

PROPENSITY SCORE BASED METHODS IN DETERMINING
THE SAFETY OF LUMBAR PUNCTURE IN COMATOSE
MALAWIAN CHILDREN

By

Lei Zhao

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Biostatistics—Master of Science

2015

ABSTRACT

PROPENSITY SCORE BASED METHODS IN DETERMINING THE SAFETY OF LUMBAR PUNCTURE IN COMATOSE MALAWIAN CHILDREN

By

Lei Zhao

Coma is a common clinical presentation for critically ill children in sub-Saharan Africa, where major differential diagnoses include cerebral malaria, viral encephalitis, and bacterial and tuberculous meningitis. Lumbar puncture (LP) is a crucial diagnostic test to distinguish between these etiologies and determine optimal treatment. However, clinicians may be hesitant to perform an LP due to concerns of precipitating herniation and death, particularly in environments where pre-procedure neuroimaging is unavailable. We performed a retrospective cohort study of the safety of LP in comatose Malawian pediatric inpatients recruited over consecutive rainy seasons from 1999-2013. Our goal is to assess whether performing an LP changed the mortality within the first 12 hours after the admission and the overall mortality during the hospitalization. Following propensity score matching, all baseline characteristics were balanced between matched pairs of children who did and did not have LPs. After matching, both 12 hour mortality and overall mortality were not significantly different between children who did and did not receive lumbar punctures. Logistic regression, stratification on propensity scores, and inverse probability weighting analyses all showed that the overall mortality has a statistically significant decrease in the group that received an LP. The results of 12-hour mortality from these methods are consistent with the matching method.

ACKNOWLEDGMENTS

This research was supported by University of Liverpool, Liverpool, Michigan State University, East Lansing and Malawi-Liverpool-Wellcome Trust, Blantyre, Malawi. This project involved huge amount of work and research including data collection, confirmation, management and analysis. I am very grateful to the help from many individuals and organizations. We would like to express gratitude to all of them.

First of all I am thankful to Dr. Chris Moxon, University of Liverpool, and Dr. Douglas Postels, Michigan State University for organizing and planning the ideas and designs of the study.

I am also very grateful to the support and instruction from my mentor Dr. Chenxi Li, who has been providing strong statistical expertise and confirming analysis results. I thank him for being the chair in my committee and providing invaluable comments to the earlier version of the manuscript.

I also want to thank Dr. Zhehui Luo and Dr. David Todem for being as my committee members. They provided insight and expertise patiently that assisted me to finish the analysis.

In the end, thanks to the support from my family and all my dear friends. Thanks to all of you for letting me become a Spartan.

TABLE OF CONTENTS

LIST OF TABLES..	v
LIST OF FIGURES	vi
1. Introduction...	1
1. 1 Study background...	1
1. 2 Confounding effect	2
2. Study Design...	4
2. 1 Treatments and outcomes...	4
2. 2 Confounders	4
3. Methods...	7
3.1 Logistic regression and propensity score...	7
3.2 Inverse probability weighting	8
3.3 Stratification...	9
3.4 Matching...	10
4. Result...	13
4.1 Descriptive statistics...	13
4.2 Impact of lumbar puncture on overall mortality	15
4.3 Impact of lumbar puncture 12-hour mortality	21
5. Conclusion and discussion...	22
REFERENCES...	24

LIST OF TABLES

Table 1 Summary statistics of all the categorical confounders... ..	5
Table 2 Summary statistics of all the continuous confounders... ..	6
Table 3 Summary statistics of propensity score in treated and untreated... ..	13
Table 4 the distribution of 12-hour mortality by LP... ..	15
Table 5 the distribution of overall mortality by LP... ..	15
Table 6 Differences of probabilities of overall death in each stratum	18
Table 7 Overall average treatment effects on overall mortality	19
Table 8 Covariates Balance table before matching... ..	20
Table 9 Covariates Balance table after matching... ..	20

LIST OF FIGURES

Figure 1 Directed Acyclic Graphs (DAG)	3
Figure 2 Distribution of propensity score by treatment... ..	14
Figure 3 Propensity score and corresponding weight... ..	16

1. Introduction

1. 1 Study background

In hospitals in sub-Saharan African presentation of a child in coma is a frequent occurrence. There are differential diagnosis including cerebral malaria, bacterial meningitis, tuberculosis meningitis, cryptococcal meningitis, viral encephalitis and drug intoxication. Lumber puncture (LP) is a critical diagnostic tool and maybe the only way to distinguish between some of these pathologies. However there is considerable controversy as to whether LPs are safe in children with decreased consciousness.

More than three thousand patients were included from a retrospective cohort study of the safety of LP in comatose Malawian pediatric inpatients from 1999-2013. Among these patients, study population with a size of nine-hundred and thirty three was selected based on:

- 1) The subjects must have a severe coma. In the data, we selected those who have coma score equal to or less than 2. In the record, the higher the score is, the more conscious the subject is.
- 2) The subject must have key covariates recorded. Clinicians believe that there are factors that would affect the decision of performing lumber puncture. We require these factors to have been recorded so that we can control for them. Otherwise, biased estimates would be obtained for the effect of lumbar puncture.

