Propensity Score Methods Using SAS®

R. Scott Leslie, MPH
MedImpact Healthcare Systems, Inc.
San Diego, CA
scott.leslie@medimpact.com

Observational Research

- Key strength: estimate effect of exposures or treatment in “real world” conditions
- Advantages
  - Data readily available, inexpensive
  - Generate quick results
  - Results more generalizable than controlled trials
  - Offer solution to limitations of RCT
    - Ethics
    - Feasibility- Costs/resources
    - Time- results lag
    - Hawthorne effect
    - External validity- patient mix
    - Small samples

Observational Research

- Key limitation: comparison groups not randomized
- Consequence: biased estimate of treatment
- Disadvantages
  - Lack of randomization
  - Differential selection- leads to differences in observed and unobserved characteristics
  - Heterogeneity of populations
  - Varying statistical analyses

Bias in Observational Studies

- Is outcome due to treatment? Or other factors?
- Limited by bias
  - Selection bias
  - Confounding
  - Reverse causality

- Selection bias
  - General definition by Rothman- a distortion resulting from the manner in which subjects are selected into the study population
  - Specified by Faries- differential probability of an individual assigned to a treatment condition and the characteristics of that individual are confounded with treatment outcomes
  - Overt (observed) and hidden (unobserved)

Rothman KJ, Greenland S. Modern Epidemiology, 3rd Edition. Lippincott Williams & Wilkins. 2008.
Faries et al, Analysis of Observational Health Care Data Using SAS, SAS Institute. 2010.

Patients eligible for Drug A or Drug B

Drug A

Patient Characteristics
Younger, More females,
Taking less medications,
Poorer medication use behavior

Drug B

Patient Characteristics
Older, More male,
Taking more medications,
Better medication use behavior

Guidance on Observational Research

- Good practices for observational studies
  - The International Society of Pharmacoepidemiology (ISPE)
  - International Society for Pharmacoeconomics and Outcomes Research (ISPOR)
- Methods for CER reviews
  - The Effective Health Care Program working document. Effective Healthcare Methods Guide for CER Reviews
  - Guide on evaluating quality CER - The GRACE Initiative (Good Research for Comparative Effectiveness)
- Guidelines on reporting observational CER studies
  - The STROBE (Strengthening the Reporting of Observational Studies) Guidelines
- Guidelines when working with patient registries – AHRQ, Registries for Evaluating Patient Outcomes: A User's Guide
- Guidelines on Systematic Reviews - Cochrane Handbook for Systematic Reviews of Interventions

Dreyer, *Epidemiology*, 2011
Sturmer, *Epidemiology*,2011

Methods to Estimate Effects

- Design stage
  - Match subjects
  - Exclusion and inclusion criteria
- Analysis stage- Use statistical techniques
  - Regression, ANCOVA, propensity scoring

- Goal = balance groups on characteristics
  - → mimic randomization or simulate random treatment assignment , “quasi randomization”
  - → more confident stating outcome is due to treatment vs. explained by other factors

D’Agostino Sr., Medical Care, 1995

Propensity Score Methods as a Potential Fix

- Traditional techniques (e.g., regression adjustment) may be limited if using too few covariates in adjustment process
- Propensity score techniques avoids limitation
  - Summarizes covariate information into a single score
- Editorial by D'Agostino (Jr. and Sr.) in JAMA
  - Use 2 methods to adjust for group differences
    - Propensity scoring- balance groups
    - Analysis of covariance- add precision

D'Agostino RB Jr & Sr, JAMA, 2007

What is the Propensity Score?

- The propensity score is the conditional probability of being treated based on individual covariates
  - Rosenbaum and Rubin demonstrated p scores can account for imbalances in treatment groups and reduce bias by resembling randomization of subjects into treatment groups

- Propensity score techniques used to compare groups while adjusting for group differences
  - Regression adjustment
  - Matching
  - Stratification (subclassification)

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.

