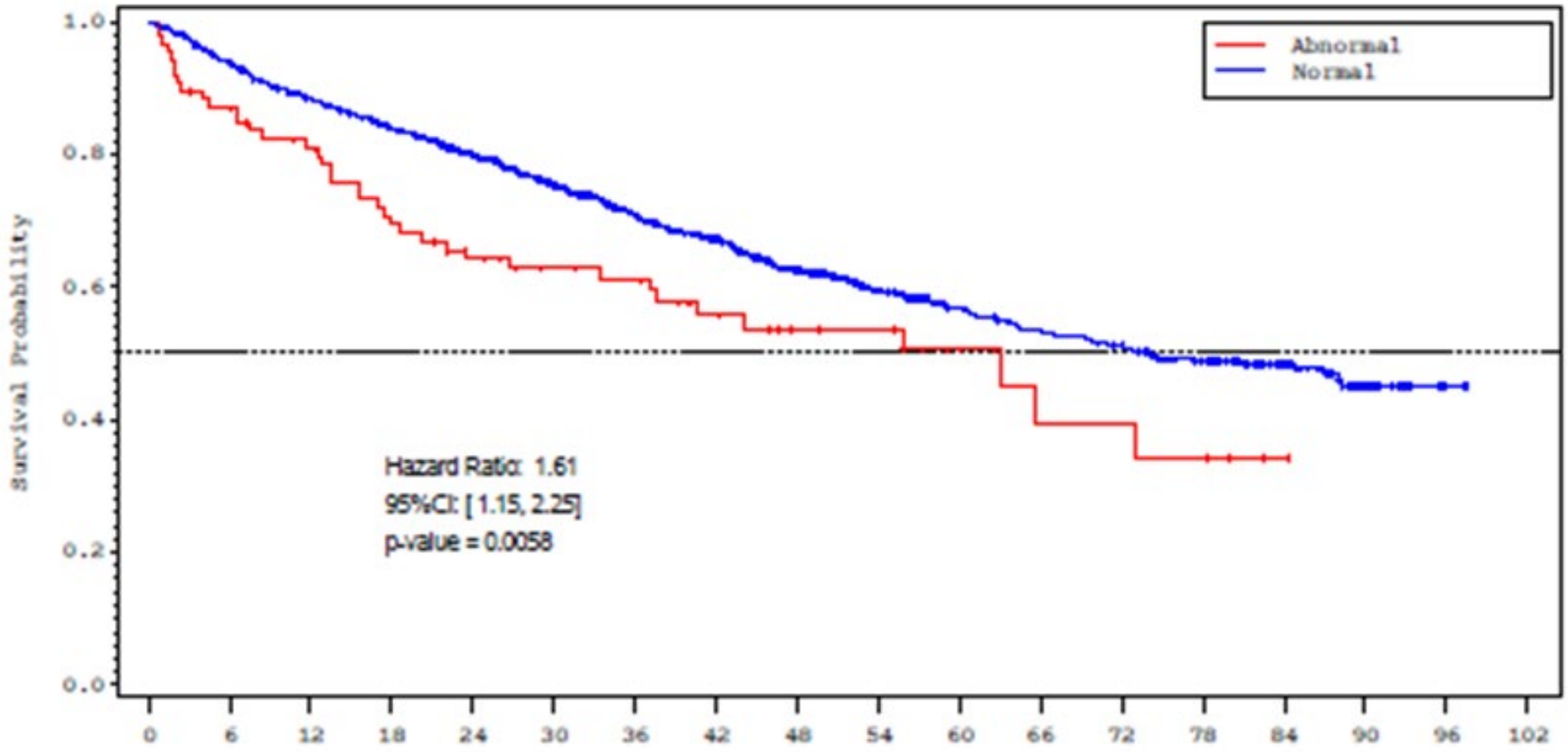
```
proc phreg data=_MIdata;
   by _Imputation_;
   class &strata. &classAdjcov /
       param=glm;
   model &time.*&censor.(0) =
       &strata &adjcov/ties=Efron covB;
   lsmestimate &strata
       'Abnormal  vs  Normal' 1  -1;
run;
```

```
proc sort data=LSMEsts out=_estHR;
      by StmtNo Label _Imputation_;
run;
```

```
proc mianalyze data=_estHR;
      by StmtNo Label;
      modeleffects Estimate;
      stderr StdErr;
      ods output
        ParameterEstimates=_estHR1;
run;
```

```
data _estHR2;
      set _estHR1;
      HR = exp(Estimate);
      HRLowerCl = exp(LCLMean);
      HRUpperCl = exp(UCLMean);
      pval = Probt;
      keep StmtNo Label HR HRLowerCl
        HRUpperCl pval;
run;
```