Purpose of the study is to see whether in a comatose child performing an LP increases the probability of death in the first 12 hours after admission and during the overall time period after correcting for other confounders. In the study, we know that the decision to perform lumbar puncture on the subjects is affected by many factors, such

as age, blood pressure, weight and height that also affect mortality. To estimate the causal effect of treatment, propensity score based methods in observational study were performed to control the effects of these confounding variables.

1. 2 Confounding effect

In a statistical model, confounding factors or confounders are those extraneous variables that both have correlation with dependent variable and independent variable. According to Hernán & Robins ^[1], confounding is the bias that appears when the treatment and the outcome share a common cause and any variable that can be used to help eliminate confounding bias is a cofounder. If we don't account for confounders in the model, the relationship between dependent variable and independent variables will be estimated incorrectly.

In equation (1), we let T be the independent variables, Y be the independent variable and $P(y|do(t))$ stands for the probability of event Y under the hypothetical intervention $T=t$.

$$P(y|do(t)) = P(y|t) \quad (1)$$

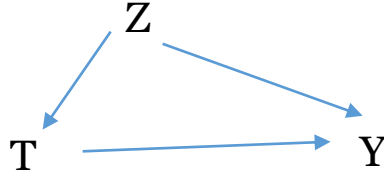
This equation clearly holds when there is no confounding factors, which is that what we observe is what really happens. However, in many observational studies, the equation doesn't hold because of confounders.

Suppose ideally we have a set of all confounders denoted as Z . This Z meets three conditions ^[2]:

- 1) It is associated with treatment.
- 2) Conditional on treatment, it is associated with outcome.
- 3) It doesn't lie on the pathway between treatment and outcome.

This definition can be explained in graph theory by directed acyclic graphs (DAG). Pearl [3] proposes the back-door criterion for nonparametric identification of causal effects, which is the effect of T on Y can be identifiable if all backdoor paths between treatment and outcome can be block. Back-door criterion can be a proper tool to determine the set of Z. In Figure 1, because the risk factors T and outcome Y share a common cause Z and there is a back door path between T and Y. This path can be blocked by controlling on Z. Then according to the three conditions and the graph, Z is a set of confounders and needs to be adjusted for.

Figure 1 Directed Acyclic Graphs (DAG)



X stands for risk factors, Y stands for outcome and Z stands for a set of variables, which are confounders in this graph.

In our study, T refers to lumbar puncture, Y refers to two kinds of mortality and Z refers to all the confounders, which are explained in details in the next section.

To get the unbiased estimate of $P(y|do(t))$, we consider equation (2)

$$P(y|do(t)) = \sum P(y|t, z_i)P(z_i) \quad (2)$$

2. Study Design

2. 1 Treatments and outcomes

In the study, cerebrospinal fluid opening pressure was used to determine whether a subject was performed the lumbar puncture or not. The majority of children have LPs performed either shortly before or shortly after admission to the research ward but, importantly, a proportion of children have not, generally for one of 5 reasons:

- 1) The clinician was concerned that the child was not stable enough to tolerate an LP the reasons for which include shock, severe respiratory distress or intractable seizures
- 2) The clinician identified papilledema on retinal examination
- 3) LP was attempted but failed for technical reasons
- 4) The parent did not consent to LP
- 5) The child regained consciousness before LP was performed.

For outcome, 12-hour mortality and overall mortality during hospitalization were taken into consideration to evaluate the treatment effects.

2. 2 Confounders

In the model setup, variables including gender, age, comatose score, blood pressure, normal heart test, weight-height Z score, respiratory disease, pulse status, blood glucose on admission, retinopathy present, plasma lactate on admission and papilledema. Among these variables, weight-height Z score was created by weight and height according to the WHO Child Growth Standards [4]. Respiratory disease was determined from sigh of grunting, deep breath and normal chest exam: if any of the

three is abnormal, then it suggests that the respiratory disease is present. Table 1 and Table 2 separately give a brief view of descriptive statistics of all the categorical and continuous confounders in the nine hundred and thirty three study subjects. The numbers in the parentheses are the statistics in treated group. We can learn that some variables have very different summary statistics and these variables are unbalanced in treated and untreated right now. So some statistical methods needed to be performed to balance the variables to lead us unbiased estimates.