Creating Propensity Scores Using PROC LOGISTIC

- Logistic regression: Used to predict probability of event occurring as a function of independent variables (continuous and/or dichotomous)

- Logistic model:

$$P(Y) = \frac{1}{1 + e^{-(\alpha + \Sigma \beta_i X_i)}}$$

- Propensity scores created using PROC LOGISTIC or PROC GENMOD
  - The propensity score is the conditional probability of each patient receiving a particular treatment based on pre-treatment variables
  - Creates data set with predicted probabilities as a variable
  - Or use logit of p score log (1/1-p)

Creating Propensity Scores: Code

```
proc logistic data = indsn;
class naive0;
model tx (event='Drug A') = age female b_hmo
   pre_drug_cnt_subset naive0 pre_refill_pct
   copay_idxdrug pre_sulf pre_htn pre_asthma
   pre_pain pre_lipo pre_depress
        /link=logit rsquare;
output out = psdataset pred = ps
   xbeta=logit_ps;
run;
```

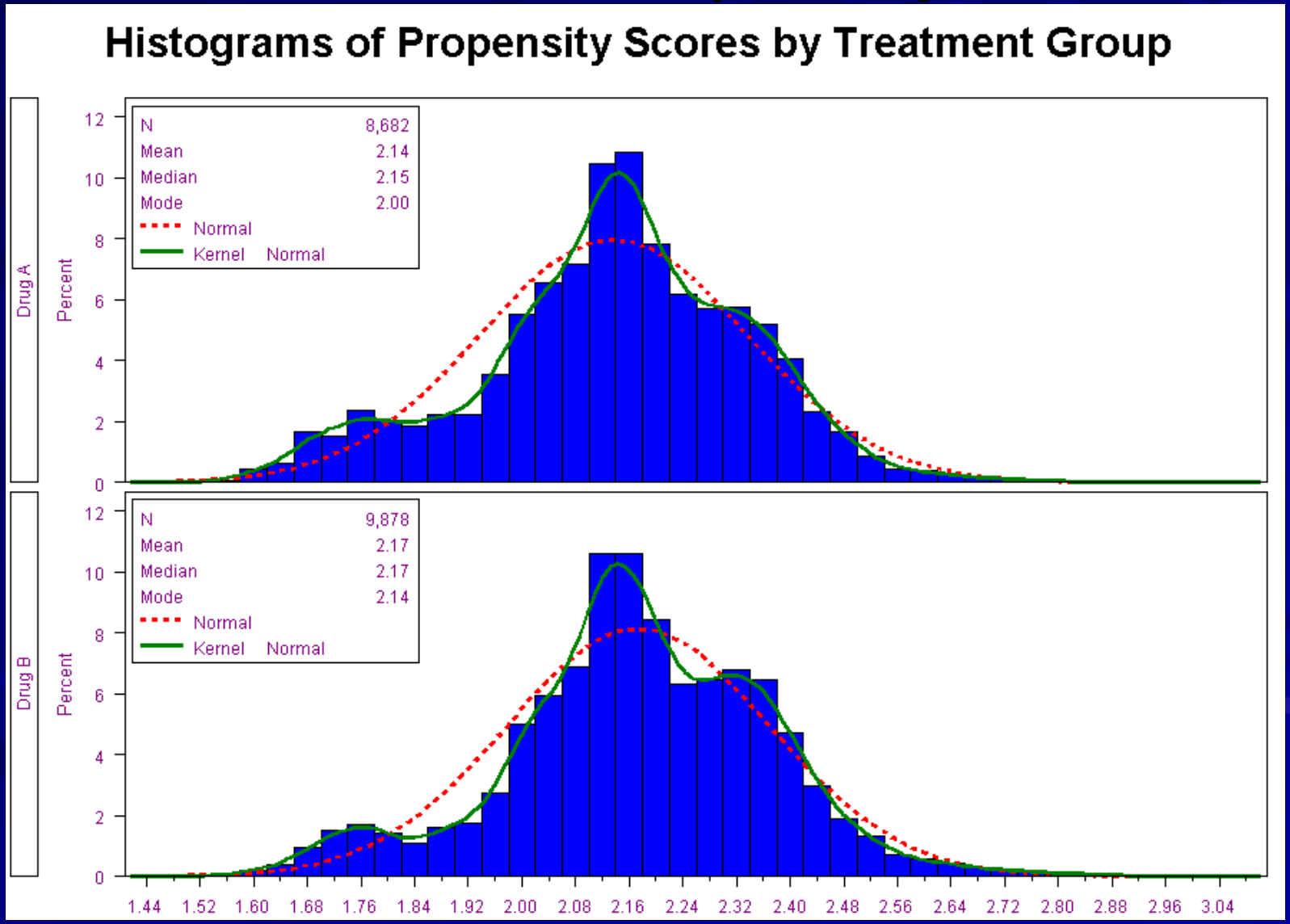
PS= predicted event probability of receiving treatment based on specified factors

