

Paper 366-2008

The Use of Propensity Scores and Instrumental Variable Methods to Adjust For Treatment Selection Bias

R. Scott Leslie, MedImpact Healthcare Systems, Inc., San Diego, CA
Hassan Ghomrawi, MedImpact Healthcare Systems, Inc., San Diego, CA

ABSTRACT

Observational studies that lack randomization of subjects into treatment groups must address selection bias to properly estimate the effect of treatment. This paper explores a propensity scoring (PS) method and instrumental variable (IV) method of controlling for treatment selection bias. Included is a discussion of the methods, the advantages and disadvantages of the two methods as well as a comparison of results when applied to a study that evaluates the effect of therapy regimen on medication adherence. The LOGISTIC procedure is used to create propensity scores and the QLIM procedure in SAS/ETS[®] is used to conduct instrumental variable analysis.

INTRODUCTION

Among the strengths of observational studies is the ability to estimate treatment effect in real world conditions. On the contrary, a limitation of observational studies is the lack of treatment assignment. Non randomized groups usually differ in observed and unobserved characteristics resulting in differential selection into treatment groups causing selection bias when evaluating the effect of treatment. Regression adjustment, matching, and stratification using propensity scores are widely used techniques to compare groups, usually comparing a treatment group to a non treatment group. Instrumental variable analysis is the standard method used to control for selection bias in economic circles. The purpose of this paper is to explain and contrast methods using an example of diabetic members taking two different therapy regimens. Treatment groups were compared by medication adherence defined as the proportion of days of medication coverage over a 1-year period.

DESCRIPTION OF STUDY

In this study, we identified 19,433 patients using oral antidiabetic therapy from a large pharmacy claims database. Patients were categorized into two drug treatment groups, A and B, with the main objective of comparing compliance and adherence rates. Compliance was measured as the proportion of days a medication was supplied over a 180 day period. Adherent patients were identified as those reaching a threshold of 80% compliance. Selection bias is believed to be a factor as the two drug treatment groups differ in patient tolerance, adverse events, and side effects which possibly influence compliance to each drug. Other variables controlled for in the analyses include demographic variables (age, gender) and previous medication use/patterns that were measured in a 6 month baseline period prior to treatment. Previous medication use was recorded by the use of specific cardiovascular, asthma, and antidepressant medications and previous medication pattern use was measured by refill patterns of maintenance type medications.

Table 1 describes the two treatment groups. PROC GLM was used to compare groups. Few covariates significantly differed.