Table 1 Summary statistics of all the categorical confounders

Categorical Variable	Value	Total Numbers (Treated)	Percent (Treated)	P-value from testing difference
Normal heart test	Normal	888 (688)	95.18 (95.56)	0.34
	Abnormal	45 (32)	4.82 (4.44)	
Respiratory disease	Not present	595 (458)	63.77 (63.61)	0.78
	Present	338 (262)	36.23 (36.39)	
Gender	Male	447 (344)	47.91 (47.78)	0.79
	Female	486 (376)	52.09 (52.22)	
Pulse Status	Low	12 (10)	1.29 (1.39)	0.25
	Normal	346 (258)	37.08 (35.83)	
	High	575 (452)	61.63 (62.78)	
Blood pressure	Low	12 (7)	1.29 (0.97)	0.63
	Normal	901 (701)	96.57 (97.36)	
	High	20 (12)	2.14 (1.67)	
Weight-height score	low weight-for-age	87 (59)	9.32 (8.19)	0.02
	Normal	846 (661)	90.68 (91.81)	
Coma score	Most severe	119 (86)	12.75 (11.94)	0.44
	Severe	366 (287)	39.23 (39.86)	
	Less Severe	448 (347)	48.02 (48.19)	
Retinopathy present	Not present	288 (237)	30.87 (32.92)	0.02
	Present	645 (483)	69.13 (67.08)	
papilledema	Not present	735 (600)	78.78 (83.33)	<0.0001
	Present	198 (120)	21.22 (16.67)	

Table 2 Summary statistics of all the continuous confounders

LP	Variable	Count	Mean	Standard Deviation	P-value from testing difference
Untreated	blood glucose on admission	215	6.27	3.38	0.24
	plasma lactate on admission		10.50	2.69	0.60
	age		41.88	29.12	0.13
Treated	blood glucose on admission	718	6.61	3.69	
	plasma lactate on admission		10.61	2.56	
	age		45.21	28.06	

3. Methods

3.1 Logistic regression and propensity score

As we discussed in the previous section, there will be 12 confounders that needed to be adjusted for. In this study, propensity score based methods are well suited to estimate the treatment effect. The technique was first published by Paul Rosenbaum and Donald Rubin in 1983 [5]. Propensity score estimation is a statistical technique that attempts to estimate the effect of a treatment by accounting for the covariates that predict receiving the treatment. It aims to decrease the bias caused by the confounders found in the study especially when there are limited amounts of observations but many confounders. It has been used in health care research, public policy investigation and other areas. This method is useful in scenarios when confounders contain many variables or may be continuous, because it will be hard to adjust for potentially high-dimensional confounder with common techniques.

Rosenbaum & Rubin showed that if we set the equation (3):

$$p(z) = P(T = 1|Z) \quad (3)$$

Then Z and T are independent given $p(z)$ and in this situation $p(z)$ is called propensity score. So a large set of confounders can be reduced to a number between 0 and 1. This relationship between treatment and confounders can be denoted in (4)

$$(T \perp Z)|p(T = 1|z) \quad (4)$$

Thus, the next thinking will be how to calculate the propensity score. We know that our treatment is a binary variable and logistics regression is a common technique to deal with binary outcome. This model was developed by D. R. Cox in 1958^[6]. If we define that given a linear combination of Z , $F(Z)$ is the probability that the dependent variable equals. In our case, we can use this regression model to get the probability of

receiving a treatment given a linear combination all the confounders, as shown in equation (5)

$$\ln \frac{F(Z_i)}{1-F(Z_i)} = \ln \frac{P(T=1|Z_i)}{1-P(T=1|Z_i)} = \beta_0 + \beta_i Z_i \quad (5)$$

Further, after we get the propensity score, we can use it as a covariate in addition to the treatment factor in a multivariate regression for the outcome of interest.

However, Rubin suggested several advice when one performs regression adjustment method:[7]

- 1) The difference in means of propensity score in the two groups should be small, unless the situation is following: (a) nearly the same variance of the covariates in the two groups, (b) symmetric distribution of covariates in the two groups, (c) sample size are nearly the same in two groups.
- 2) The ratio of the variances of the propensity score in the two groups must be closed to one.
- 3) After adjusting for the propensity score, the ratio of the variances of the residuals of the covariates should be close to one.

3.2 Inverse probability weighting

A statistical technique for calculating standardized statistics to a population different from that in which we collect the data can is call inverse probability weighting. In observational study, the sampling probability is usually different over subjects, so we need to use this probability to weight the observations [8]. Using the propensity score, inverse probability weighting creates a weighted population, in which the distribution of the covariates is independent with the treatment assignment.