Adjusted Inferential Analysis for those Normal Versus those Abnormal (De-identified) on Reported Measure



Number at Risk:

Group	0 Months	6 Months	12 Months	18 Months	36 Months	54 Months	72 Months	90 Months
Abnormal	76	64	56	46	31	15	5	0
Normal	1052	969	883	817	406	214	152	25

Covariates: Cohort, Age group, BMI Category, del(17p), Platelet counts

Exact Matching

A Disease Related Factor	A Frailty or Co-morbidity Factor	N Control	N Target Therapy	Matched per Group
Level 1	Level 1	65	32	32
Level 2	Level 1	21	28	21
Level 3	Level 1	18	12	12
Level 1	Level 2	27	22	22
Level 2	Level 2	29	40	29
Level 3	Level 2	31	51	31
	Totals →	191	184	147

Can be useful when data on all except for a few core factors are unavailable.
Should be the variables most relevant to outcome.
Often erroneously, something perhaps unrelated, such as gender or race is used.
Very easy to communicate to non-statisticians. Leads to loss of data records. Though stratified analyses could be conducted.

Propensity Score (PS) Methods - More than 2 groups

- For a large number p of covariates ***X***, the results assume that the probabilities of receiving the *k* treatments (*z*=1, 2 …*k*) can be determined without bias.
- Consider vector propensity 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 ***x***.
- Proof of the result
  - Let $\mathbf{e_s}(\boldsymbol{x})$ equal some vector value ***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 ***g***, ***x*** can vary over the set χ_g 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 $\mathbf{e_s}(\boldsymbol{x})$.
- 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 ***s*** referring to a curated sampling process in observational data to be described in our next slide.

Curated Sampling in Observational Data

- 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.

Curated Contrast

- With many therapy options, one investigator may do a data carve-out of subjects on 5 therapy groups and another may pick 3 or 4 which may or may not overlap with the 5.
- Relative group effects of interest may involve further subsets, such as pairwise comparisons of the groups or the comparison of the outcome in one group to that in two others.
- Such relative groups effects are usually assessed using contrasts, in general settings where bias is not expected, usually among randomized groups.
- We refer to contrasts in the observational settings having neither randomization nor random sampling, as curated 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.

Conditioning on the Contrasted Inference as Well

- 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!

Curated Contrast Effect (CCE)

- For *k* groups, a contrast is defined as
  - a vector $\boldsymbol{c}^T = [c_1, c_2, .., c_k]$
  - where $\sum_{i=1}^{k} c_i = 0$ and $\sum_{i=1}^{k} |c_i| = 1$.
- Let C_k be the set of all $\boldsymbol{c}$ meeting these conditions.
- For propensities {0.1, 0.2, 0.7} and contrast [0.5, -0.5, 0], conditionally we get

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

- So, absolute value $|c_i|$ is interpretable as the proportion randomly selected in Group *i* independent of the sampling process.
- The probability $P(z = i|\boldsymbol{c}) = |c_i|$.
- We will refer to inferences drawn in this setting as Curated Contrast Effects.

Unconditional and Conditional Weights

Unconditional on Contrasted Inference

- Balancing Score

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

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

- Expression for the Sample Weights

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

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

- 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.