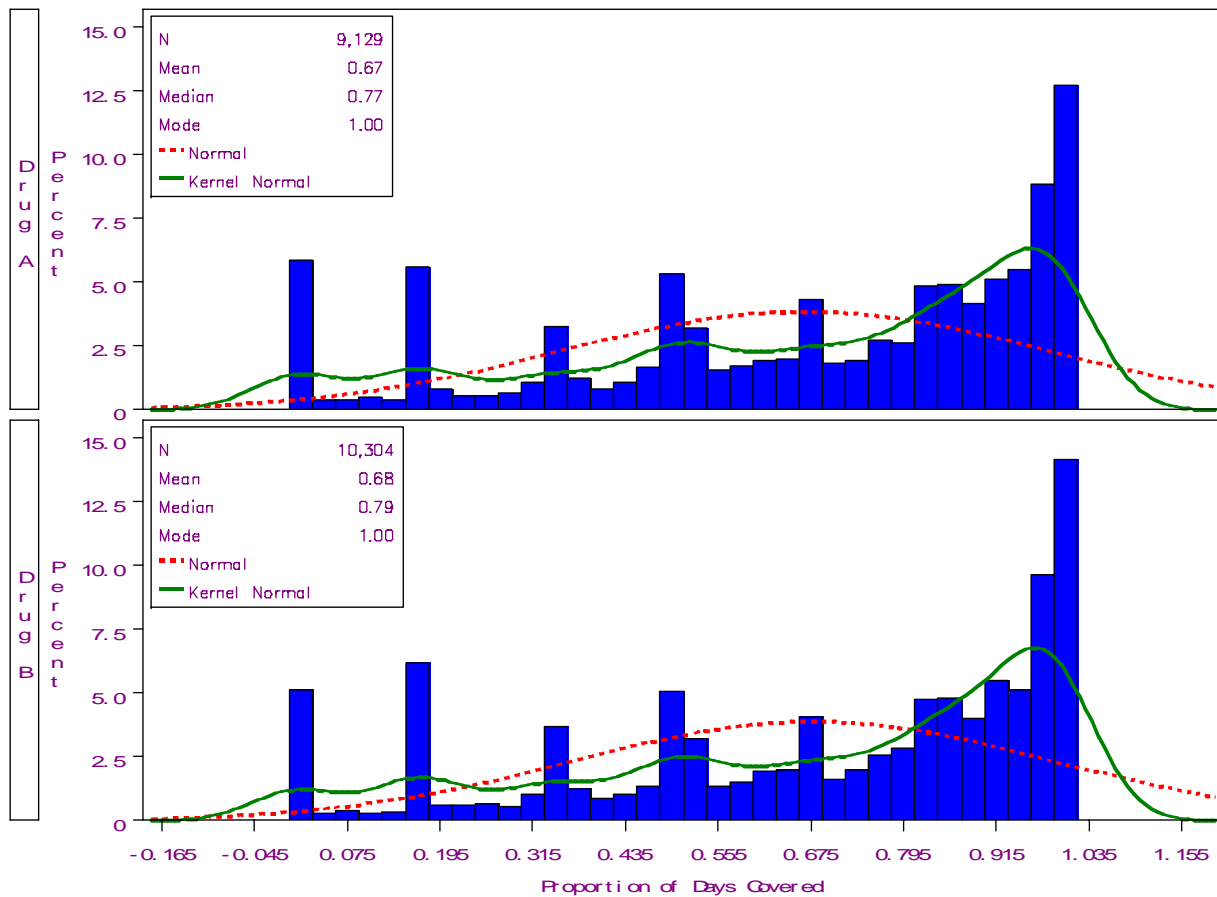
Table 1. Unadjusted Demographic and Baseline Measures

UNADJUSTED VALUES		Drug A	Drug B
Member Count		9,129	10,304
Age	Mean	57.7	57.5
	SD	10.5	10.7
Female	%	43.1%	43.4%
HMO	%	73.3%	75.3%
# of Drugs Utilized	Mean	4.9	5.0
	SD	3.0	3.1
Maintenance Medication Refill *	%	59.6%	60.5%
Prior Use *			
Naive/Naive	%	8.8	10.1
Experienced/Experienced	%	77.5	80.4
Naive/Experienced User	%	13.7	9.5

UNADJUSTED VALUES		Drug A	Drug B
Sulfonylurea *	%	55.3%	51.3%
Hypertension	%	79.1%	78.4%
Lipid Irregularity	%	68.3%	68.5%
Pain Management *	%	23.6%	24.9%
Antidepressant	%	19.5%	18.4%
Asthma	%	9.3%	9.9%

* $p < .05$

Below is a distribution of the outcomes for the two treatment groups. Mean compliance for drug A and drug B was 0.67 and 0.68, respectively. Although statistically different ($p = 0.0071$), the difference between groups is less than 1 percentage point and the distributions are very similar. Since compliance was measured as the proportion of days the medication was supplied over the 180 day period, the maximum allowable value was 1.



PROPSENSITY SCORE ANALYSIS

The propensity score is the conditional probability of each patient receiving a particular treatment based on pre-treatment variables. Using the LOGISTIC procedure, propensity scores were calculated based on the covariates listed in Table 1. The objective was to balance the treatment groups so to reduce bias of treatment selection and obtain better idea of treatment effect on the outcome of compliance. The logit function is specified in the LINK option to fit the binary logit model and the RSQUARE option assesses the amount of variation explained by the independent variables. The propensity score is output to data set named "psdataset". The predicted probabilities are output to a variable named "ps".