In our study, we use the weighting method that the inverse of propensity score is the weight for each treated subject and the inverse of (1-propensity score) is the weight for each untreated subject. Let A_i be the indicator denoting whether subject i is treated or not, we have the weight for subject i ,

$$w_i = \frac{A_i}{p(T=1|z_i)} + \frac{1-A_i}{1-p(T=1|z_i)} \quad (6)$$

We define $E(Y^1)$ as the counterfactual mean, which is the outcome that would have been observed if we had intervened by setting the treatment value equal to 1. We want to show that the IP weighting mean is equivalent with counterfactual mean ^[1]. The IP weighing mean is defined as $E\left[\frac{Y}{p(T=1|z_i)}\right] = E\left[\frac{Y}{p(z_i)}\right]$, then (7)

$$\begin{aligned} E\left[\frac{Y}{p(z_i)}\right] &= E\left[\frac{Y^1}{p(z_i)}\right] \\ &= E\left\{E\left[\frac{Y^1}{p(z_i)}|Z_i\right]\right\} \\ &= E\left\{E\left[\frac{1}{p(z_i)}|Z_i\right]E[Y^1|Z_i]\right\} \quad (7) \end{aligned}$$

We know $E\left[\frac{1}{p(z_i)}|Z_i\right]=1$, so (7) becomes

$$E\{E[Y^1|Z_i]\} = E[Y^1]$$

The same proof can be done for $E(Y^0)$. In this way, our interest $E(Y^1) - E(Y^0)$ can be estimated.

3.3 Stratification

In stratification method, the common way is to create intervals according to the scale of propensity score. All the subjects are ranked based on their propensity scores.

Usually the number of strata depends on the sample size and the proportion of the

overlap between the propensity scores of treatment and non-treatment [8].

Rosenbaum and Rubin claimed that stratification on propensity score could remove 90% of the bias due to measured confounders using 5 strata. A reduction in bias should appear when one increase total numbers of strata. Within each strata, the treated and untreated should have a nearly similar values of propensity score. That is to say, the covariates are distributed similarly in two groups in same strata. Then the treatment effects are estimated in each interval or strata with assigned weigh and a weighted average of the effects will give an overall estimate of the treatment effects. The weight is commonly determined by the fraction of the treated subjects in each interval. In this study, we define the stratum-specific weight: [9]

$$w_i = \frac{1}{se_i^2}$$

$$ATE = \frac{\sum w_i(\bar{Y}_{1i} - \bar{Y}_{0i})}{\sum w_i} \quad Var(ATE) = \frac{1}{\sum w_i} \quad (8)$$

where se_i^2 is the variance of the difference between means in ith strata. So by this definition, we can get the estimate of average treatment effect and the variance of it as shown in (8).

3.4 Matching

From (4), for a given propensity score, exposure to treatment is random and treated and untreated subjects should be on average observationally identical. Matching is another way to adjust the confounding effects. The goal is that we want to build a population in which Z variables have the same distribution in treated and untreated. According to the definition of propensity score, matching on the variables Z is

equivalent to matching on propensity score. Propensity score matching reduces the dimensionality of matching to a single dimension.

In general, there are several ways to match each treated to one or more untreated: nearest neighbor matching, caliper matching, Mahalanobis metric matching and exact matching. In this study, we will use nearest neighbor matching (NNM) based on the PS distance, shown in equation (9)

$$D_{ij} = |p_i - p_j| \quad (9)$$

NN match treated and untreated subjects taking each treated subject and searching for the untreated subject with the closest propensity score. Usually a subject can be a best match for more than one subject in another group. The advantage of NNM is that it can be simply performed. However, even if the propensity scores from two subjects are very close, they can still have very different distributions, especially in some key confounders. To ensure the quality of matching, balance of the covariates will be checked after matching. We will discuss more about this in the result section. A propensity score estimator for average treatment effect can be defined as:

$$\hat{\tau} = \frac{1}{N} \sum_{i=1}^N (2T_i - 1) \left(Y_i - \frac{1}{M} \sum_{j \in \Lambda_M(i)} Y_j \right) \quad (10)$$

Where M is the total numbers of matches for one subject and $\Lambda_M(i)$ is the set of matches for subject i. It can be defined as:

$$\Lambda_M(i) = \{ j = 1, \dots, N : T_j = 1 - T_i, \left(\sum_{k: T_k = 1 - T_i} I_{|p(Z_i) - p(Z_k)| \leq |p(Z_i) - p(Z_j)|} \right) \leq M \} \quad (11)$$

The variance estimator is:

$$\hat{\sigma}_{\tau, adj}^2 = \hat{\sigma}_{\tau}^2 + \hat{c}_{\tau}' \hat{V}_{\gamma} \hat{c}_{\tau} \quad (12)$$

Where \hat{V}_γ is the treatment model coefficient variance- covariance matrix, \hat{c}_τ is the term used to adjust the variance based on the covariance between confounders and the outcome. The details of the expression are in Abadie, A., and G. W. Imbens 2012^[10].

And

$$\begin{aligned} \hat{\sigma}_\tau^2 = & \frac{1}{N} \sum_{i=1}^N ((2T_i - 1)(Y_i - \frac{1}{M} \sum_{j \in \Lambda_M(i)} Y_j) - \hat{\tau})^2 \\ & + \frac{1}{N} \sum_{i=1}^N ((\frac{K_i}{M})^2 + \frac{2M-1}{M} (\frac{K_i}{M})) \hat{\sigma}^2(T_i, p(Z_i)) \end{aligned} \quad (13)$$

In which we define K_i is the number of times that subject i is used as a match.