Estimating P Scores

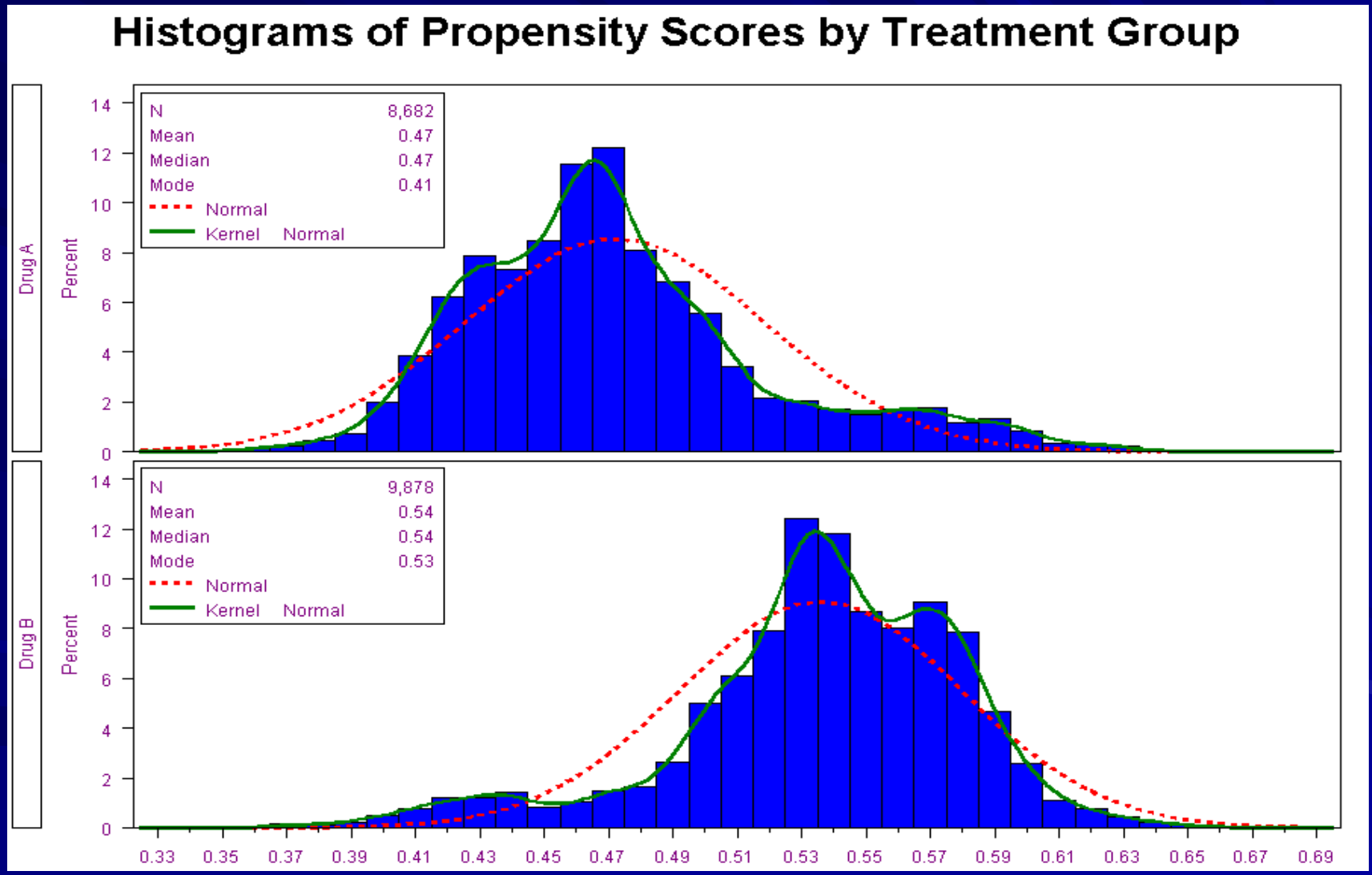
- Propensity score is the conditional probability of each patient receiving a particular treatment based on pre-treatment variables
  - More covariates better than less (Austin, 2007)
  - Include characteristics that are unbalanced b/w treatment groups
  - Success: Did it balance treatment groups?
  - Michael Doherty SAS paper/macro