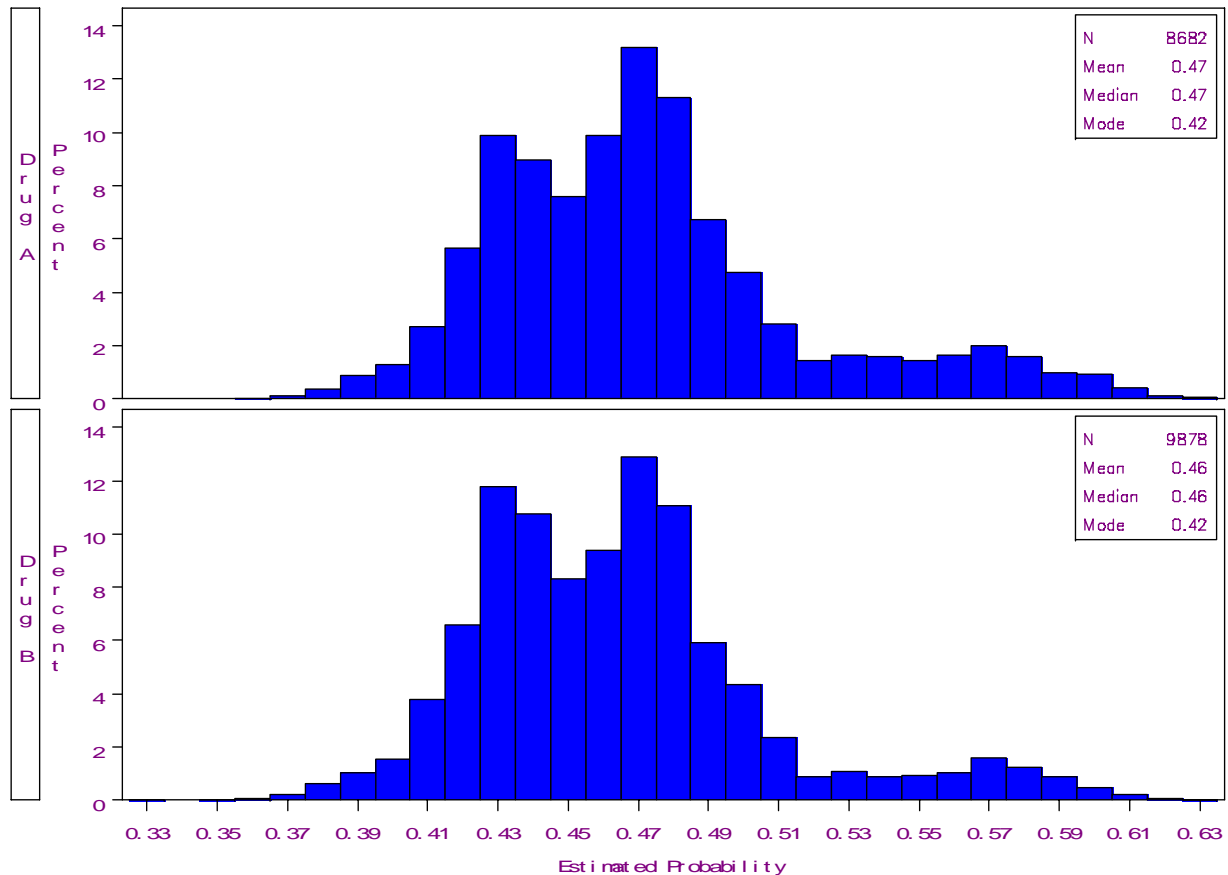
```

proc logistic data = wuss;
class preuser;
model tx (event=first) = age female pre_drug_cnt_subset preuser maintrefillratio
      copay_idxdrug pre_sulf _0106 _0109 _0112 _0113 _0149
/link=logit rsquare;
output out=psdataset pred=ps;
format preuser $preuser.;
run;

```

After creating the propensity scores, an evaluation of the distributions by treatment group checks for sizeable overlap among the groups demonstrating that the groups are comparable.