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

Conditional on Contrast

- Balancing Scores

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

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

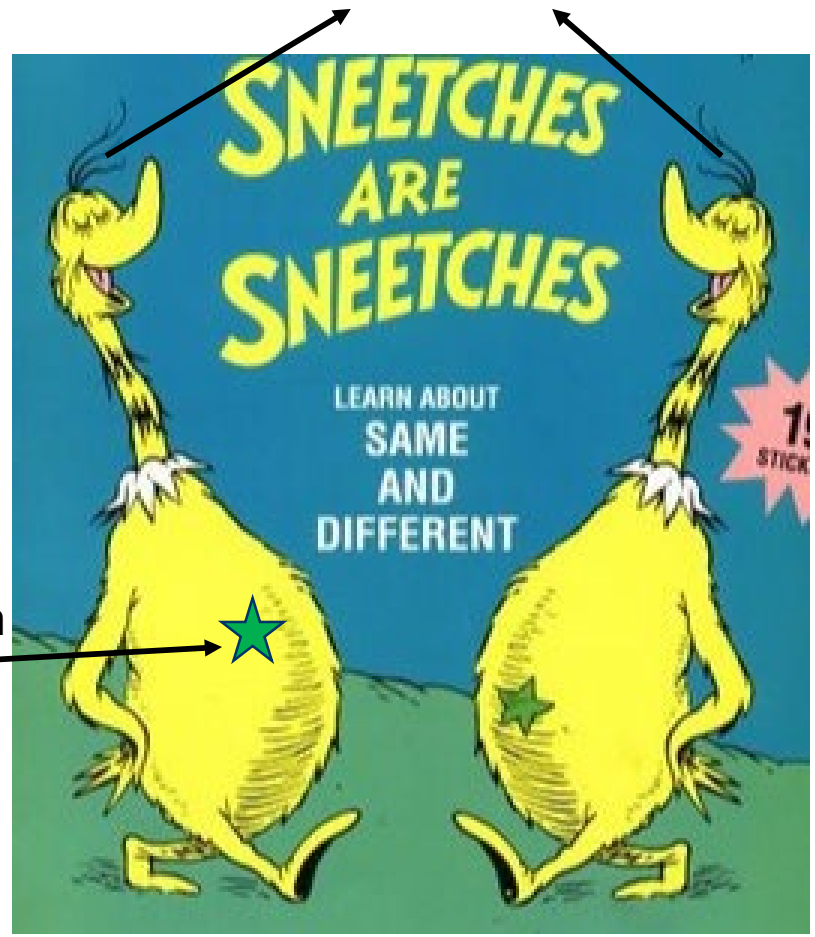
$$= \frac{P(z=i\mid \boldsymbol{x},\boldsymbol{s}) * P(z=i\mid \boldsymbol{c})}{\sum_{i=1}^{k} P(z=i\mid \boldsymbol{x},\boldsymbol{s}) * P(z=i\mid \boldsymbol{c})}$$

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

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

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

The Sneetch* Example



*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 treatments for hair growth where there are biases based on the Plain/Star Belly (Covariate).

Illustrative Calculator*

Unconditional on Inference

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

Conditional on Contrast Inference

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

The sum of weights in a group add up to the group sizes in this context, as is achieved in outcome analysis through the scaling of the inverse PS weights by the average weight in a group.

*Interactive calculator is available at https://resourcetepee.com/free-statistical-calculators/observational-data/contrast-effects-in-curated-observational-data/

Propensity score (p), logit and inverse probability of treatment weighting (IPTW)

- 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].
- 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 for the ith subject then the inverse probability of treatment weighting w_i is defined as
- $w_i = 1/p_i$ for a target therapy subject