4. Result

4.1 Descriptive statistics

Based on logistic regression, we put lumbar puncture as the dependent variable and the 12 confounders as independent variables. Table 3 displays the descriptive statistics of propensity scores in two groups.

Table 3 Summary statistics of propensity score in LP and Non-LP

Estimated Probability of Propensity Scores					
LP	N	Mean	Std Dev	Minimum	Maximum
0	215	0.7224326	0.1241450	0.3492940	0.9034260
1	718	0.7838356	0.0923648	0.3119707	0.9240525

In the original data, there were more than three thousand subjects recorded in the clinical center in Malawian. Only half of the population has been included in the study because others present missing values in confounders, which lead to missing values in propensity scores. Also the means of propensity score in two groups are not large so that logistic adjustment regression will likely work.

Figure 1 displays the distribution of propensity score by treatment. The mean propensity score in treated is 0.78 with SD 0.09 and in untreated is 0.72 with SD 0.12. The upper part in the figure shows the distribution of propensity scores in the untreated and the lowers shows the distribution in treated group. Visually, we can say there is a good overlap of propensity scores in distributions among treated and untreated.

Figure 2 Distribution of propensity score by treatment

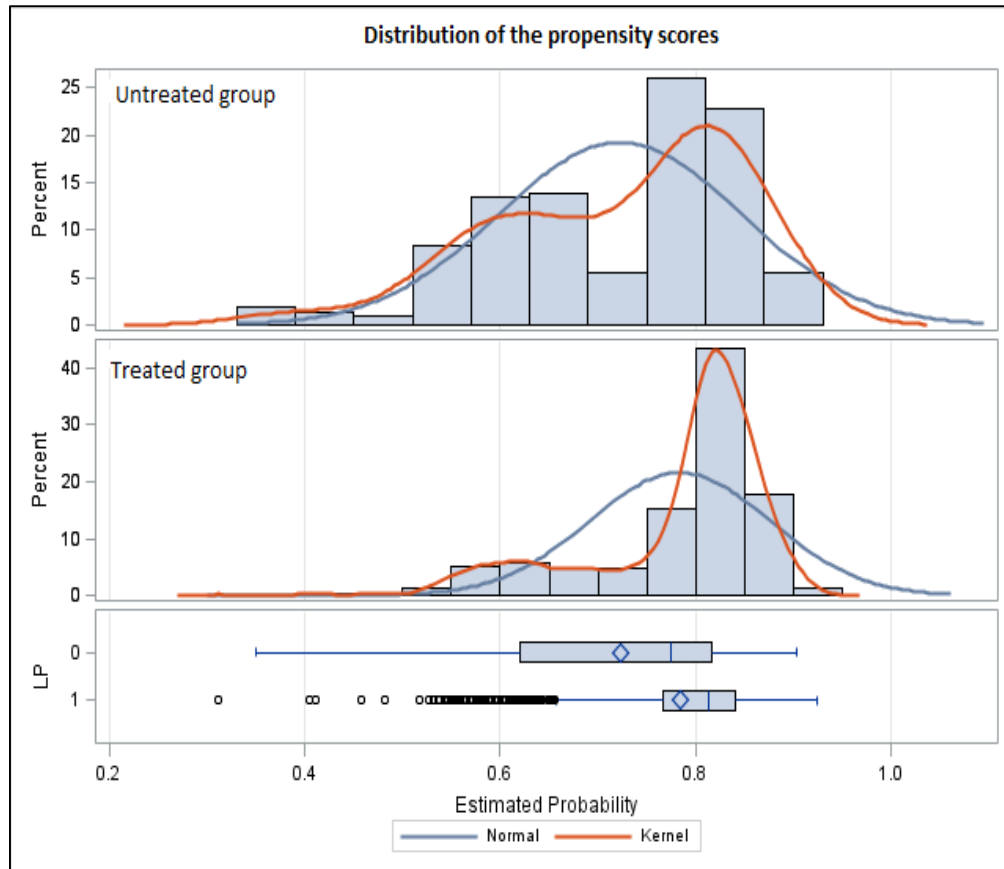


Table 4 and Table 5 display the distribution of two outcomes by treatment. 126 subjects died in the first 12 hours and 138 died in the overall time. Among them, 18.1% LPs died in the first 12 hours and 12.1% non-LPs died in the 12 hours. Based on these two tables, we were able to compute the unadjusted treatment effects of lumbar puncture on both 12 hour mortality and overall mortality. LPs lowered the 12 hour mortality by 0.08 (p-value 0.0234) and by 0.10 (p-value 0.0002). For both outcomes, LPs had significant protective effects and details are displayed in Table 4 and Table 5.