Histograms of Propensity Scores by Treatment Group



A propensity score weight, also referred to as the inverse probability of treatment weight (IPTW), is calculated as the inverse of the propensity score. The treatment selection model above modeled the propensity to receive drug a. For those patients receiving drug b, the propensity score would be $1 - ps$ and the propensity score weight would be the inverse of $1 - ps$.

```

data psdataset;
set psdataset;
if druga=1 then ps_weight=1/ps;else ps_weight=1/(1-ps);
run;

```

PROPENSITY SCORE WEIGHTED OUTCOME MODEL

Next, a propensity score-weighted linear regression model was fitted to compare drug treatment on the outcome of compliance while controlling for other covariates. The GLM procedure is used to take advantage of continuous and categorical covariates. The LSMEANS statement computes the least-squares means for the treatment variable allowing for multiple comparisons. The ADJUST=TUKEY option uses the Tukey-Kramer method to adjust the least-square means and the PDIF and CL options give the p values and corresponding confidence intervals for the differences in the least-squares means.

```
proc glm data=psdataset;
class tx preuser;
model p_dayscovered = tx age female copay_idxdrug pre_drug_cnt_subset preuser
                    maintrefillratio pre_sulf _0106 _0109 _0112 _0113 _0149
/solution;
lsmeans tx/OM ADJUST=TUKEY PDIF CL;
weight ps_weight;
format preuser $preuser.;
quit;
```

Results of the model above showed no difference between treatment groups ($p = 0.2066$). Compared to the unadjusted compliance means (a model with compliance as the dependent variable and drug treatment as the only independent variable) shows the effect of controlling for confounding. The increase in compliance from the unadjusted means (0.6673 and 0.6793) to the adjusted means (0.7032 and 0.7082) was probably due to the large number of patients with prior use of the drug. These patients had higher compliance values than those new to therapy and therefore when this variable is placed in the model as a covariate the mean compliance values rise.

Table 2. Mean Compliance by Treatment: Unadjusted vs. PS Adjusted Models

Model	Treatment	Compliance Outcome	95% Confidence Limits	
Unadjusted Model P = 0.0071	Drug A	0.6673	0.6601	0.6736
	Drug B	0.6793	0.6733	0.6852
Propensity Score Adjusted Model P = 0.2066	Drug A	0.7032	0.6975	0.7089
	Drug B	0.7082	0.7029	0.7135

INSTRUMENT VARIABLE ANALYSIS

Instrument variable (IV) analysis is commonplace in economics and social sciences. As opposed to traditional risk adjustment methods that rely on observable measures, IV methods factor in unmeasured or unobserved factors as the source of confounding. The goal of IV analyses is to find instruments that are correlated with treatment selection but are not directly correlated with the outcome variable. When variables are found, the IV creates variance to estimate the effect of treatment on the outcome. A special issue of Health Service Research (*HSR* 2000: 35 [5]) has an overview of IV by McClellan and Newhouse.