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

Stabilized weights

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

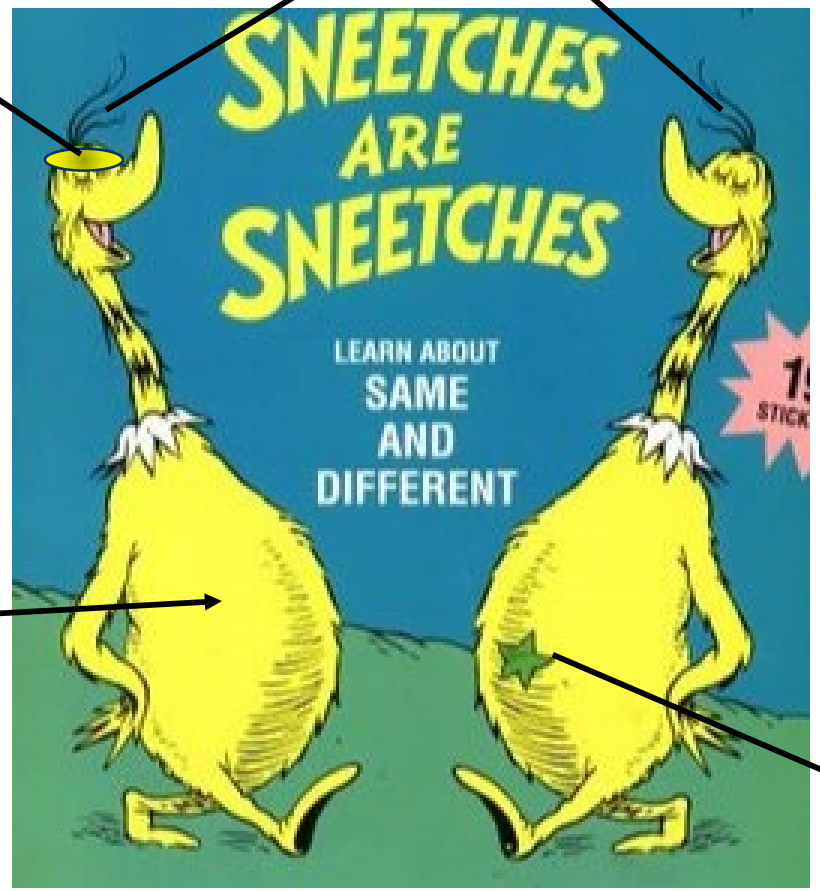
  w_i P(subject in Control)/(1-p_i) for a control subjects

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

The Sneetch* Example: Simulation** Study

Three Hair Growth Treatments

Bald Spot Luminosity ~ Exponential with a mean of 0.5 watts



Plain or Star Belly ~ Bernoulli with p =0.4

Belly Volume ~ Exponential With Mean of 50 cu. inches.

*Snapshot from goodreads.com

1. Simulate covariate values on a million subjects in the population
2. Generate treatments using coefficients predicting treatment
3. Generate response (> 25% reduction), a continuous outcome (% reduction in bald area) and time-to event outcome (time to complete baldness)
4. Pick two curated samples with one selecting 100, 220 and 310 simulated subjects and other selecting after replacement 200, 125 and 75 having treatments 1, 2 and 3 respectively
5. Find conditional (one per contrast) and unconditional weights (one for both contrasts)
6. Use in outcome models and compare estimated treatment effects versus parameters in the simulation of outcome in #3 above.

**The complete derivations of the weighting conditional on the Contrasted inference and the proposed simulation are in a technical report by the author on Research Gate.

Simulation Parameters for the Treatment Bias Model.

Treatment Bias Model	X1 (Belly Star)	X2 (Belly Volume)	X3 (Bald Patch Luminosity)
LOGIT (Trt 1 vs Trt 2)			
Odds Ratios	1.25	0.95	1.05
Coefficients Corresponding to Odds Ratio Above	0.2231	-0.0513	0.0488
LOGIT (Trt 1 vs Trt 3)			
Odds Ratios	0.9	1.10	1.2
Coefficients Corresponding to Odds Ratio Above	-0.1054	0.0953	0.1823

Covariate values generated for a million simulated subjects followed by Treatment by using the model coefficients above.

Simulation Parameters for the Outcome Models.

Treatment Bias Model	Trt 1 vs Trt 2 Contrast: [0.5, -0.5, 0]	Trt 1 vs Trt 3	Trt 1 vs {Trt2 and Trt 3}. Contrast: [0.5, -0.25, -0.25]	X1 (Belly Star)	X2 (Belly Volume)	X3 (Bald Patch Luminosity)
Odds Ratio of Response vs Non Response (X1)	1.6	1.4	1.497	1.05	0.95	0.8
Coefficients for LOGIT of X1 above	0.470	0.3365	0.4033	0.0488	-0.0513	-0.2231
Regression Coefficients for X2 (Percent Reduction in Bald Area)	10	7	8.5	2	-1.5	-2.5
Hazard Ratio for Time to Complete Baldness (X3)	0.6	0.75	1.491	0.95	1.25	1.35
Coefficients for Cox Regression for X3 above	-0.5108	-0.2877	-0.3993	-0.0513	0.2231	0.3001

Outcome values generated for a million simulated subjects. For TTE, we used 5 year follow-up (20 Sneetch years), 1.5 years uniform enrollment, exponential distributions and a baseline hazard (X1 =0 and X2 and X3 at mean values) corresponding to 3 years.