Rosenbaum P, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983:70:41-55

Distribution of Propensity Scores



Distribution of Propensity Scores



Distribution of P Scores: Code

```
proc univariate data=psds plot;
title 'Histograms of Propensity Scores by Treatment Group';
var ps;
class tx;
histogram ps / ctext=purple cfill=blue
    kernel (k=normal color=green w=3 l=1)
    normal (color = red w=3 l= 2)
    ncols=1 nrows=2;
inset n='N' (comma6.0) mean='Mean' (6.2)
    median='Median' (6.2) mode='Mode'(6.2)
    normal kernel(type) / position=NW;
run;
```

Propensity Score Methods

- Once the propensity score is calculated what to do you with them?

- 3 common methods as stated by Rosenbaum and Rubin, 1984
  - Regression adjustment
  - Stratification (subclassification)
  - Matching

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.

Regression Adjustment Method

- Use p score as a covariate outcome model
  - Or use propensity score weights as the inverse of propensity score
- Use PROC GLM or PROC LOGISTIC to model outcome
  - Add independent variables believed to confound outcome
- Second step of 2 stage process
  1. Use propensity scores to balance groups
  2. Use ANCOVA modeling to create precisions

Regression Adjustment: Code

- Model continuous outcome adjusting for p scores

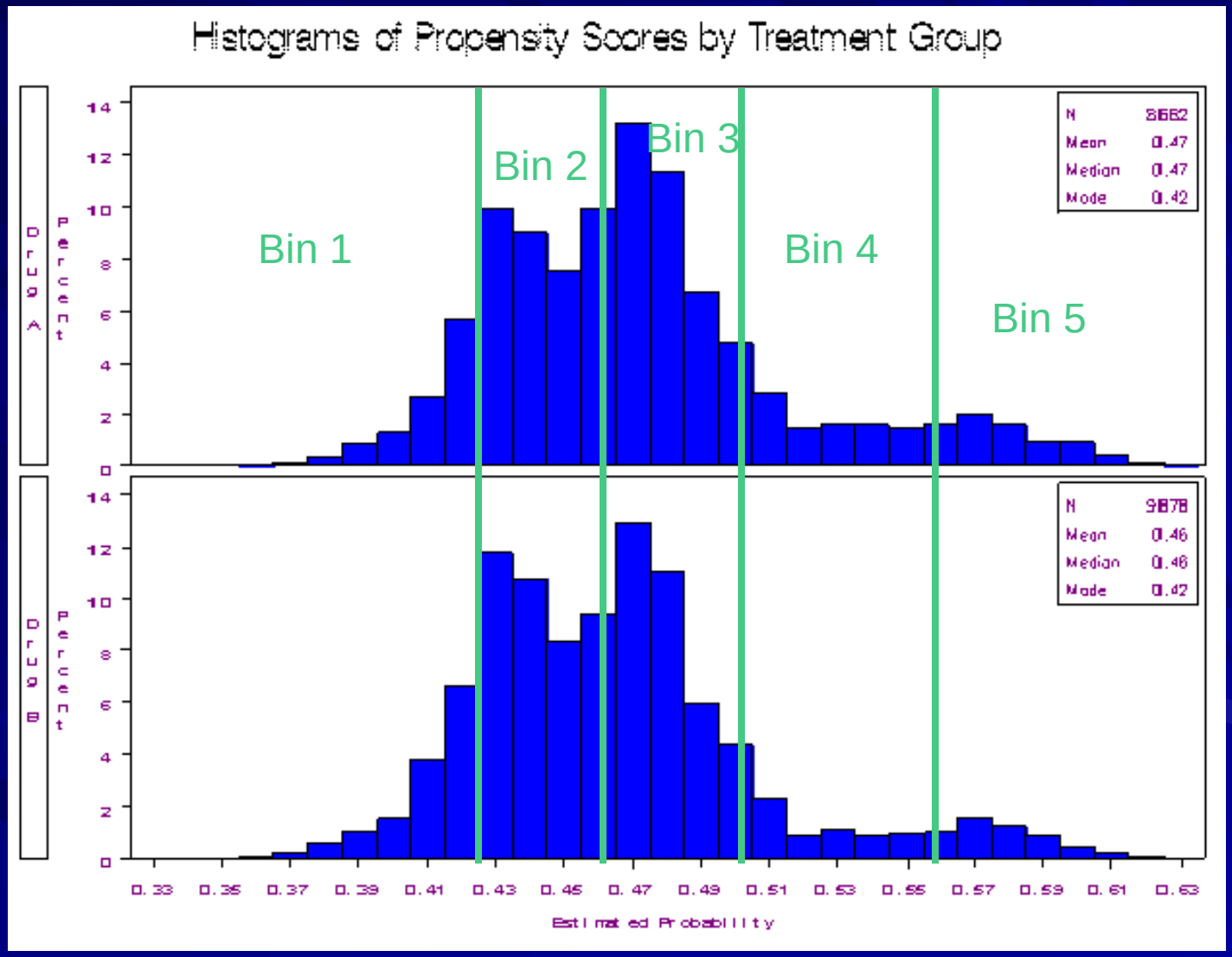
```
/*create p score*/
proc logistic data = indsn;
class naive0;
model tx (event='Drug A') = /*pre_tx_vars*/  ivar1 ivar2
    /link=logit rsquare;