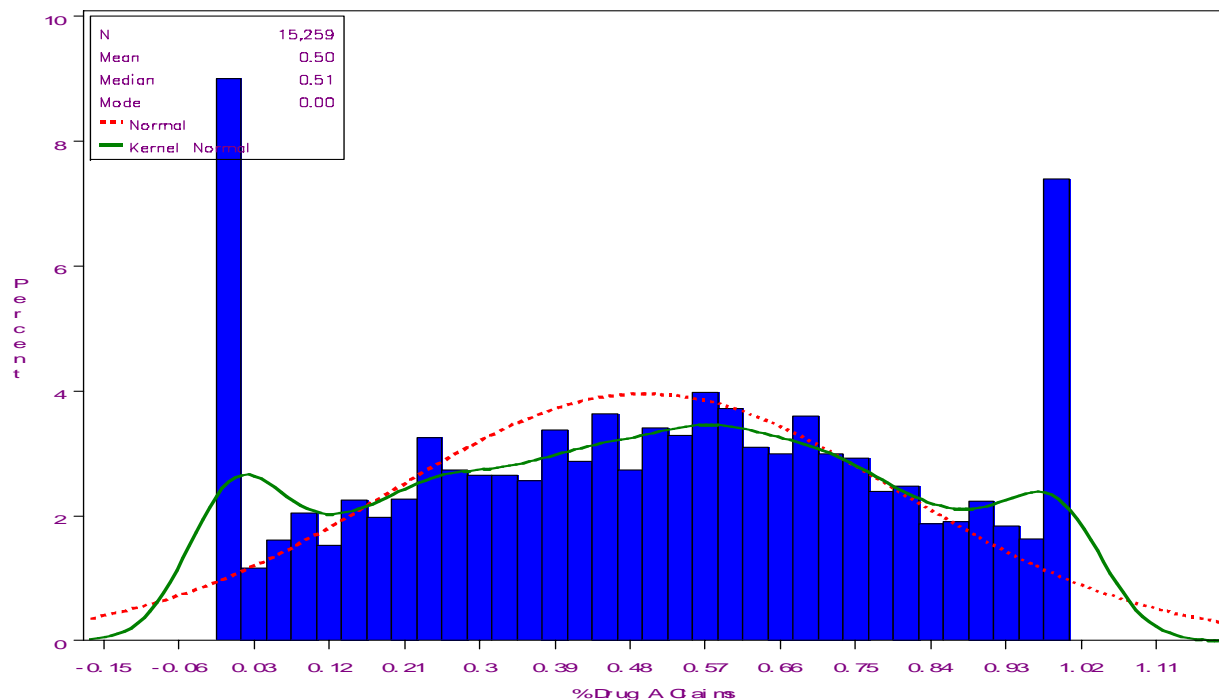
For the study example presented here, prescriber patterns in the preceding year was used as the IV because it was hypothesized that prescribers with a tendency to prescribe drug A will continue to prescribe drug A. It was measured by the percentage of claims prescribed for drug A (# of claims for drug A / # of claims for drug A and B) in the preceding year. Each prescriber was identified from the first prescription for each patient. Claims for these prescribers for the two drugs were extracted to calculate the proportion of claims for drug A. Some prescribers were not identified correctly and some prescribers had no history of prescribing either drug and therefore these patients were excluded from the analysis.

Testing the IV first used a LOGISTIC procedure to prove that it was predictive of treatment. There was a very large relationship between previous prescribing patterns (IV) and the drug that they prescribed to the patient (treatment selection). Second, prescriber behavior was not correlated with the outcome of compliance. That is, prescriber's tendency or preference to prescribe drug A or drug B is not a predictor of compliance. Using the CORR procedure, the correlation between the IV and compliance was slightly negative ($Rho = -0.0071$) and insignificant ($p = 0.3796$).

A two stage IV process first uses instrument variables and other covariates to predict the treatment. A second stage estimates the outcome by the predicted treatment (from the first model) and other covariates (Angrist, J., G. Imbens, and D. Rubin). This two stage approach has the advantage of incorporating the predicted treatment into the outcome model as it represents the portion of treatment selection related to prescriber patterns.

Distribution of the IV shows adequate variance.

Distribution of Instrument Variable



USING PROC QLIM IN SAS ETS®

The two stage IV process can be done in one step using the QLIM procedure in SAS ETS®. SAS ETS®, Econometrics Times Series, module is useful in analyzing time series data (data with processes that take place over time). It is widely used in economic analysis and financial modeling. The QLIM (Qualitative and Limited Dependent Model) can analyze models where the dependent variable has limited values. This fits this example as the dependent variable is choice of two drugs.

```
proc qlim data = finall;
class preuser;
model druga = pct_druga age female copay_idxdrug pre_drug_cnt_subset preuser
      maintrefillratio pre_sulf _0106 _0109 _0112 _0113 _0149 /discrete;

model p_dayscovered = age female copay_idxdrug pre_drug_cnt_subset preuser
      maintrefillratio pre_sulf _0106 _0109 _0112 _0113 _0149
      /select(druga=0);
output out=druga prob proball predicted;
run;
```

The first MODEL statement is a selection equation that uses the probit model and creates the predicted probability of each subject receiving treatment. The second MODEL statement is the outcome equation which uses regression to model the outcome. This assesses the effect of the treatment group while controlling for the probability produced from the first equation. Output includes the following table of parameter estimates. Of most importance are the treatment selection/IV parameter (druga.pct_druga) and the correlation parameter (_Rho). The druga.pct_druga parameter estimate indicates a strong effect of the IV on treatment selection ($p < .0001$) and the _rho parameter estimate indicates that treatment selection bias had no effect on the outcome.