Parameters used and Estimates to be obtained using Unconditional and Conditional Weights in Simulation Study

	Population Parameters in Simulation		Estimates (CI) Using Unconditional Weights (One per sample)		Estimates (CI) Using Weights Conditional on Inference (one per contrast and sample)	
Treatment Bias Model	Trt 1 vs Trt 2. Contrast: [0.5, -0.5, 0]	Trt 1 vs {Trt2 and Trt 3}. Contrast: [0.5, -0.25, -0.25]	Trt 1 vs Trt 2. Contrast: [0.5, -0.5, 0]	Trt 1 vs {Trt2 and Trt 3}. Contrast [0.5, -0.25, -0.25]	Trt 1 vs Trt 2. Contrast: [0.5, -0.5, 0]	Trt 1 vs {Trt2 and Trt 3}. Contrast: [0.5, -0.25, -0.25]
Curated Sample selecting 100, 220 and 310 simulated subjects in treatments 1, 2 and 3 respectively						
Odds Ratio for Response (X1)	1.6	1.497				
Regression Coefficients for % Reduction in Bald Area (X2)	10	8.5				
Hazard Ratio for Time to Complete Baldness (X3)	0.6	1.491				
Curated Sample selecting 200, 125 and 75 simulated subjects in treatments 1, 2 and 3 respectively						
Similar side-by-side assessment to the one above for the first curated sample						

- Analytical assessment of variance across weighting schemes.
- Simulation of Matching using PS, and the conditional and unconditional weights are being planned.
- Sample size assessment under variable ratio matching using Lachin (2008) for Conditional Logistic regression for binary data are at:

https://resourcetepee.com/free-statistical-calculators/observational-data/sample-size-for-matched-case-control-and-cohort-studies/

Final Notes on Observational Data

- Data with limitations. Considerable Missing data and Treatment bias issues.
- Improvements may be possible through some standards, which do not increase administrative overload at clinics, or lead to the performance of unneeded procedures.
- In regulatory setting may be useful for rare disease indications, support for accelerated approvals as an add on comparator to single arm interventional trials, and possibly for secondary label approvals.
- Pre-specified pathway through data for analyses to be credible, while specifying sensitivity analyses arounds variants on estimands and methods
- Blinding of analyst to outcome while decisions are being made about analyses
- DONE! Some balding sneetches are waiting!

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.
- Roderick J.A. Little and Doanld B. Rubin (2002). Statistical Analysis with Missing Data. 2nd edition. John Wiley and Sons, Inc., Hoboken, New Jersey.
- 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
- Austin PC. Statistical Criteria for Selecting the Optimal Number of Untreated Subjects Matched to Each Treated Subject When Using Many-to-One Matching on the Propensity Score. Am J Epidemiol. 2010;172(9), 1092-1097.
- 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.
- Lachin JM. Sample size evaluation for a multiply matched case-control study using the score test from a conditional logistic (discrete Cox PH) regression model. Stat Med. 2008;27(14), 2509-2523.
- National Research Council. (2010). The Prevention and Treatment of Missing Data in Clinical Trials. Panel on Handling Missing Data in Clinical Trials. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
- Imai K, Ratkovic M. Covariate balancing propensity score. J. R. Statist. Soc. B (2014) 76, Part 1, pp. 243–263
- The Sneetches and Other Stories. Dr Suess. Random House, 1961.

Hats of Various Sizes and Colors: Questions

Acknowledgements:

Various Statisticians and programmers in Medical Affairs Multiple Myeloma at Celgene assisted with the slides and the analysis in this slide deck.

Feedback on analyses of observational data from the medical affairs clinical team at Celgene, observational data study steering committees and the data science group is acknowledged.

The support of the Celgene statistics department and it's leadership is much appreciated.