output out = ps_dataset  pred = ps xbeta=logit_ps;
run;

/*outcome model adjusting for p score*/
proc glm data = ps_dataset ;
class tx;
model pdc = tx ps /solution;
lsmeans tx / om adjust = tukey pdiff cl;
quit;
```

Stratifying by P Score: Objective

- Stratification, subclassification or binning involves grouping subjects into strata based on subject's observed characteristics
- Use calculated p scores to place subjects into strata
- Objective = subjects in the same stratum are similar in the characteristics used in the propensity score development process

Stratifying Propensity Scores



Cochran, Biometrics, 1968 - 5 strata can remove 90% of the bias

Stratifying by P Score: Code

/*create 5 quintiles of p scores */

```
proc rank data = psdataset groups=5
   out = rank_ds;
ranks rank;
var ps;

data quintile;
set rank_ds;
quintile = rank + 1;
run;
```

Stratifying by P Score: Estimate Effect

- Result of code is 5 bins of homogenous subjects
  - Check differences between treatment groups
  - Sensitivity analysis if distributions don’t overlap
- Outcomes can be compared within the 5 subclasses
- Calculate weighted mean of the subclasses to report an overall treatment effect

Stratifying by P Score: Code

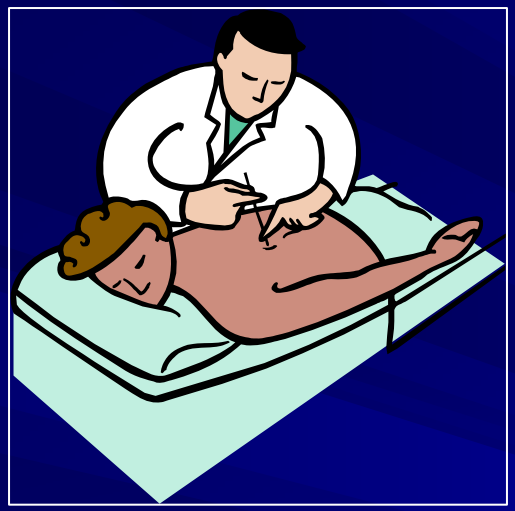
- Model continuous outcome adjusting for p scores

```
/*outcome model adjusting for quintile of p score*/
proc glm data = quintile;
class tx;
model pdc = tx quintile /solution;
lsmeans tx / om adjust = tukey pdiff cl;
quit;
```