Table 3. Parameter Estimates from QLIM Procedure

Parameter Estimates					
Parameter		Estimate	Standard Error	t Value	Approx Pr > t
p_dayscovered.Intercept		0.574566	0.022066	26.04	<.0001
p_dayscovered.age		0.001004	0.000323	3.11	0.0019
p_dayscovered.female		-0.040598	0.006630	-6.12	<.0001
p_dayscovered.copay_idxdrug		-0.000979	0.000127	-7.73	<.0001
p_dayscovered.pre_drug_cnt_subset		-0.001481	0.001226	-1.21	0.2271
p_dayscovered.preuser	Naive/Naive	-0.013398	0.016848	-0.80	0.4265
p_dayscovered.preuser	Experienced/Experienced	0.109298	0.009766	11.19	<.0001
p_dayscovered.preuser	Naive/Experienced	0	.	.	.
p_dayscovered.maintrefillratio		0.053654	0.010917	4.91	<.0001
p_dayscovered.pre_sulf		-0.004019	0.006537	-0.61	0.5388
p_dayscovered._0106		0.009691	0.008293	1.17	0.2426
p_dayscovered._0109		0.028269	0.007202	3.93	<.0001
p_dayscovered._0112		-0.048821	0.008057	-6.06	<.0001
p_dayscovered._0113		-0.004750	0.008412	-0.56	0.5723
p_dayscovered._0149		0.000940	0.011086	0.08	0.9324
_Sigma.p_dayscovered		0.263242	0.002246	117.23	<.0001
druga.Intercept		-1.373643	0.084455	-16.26	<.0001
druga.pct_druga		2.983973	0.046337	64.40	<.0001
druga.age		0.000251	0.001193	0.21	0.8336
druga.female		-0.028216	0.024800	-1.14	0.2552
druga.copay_idxdrug		0.000072619	0.000468	0.16	0.8768
druga.pre_drug_cnt_subset		-0.000219	0.004531	-0.05	0.9614
druga.preuser	Naive/Naive	-0.284848	0.060922	-4.68	<.0001
druga.preuser	Experienced/Experienced	-0.214734	0.038712	-5.55	<.0001
druga.preuser	Naive/Experienced	0	.	.	.
druga.maintrefillratio		-0.069171	0.040817	-1.69	0.0901

Parameter Estimates					
Parameter		Estimate	Standard Error	t Value	Approx Pr > t
druga.pre_sulf		0.111517	0.024279	4.59	<.0001
druga._0106		0.024808	0.031059	0.80	0.4244
druga._0109		-0.072252	0.027340	-2.64	0.0082
druga._0112		-0.058125	0.029630	-1.96	0.0498
druga._0113		0.024990	0.031566	0.79	0.4286
druga._0149		-0.010035	0.041075	-0.24	0.8070
_Rho		-0.034112	0.029647	-1.15	0.2499

When this is done for each treatment, the predicted value of compliance is output for each treatment. Predicted compliance if all members were on drug A was 0.7195 and the predicted compliance if all patients were on drug B was 0.7237.

COMPARISON OF THE TWO METHODS

The main difference between the two methods is that propensity scoring uses observable measures to construct a weight based on selection where IV methods rely on an instrument variable made from unmeasured or unobserved factors. An advantage of IV is that it accounts for unmeasured factors correlated with the outcome, this is especially helpful analyzing data sets that were not created for the purposes the research question. Weakness of IV is that the instrument can be challenging to find and difficult to validate and must be agreed upon by subject matter experts.