Table 4 the distribution of 12-hour mortality by LP

Table of the distribution of 12-hour mortality by LP			
	12-hour mortality		
LP	0	1	total
0	179 (81.9%)	39 (18.1%)	215
1	631 (90%)	87 (10%)	718
Total	807	126	933
Unadjusted treatment effect -0.06 with 95% CI(-0.11, -0.01)			

Table 5 the distribution of overall mortality by LP

Table of the distribution of overall mortality by LP			
	overall mortality		
LP	0	1	total
0	166 (77.3%)	49 (22.7%)	215
1	629 (87.6%)	89 (12.4%)	718
Total	795	138	933
Unadjusted treatment effect -0.10 with 95% CI(-0.16, -0.05)			

4.2 Impact of lumbar puncture on overall mortality

Before we performed propensity score methods to analyze the treatment effect, a basic logistic regression model was used to evaluate in the interest. LPs lower the overall mortality by 0.06 with confidence interval (0.01, 0.12) and p-value 0.02.

Analysis under inverse probability weighting:

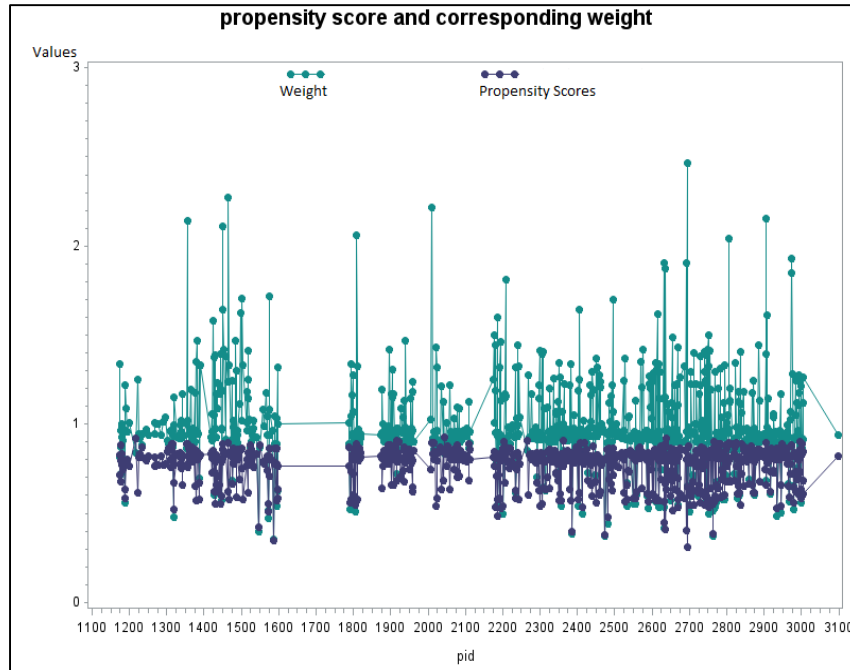
From the inverse probability weighting method we mentioned in the previous section, the goal of it is to create weighted population in which covariates and treatment are independent. We can make sure that in the weighted population the probability of receiving a treatment is independent from confounders and we achieve this by

assigning treatment with same probability for everyone. With the weight $W_i = \frac{1}{p(T=1|z_i)}$, we create a population in which subjects have a probability equal to 1 to receive treatment (the same for untreated). However, we can also assign different probabilities as long as the treatment and confounders are independent. This is called stabilized weight, shown in equation (14), where $P(T)$ is the observed probability of receiving treatment and i and j refer to treated and untreated [1].

$$W_i = \frac{P(T)}{p(T=1|z_i)} \quad W_j = \frac{1-P(T)}{1-p(T=1|z_j)} \quad (14)$$

Figure 3 displays the relationship between propensity score estimates and the corresponding weight. Because we use the stabilized weight, the expected mean of weight is 1.

Figure 3 Propensity score and corresponding weight



With the weighted population, we get the estimate of odds ratio for lumbar puncture on overall mortality. After we obtained the estimates of parameters, we were able to express the result in predicted probability, the probability to die in the overall time is 0.186 with CI (0.139, 0.243) for the untreated and the probability to die for the treated is 0.131 with CI (0.108, 0.157). To get the point estimate for the variance of the difference, we use bootstrap to find the standard error of the difference in predicted probability. We sample 933 subjects from our study population for 500 times. In each sample, we obtain one estimate for the difference in predicted probabilities. We define the difference in original study population is \widehat{D} and the one from i th sample is \widehat{D}_i . Then the standard error is:

$$s.e. = \frac{1}{500} \sum_{i=1}^{500} (\widehat{D}_i - \widehat{D})^2 \quad (15)$$

Finally we have an estimate of average difference -.055 with CI (-0.081,-0.034), P-value= 0.003. LPs had a slightly protective effect on overall mortality.