P Score Matching

- Matching groups by p scores can balance groups on covariates
- Subjects are matched by single score vs. by one or more variables (traditional direct matching)
- Challenges
  - incomplete matching (can’t find a match)
  - inexact matching (how close is a match)

Matching by Characteristic



cases → controls

treated → untreated

intervention → non intervention

- Match on single or multiple characteristics
  - e.g., age, gender, disease severity, health plan, etc.
- 1:1 or 1:many

P Score Matching

I need a match! Does anyone have a propensity score near 0.824?

My propensity score is 0.859. Is that close enough?

P Score Matching Methods

- Techniques
  - Stratified
  - Nearest neighbor
  - Radius matching
  - Caliber
  - Kernal matching
  - Mahalanobis metric
- Replacement- back in pool for further possible matching
- W/o replacement or greedy algorithm- find match and keep it
- Which is appropriate? Literature offers some guide
  - With replacement when matching pool is small
  - 2 to 1 match if control group is large
  - Ease of calculation
- Goal- Increase balance between groups

Baser, Value *in Health*, 2006; Austin, Biometrical Journal, 2009

PS Matching Using Greedy Algorithm

- Example of case-control match using a greedy matching algorithm
- Nearest available pair method
- Reducing the non matches and inexact matches
- P scores used to balance treated and untreated groups

Parsons, Lori. 2000. “Using SAS® Software to Perform a Case Control Match on Propensity Score in an Observational Study”. *Proceedings of the Twenty-Fifth Annual SAS Users Group International Conference*, Indianapolis, IN, 214-26.

PS Matching Using Greedy Algorithm

Table 1: Original Population

	Early Intervention N (%)	Conservative N (%)	p-value
Total Patients	**2,402**	**17,735**	
Age (Mean±sd)	61.3 ±12.2	68.2±13.0	<0.0001
Male Gender	1,744 (72.6)	10,914 (61.5)	<0.0001
White Race	2,079 (91.8)	15,002 (88.4)	<0.0001
Hx Angina	444 (18.5)	4,441 (25.0)	<0.0001
Hx MI	574 (23.9)	5,382 (30.3)	<0.0001

Table 2: Greedy 5 to 1 Digit Matched Population

	Early Intervention N (%)	Conservative N (%)	p-value
Total Patients	**2,036**	**2,036**	
Age (Mean±sd)	61.9 ±12.0	61.7±13.3	0.5405
Male Gender	1,452 (71.3)	1,445 (71.0)	0.8087
White Race	1,865 (91.6)	1,858 (91.3)	0.6952
Hx Angina	390 (19.2)	381 (18.7)	0.7189
Hx MI	488 (24.0)	491 (24.1)	0.9124

Summary

- Propensity score as the conditional probability of treatment (or desired event) summarizes observed values into a single score
- P scores uses:
  - Match subjects
  - Stratify subjects
  - As a covariate
- Purpose = balancing groups to remove bias when assessing treatment effect on outcomes

Advantages

- Summarizes observed values into a single score less sensitive to model misspecification
  - Traditional techniques may be limited if accounting for only a few covariates
- P scores can diagnose comparability of groups before modeling stage
  - Distributions overlap?
    - If comparison groups are too different >>>difficult to balance groups
- P score is more robust approach
  - Address selection bias and offers precision

Assumptions & Disadvantages

- Assumptions
  - All covariates that affect both treatment and outcome must be included in the model. How do you determine this?
  - All patients have a non zero probability of receiving each treatment

- Disadvantages
  - Incorporates observed characteristics and thus doesn't account for unobserved factors, e.g., patient attitudes, socioeconomic status, and education level
    - Modified if unobserved factors are correlated to observed factors
  - Large samples sizes may be needed to establish adequate variance in covariate distributions