The following table compares the results from the methods. First, unadjusted mean compliance for drug A and drug B was 0.67 and 0.68, respectively. When adjusting for covariates (Adjusted model) without using propensity scoring, the means increased as the model controls for the various factors. Next, the propensity score models evaluate the effect of selection process. Least-squares means from this model are similar to the adjusted model. This suggests that selection bias was not a factor or was not captured as it exists. The last model using IV techniques also found a lack of treatment selection effect. Limitations to study design that are present for both methods include the possibility that actual compliance may differ from observed compliance. Also, there may well be additional predictors of treatment selection and adherence that are not captured including, patient attitudes, socioeconomic status, education level, and number of other medications used.

Table 4. Mean Compliance by Treatment: Unadjusted vs. Adjusted Models

Model	Treatment	Compliance Outcome	95% Confidence Limits	
Unadjusted Model				
	Drug A	0.6673	0.6601	0.6736
	Drug B	0.6793	0.6733	0.6852
Adjusted Model				
	Drug A	0.7027	0.6970	0.7083
	Drug B	0.7072	0.7019	0.7125
Propensity Score Adjusted Model				
	Drug A	0.7032	0.6975	0.7089
	Drug B	0.7082	0.7029	0.7135
IV Adjusted Model				
	Drug A	0.7195	0.7226	0.7206
	Drug B	0.7237	0.7184	0.7249

CONCLUSION

This paper presents 2 methods commonly used to control for selection bias in observational studies. An inverse propensity score weighted method is compared to an instrument variable method. Each method has strengths and weaknesses. Although IV methodology is less known and less widely used in the health sciences, it can be used in

conjunction with traditional risk adjustment techniques, such as propensity score matching, propensity score subclassification, and multivariable logistic regression, to reduce bias and better describe the effect of treatment.

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.
- D'Agostino, R. 1998. "Tutorial on Biostatistics: Propensity Score Methods for Bias Reduction in the comparison of a treatment to a non-randomized control group". *Statistics in Medicine* 17, 2265-2281.
- Kurth, T., et al. 2006. "Results of Multivariate Logistic Regression, Propensity Matching, Propensity Adjustment, and Propensity-based Weighting under Conditions of Nonuniform Effect". *American Journal of Epidemiology* 206: 163:262-270.
- Bao, Y., Duan, N., Fox, A. 2006 "Is Some Provider Advice on Smoking Cessation Better Than No Advice? An Instrument Variable Analysis of the 2001 National Health Interview Survey". *Health Services Research* 41:6, 2114-2135
- Angrist, J., G. Imbens, and D. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91: 444-55.
- Hogan, J. et al. 2004. "Instrument Variables and Inverse Probability Weighting for Causal Inference from Longitudinal Observational Studies". *Statistical Methods in Medical Research* 13, 17-48.
- Cole, J. et al. 2006 "Drug Payment and Adherence in Chronic Heart Failure: Effect on Cost and Outcomes". *Pharmacotherapy* 26 (8), 1157-1164.
- Pasta, David J. 2000. "Using Propensity Scores to Adjust for Group Differences: Examples Comparing Alternative Surgical Methods". *Proceedings of the Twenty-Fifth Annual SAS Users Group International Conference*, Indianapolis, IN, 261-25.
- 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_91/stat_ug_7313.pdf
- SAS Institute Inc. 2004. "SAS Procedures: The QLIM Procedure". *SAS OnlineDoc*® 9.1.3. Cary, NC: SAS Institute Inc.
http://support.sas.com/documentation/onlinedoc/91pdf/sasdoc_91/stat_ug_7313.pdf

ACKNOWLEDGMENTS

The author would like to thank Hassan Ghomrawi for his expertise in instrument variable analysis and his review of the paper.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

R. Scott Leslie
MedImpact Healthcare Systems, Inc.
10680 Treena Street
San Diego, CA 92131
Work Phone: 858-790-6685
Fax: 858-689-1799
Email: scott.leslie@medimpact.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. © indicates USA registration.

Other brand and product names are trademarks of their respective companies.