Analysis under regression adjustment:

After correcting the confounders, we got the propensity score for each subject and we took it as a covariate into the logistic model. After we obtained the estimates of parameters, we were able to express the result in predicted probability. Similarly, using bootstrap we got the estimate for treatment effect: LPs reduce the overall mortality by 0.06 with 95% confidence interval 0.01 to 0.12 (p-value=0.02). The

result is consistent with the one from IPW method: LP has a slight protective effect on overall mortality.

Analysis under stratification:

According to the propensity we got before, we created five stratum containing both patients in treatment group and non-treatment group. Each stratum was created based on the scale of propensity scores. The sample sizes in five stratum are similar with each other. In addition to this, treated and untreated subjects would have similar distributions in each stratum. We also tested the balance in each stratum through t-test for each covariate. The weight of each stratum was calculated according to the standard error in each stratum. The specific formula was introduced in the method section. In the Table 6 below, there are the differences between probabilities of death between treatment group and non-treatment group in each stratum. From the confidence interval we cannot say there is a significant difference between two groups in each stratum. After taking the weight of each stratum into consideration, Table 7 displays the overall average treatment effects and here we can conclude the treatment lowers the overall mortality by 5.54% with CI (0.2%, 10.9%).

Table 6 Differences of probabilities of overall death in each stratum

Stratum	Treatment effect in each stratum	Lower CL	Upper CL
1	-0.0991	-0.2309	0.0326
2	-0.0651	-0.1905	0.0603
3	-0.0541	-0.1747	0.0665
4	-0.0674	-0.1180	0.1070
5	0.00547	-0.1807	0.0459

Table 7 Overall average treatment effects on overall mortality

Overall treatment effect	Stander error	Lower CI	Upper CI
-0. 055493	0. 027212	-0. 10883	-0. 002156978

Analysis under matching:

To find the average treatment effect, we perform one to one matching following nearest neighbor method. Each treated unit is matched with an untreated unit based on having the closest propensity score. So in total we find 215 treated units to match all the untreated units and 718 untreated units to match all the treated. To see if the covariates distribute similarly in the treated and untreated, a covariate balance table was created with using standardized difference as a test measurement. It has been suggested that if the standardized difference is greater than 10%, there is a meaningful imbalance in covariates in two groups ^[12].

Table 8 and 9 display the balance tables before matching and after matching. We notice that in Table 8 among confounders, age weight, height and papilledema have more than 10% standardized difference, shown in (16) and (17)

For a continuous variable, ^[13]

$$d = \frac{(\bar{Z}_{treatment} - \bar{Z}_{non-treat})}{\sqrt{\frac{s_{treat}^2 + s_{non-treat}^2}{2}}} \quad (16)$$

Where \bar{Z} is the mean in each sample and S^2 is means the sample variance in each group.

For a dichotomous variable,

$$d = \frac{(\hat{p}_{treatment} - \hat{p}_{non-treat})}{\sqrt{\frac{\hat{p}_{treat}(1 - \hat{p}_{treat}) + \hat{p}_{non-treat}(1 - \hat{p}_{non-treat})}{2}}} \quad (17)$$

where \hat{p} is the mean in each sample.

After matching, these variables all have standardized differences less than 10% and most standardized differences become smaller.

Table 8 Covariates Balance table before matching

Untreated	Variable	Mean	Treated	Mean	SD
215	Age	41.8883	718	45.293	0.1190
	Sex	0.5116		0.5236	0.0240
	Coma sc	1.3209		1.3621	0.0582
	Bp stat	1.0139		1.0069	0.0335
	Resp dis	0.3534		0.3649	0.0237
	N heart	0.9395		0.9554	0.0711
	Pulse stat	1.5674		1.6142	0.0908
	Wh zscore	-1.1573		-1.058	0.0720
	Adm gluc	6.27723		6.6169	0.0958
	Papilledema	0.3628		0.1671	0.4538
	plasma lactate on admission	10.61		10.5	0.04189
	Retinopathy	0.7210		0.6708	0.20573

Table 9 Covariates Balance table after matching

Untreated	Variable	Mean	Treated	Mean	SD
933	Age	43.9346195	933	44.3772	0.0153
	Sex	0.5423365		0.5305	0.0236
	Coma sc	1.437299		1.4155	0.0309
	Bp stat	0.9989282		1.0086	0.0481
	Resp dis	0.3622722		0.3536	0.0178
	N heart	0.9442658		0.9453	0.00469
	Pulse stat	1.5344		1.5442	0.019
	Wh zscore	-1.0886007		-1.0686	0.0152
	Adm gluc	6.4026795		6.6020	0.0555
	Papilledema	0.2079314		0.2154	0.0183

Table 9 (cont'd)

Untreated	Variable	Mean	Treated	Mean	SD
933	plasma lactate on admission	10.7002	933	10.7135	0.0258
	Retinopathy	0.7325		0.7256	0.0134

With the good balanced distribution of covariates in treatment group and non-treatment group, we can implement the nearest-neighbor matching estimator for the average treatment effect. The estimate of the average treatment effect is -0.0525 with a confidence interval (-0.119, 0.143). This suggests that lumbar puncture doesn't have a significant effect on overall mortality.