Conclusion

- Selection bias may create biased estimate of your outcome in observational studies
- P score methods used to adjust for selection bias
- Use with traditional risk adjustment techniques to reduce bias and better describe the effect of exposure on outcomes
- Minimizes bias, not total adjustment
- Observables vs. unobservables: Instrumental variable method account for unobservables
- Use multiple methods and consistent results add robustness of research

Questions and Comments

- Thank you, BASUG and BASUG Officers
- Special thanks to,
  - Bridget Neville
  - Karen Olsen

References

- Austin PC. Some methods of propensity-score matching had superior performance to others: results of an empirical investigation and Monte Carlo simulation. Biometrical Journal. 2009;51:171-184.
- Baser O. Too much ado about propensity score models? Comparison of types of propensity score matching. *Value in Health*. 2006; 9(6):677-385.
- Cochran WG. The effectiveness of adjustment b subclassification in removing bias in observational studies. *Biometrics*. 1968; 24:295-313.
- D'Agostino RB Sr, Kwan H. Measuring effectiveness: what to expect without a randomized control group. *Medical Care*. 1995;195:33 (4 suppl): AS95-AS105.
- D'Agostino RB, Jr, D'Agostino RB, Sr. Estimating treatment effects using observational data. *JAMA*. 2007;297 (3): 314-316.
- D'Agostino RB. Tutorial on Biostatistics: Propensity Score Methods for Bias Reduction in the comparison of a treatment to a non-randomized control group. *Statistics in Medicine*. 1998;17:2265-2281.
- Doherty M. Automating the process of choosing among highly correlated covariates for multivariable logistic regression. Proceedings of the 2008 Western Users of SAS Software Conference, Los Angeles, CA.
- Dreyer NA. Making observational studies count. Epidemiology. 2011; 22(3):295-297.
- Glynn RJ, Schneeweiss S, Sturmer T. Indications for propensity scores and review of their use in pharmacoepidemiology. *Basic & Clinical Pharmacology & Toxicology.* 2006, 98, 253–259.
- Faries, Douglas, Andrew C. Leon, Josep Maria Haro and Robert L. Obenchain. 2010. Analysis of Observational Health Care Data Using SAS®.Cary, NC: SAS Institute Inc.

References cont.

- Parsons, Lori. 2000. “Using SAS® Software to Perform a Case Control Match on Propensity Score in an Observational Study”. *Proceedings of the Twenty-Fifth Annual SAS Users Group International Conference,* Indianapolis, IN, 214-26.
- Rosenbaum PR, Rubin DB, The central role of the propensity score in observational studies for causal effects. Biometrika. 1983;70:41-55.
- Rosenbaum P.R. and Rubin D.B. 1984. “Reducing bias in observational studies using subclassification on the propensity score. J Am Stat Assoc. 1984;79:516-524.
- Rothman KJ, Greenland S. Modern Epidemiology, 3rd Edition. Lippincott Williams & Wilkins. 2008.
- Schlesselman JJ. Case-Control Studies: Design, Conduct, and Analysis. New York, Oxford University Press, 1982
- Shah BR , Laupacis A, Hux JE, Austin PC. Propensity score methods gave similar results to traditional regression modeling in observational studies: a systematic review. J. Clin. Epidemiol. 2005;58: 550–559.
- Stürmer T, Joshi M, Glynn RJ, Avorn J, Rothman K, Schneeweiss S. A review of the application of propensity score methods yielded increased use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. J. Clin. Epidemiol. 2006
- Sturmer T, Funk MJ, Poole C, Brookhart MA. Nonexperimental comparative effectiveness research using linked healthcare databases. *Epidemiology*. 2011; 22(3):298-301.

SAS References

- SAS Institute Inc. 2004. “SAS Procedures: The LOGISTIC Procedure”. *SAS OnlineDoc® 9.1.3.* Cary, NC: SAS Institute Inc. http://support.sas.com/documentation/onlinedoc/91pdf/sasdoc_913/base_proc_8977_new.pdf

- SAS Institute Inc. 2004. “SAS Procedures: The GLM Procedure”. *SAS OnlineDoc® 9.1.3.* Cary, NC: SAS Institute Inc. http://support.sas.com/documentation/onlinedoc/91pdf/sasdoc_913/base_proc_8977_new.pdf