4.3 Impact of lumbar puncture on 12-hour mortality

Before we performed propensity score methods to analyze the treatment effect, a basic logistic regression model was used to evaluate in the interest. LPs lower the 12 hour mortality by 0.04 with confidence interval (-0.01, 0.09) and p-value 0.15.

By using propensity score based methods, we got the estimate when outcome variable is 12-hour mortality. Treatment effect is -0.041 with confidence interval (-0.094, 0.010), p-value=0.11 in the weighed population and -0.038 with confidence interval (-0.091 0.016), p-value 0.16 in the regression adjustment method.

The average treatment effect is -0.0371 with confidence interval (-0.091, 0.016) from stratification and -0.053 with confidence interval (-0.116, 0.011), p-value=0.26 from matching. All the results are consistent showing there is no significant treatment effect on 12-hour mortality from lumbar puncture.

5. Conclusion and discussion

Randomized controlled trial is considered to be the ideal approach to get the average treatment effects. However, in observational study, subjects are not randomly assigned as the treated or untreated. In this study, the decision of performing lumbar puncture can be affected by other symptoms. Historically, researchers have been using regression adjustment to control for other factors. However, those procedures could be hard to implement when there are many factors to be adjusted. Especially some of them are often continuous variable and undefined variable. Propensity score based method is able to reduce the confounding effects by boiling down a large set of confounders to a probability. In this study, we performed inverse probability weighing, regression, stratification and matching based on propensity score. We find out that 12-hour mortality is not significantly different among those who received the lumbar puncture and who didn't receive the lumbar puncture. For overall mortality, inverse probability weighing, regression, and stratification give that there is a slight protective effect from lumbar puncture. But in matching, the average treatment effect is -0.0525 with a confidence interval (-0.119, 0.143), which is not significant. From the matching table, we achieved balanced the distribution of confounders in treatment and non-treatment. In that scenario, it is closer to an ideal random clinical trial and the estimate of the average treatment effect would be more accurate. We also compared the results from propensity score based methods and basic regression model. The results were consistent and propensity score based showed a stable and good performance in evaluation of treatment effects. When the number of covariates is large, propensity score based methods would have better performance.

However, there are limitations and pitfalls in propensity score based methods. First, all the confounders we used are from observed data and experience from clinicians and in this way those unmeasured confounders cannot be controlled. In our study, there are 12 confounders that may cause bias on our outcome. But there may be other unobserved factors that would affect the decision of lumbar puncture. We can barely do something to them. What is more, in our study, we observed massive missing values in some key confounders. Due to this, more than half subjects were excluded from the study. This may lead to decrease in power and lose of useful information.

REFERENCES

REFERENCES

1. Hernán, M. A. & Robins, J. M. (2013). Causal Inference.
2. Pearl, J., (1993). "Aspects of Graphical Models Connected With Causality," In Proceedings of the 49th Session of the International Statistical Science Institute, pp. 391 - 401.
3. Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, 82(4), 669–688.
4. WHO Multicentre Growth Reference Study Group. WHO Child Growth Standards: Length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: Methods and development. Geneva: World Health Organization, 2006 (312 pages).
5. Rosenbaum, Paul R.; Rubin, Donald B. (1983). "The Central Role of the Propensity Score in Observational Studies for Causal Effects". *Biometrika* 70 (1): 41–55. doi:10.1093/biomet/70.1.41.
6. Cox, DR (1958). "The regression analysis of binary sequences (with discussion)". *J Roy Stat Soc B* 20: 215–242.
7. Rubin, D. B. (2001). Using propensity scores to help design observational studies: application to the tobacco litigation. *Health Services & Outcomes Research Methodology*, 2, 169–188.
8. Rosenbaum P.R., Rubin D.B. Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*. 1984;79:516–524.
9. Rheta E. Lanehart, Patricia Rodriguez de Gil, Eun Sook Kim, Aarti P. Bellara, Jeffrey D. Kromrey, and Reginald S. Lee Propensity Score Analysis And Assessment Of Propensity Score Approaches Using Sas® Procedures
10. Abadie, A., and G. W. Imbens 2012. Matching on the estimated propensity score. Harvard University and National Bureau of Economic Research. <http://www.hks.harvard.edu/fs/aabadie/pscore.pdf>.
11. Robins, JM; Rotnitzky, A; Zhao, LP (1994). "Estimation of regression coefficients when some regressors are not always observed". *Journal of the American Statistical Association* 89 (427): 846 866. doi:10.1080/01621459.1994.10476818
12. AustinPC, GrootendorstP, AndersonGM. A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study. *Stat Med*. 2007;26(4):734-753.
13. Flury B.K., Riedwyl H. Standard distance in univariate and multivariate analysis. *The American Statistician*. 1986;40:249–251.