J. Appl. Environ. Biol. Sci., 7(8)170-177, 2017

ISSN: 2090-4274
Journal of Applied Environmental and Biological Sciences
www.textroad.com

Propensity Score Stratification Using Logistic Regression Bootstrap in the Case of Peripheral Diabetic Neuropathy

Bambang Widjanarko Otok*, Purhadi, Shofi Andari, Ingka Rizkyani Akolo

Laboratory of Environmental and Health Statistic, Department of Statistic,
Institut Teknologi Sepuluh Nopember (ITS), Surabaya (Indonesia)

Received: April 19, 2017
Accepted: July 2, 2017

ABSTRACT

Propensity score is the conditional probability to get a specific treatment based on the observed covariat. In the analysis of the study cohort studies, this method is used to reduce the bias in the estimation of treatment effects on the data is the observation due to confounding factors. If the treatment is binary form, logistic regression model is one of the estimation of the value of the propensity score is exactly because easily in the estimation and interpretation. The Data used in this research is the patient medical record secondary data about the case of Diabetic Neuropathy Peripheral (NDP). The purpose of this research is to examine the estimates of the propensity score based on the binary logistic regression with bootstrap and analyzing Propensity Score Stratification (PSS) bootstrap in obtaining estimates of Average Treatment Effect (ATE) in the case of Diabetic mellitus especially for NDP. The results of the estimation of the parameters with the MLE method does not close the form and done iteration Newton-Raphson, next iteration Newton-Raphson resampling results as much as B times and obtained bootstrap parameters. While the results of the bootstrap PSS analysis show that the 5 strata provide standard errors that are smaller and the largest bias reduction (89,83%) compared to other strata. In addition, from logistic regression model it is known that the status of hypertension patients, long patients suffer Diabetes Mellitus (DM) is the variables that impact directly and indirectly on the status of NDP Diabetic, while the status of obesity patients and serum HbA1C Diabetic is the variables that affect only directly to the status of NDP patients.

KEYWORDS - Bootstrap, Confounding, DM , Propensity Score Stratification, Peripheral Diabetic Neuropathy

INTRODUCTION

The one research standard is the existence of randomisasi experiments, namely lay the subject of research in treatment groups or control groups based on random. In general randomness is needed on research so that the assumption of independency are met so that the effect of the bias can be minimised. But in the field of health research involving human life random may not always practiced. In some conditions, researchers usually use design non-experimental/observation studies that in many cases are vulnerable to selection biases and the impact on the results of the estimation of treatment effects that doubted.

Propensity score (PS) is one of the methods that can reduce the bias from *confounding*. *Propensity score* is defined as a conditional probability to receive treatment based on previous characteristics [1]. There are two main things in the propensity, namely the estimation of the value of the propensity and the propensity method is used in a case of health. For data category, logistic regression model is one model that is often used for the estimation of the value of the propensity for estimation more easily than other models [2]. Some research about methods based on propensity score, [3] use PS Matching (PSM) and PS Stratification (PSS) to reduce the bias in the comparison of treatment groups and control for the case of drugs, [4] compare 4 score propensity method (PSM, PSS, covariate adjustment PS and PS Weighting) to reduce the systematic differences between treatment groups and control in the case of Smoking obtained the conclusion that the methods of community participation is the best method.

From the research above it is known that the bias that is produced from the PSS method is still large compared with other propensity method. This is because in the PSS method there is a grouping of observations into strata based on the order in which the value of the estimates of the propensity score resulted in response in each class and between classes are not free to each other and the estimation of the propensity score is unknown also affect the variance estimates to conclusions [5]. Therefore, PSS method combined with the bootstrap method for the estimation of treatment effects to minimize confounding bias produced [5]. Bootstrap introduced the first time by Efron, 1979 [6]. Bootstrap is a resampling based method sample data with the conditions of the return on the data in the complete statistics of the size of a sample with the help of a computer [6]. Advantages of the bootstrap partition is able to overcome the problems of statistics without the complex mathematical calculations, without any assumptions and only based on the existing data.

In the last few years of the disease is not transmitted to the attention in the field of health because it is one of the causes of the increasing number of death. One of the disease prevalence contagious not high enough in the world is Diabetes Mellitus (DM). In Indonesia, patients with diabetes mellitus has reached 8.4 million in 2000 and is expected to be around 21.3 million in 2030. The high number of such patients make Indonesia occupies the fourth sequence of the world after the United States, India and China [7]. Diabetes mellitus is a chronic metabolic disturbances diseases as a result of the pancreas does not

***Corresponding author:** Bambang Widjanarko Otok, Laboratory of Environmental and Health Statistic, Department of Statistic, Institut Teknologi Sepuluh Nopember (ITS), Surabaya (Indonesia). email: dr.otok.bw@gmail.com

produce enough insulin or the body cannot use the insulin produced effectively [8]. In patients with diabetes mellitus there are two types of vascular complication that may arise, namely macrovascular complications for example of arterial disease coroner and complications of microvascular i.e. diabetic neuropathy. For neuropathy is the main complications that often found in patients with diabetes mellitus especially in patients with diabetes mellitus type-2 [9]. The risks faced by patients with diabetes mellitus with neuropathy diabetic among others infection repeatedly, peptic not healed and healed and amputation finger or foot. These conditions lead to increased morbidity and mortality, patients with diabetes mellitus with neuropathy diabetic. Most neuropathy is often found in patients with diabetes mellitus who aged more than 50 years and rarely found on the age under 30 years [10].

Diabetes mellitus is a metabolic disease that is a collection of symptoms that arise in a person due to increased blood glucose levels above normal values. Diabetic neuropathy (ND) is a condition where nerves edge malfunctioning caused by damage Cellular and Molecular because DM disease [11]. Diabetic Polineuropatic describes the involvement of many nerves edge and the distribution of the general symmetric bilateral includes motoric disorders, sensory apparatus, or autonomous [12]. The number of this neuropathy increased simultaneously with the length of the suffering of Diabetes Mellitus and age patients. According to the [13][14] factors that affect the NDP is long suffering from diabetes mellitus, smoking hypertension, age, HbAIC levels, gender, increased body mass (obesity), the use of alcohol CAGE high, the serum albumin level is low in the treatment of insulin, hyperglikemia. While according to [11][15] factors that affect the NDP is glucose, lipid and amino acids, hypertension, smoking long DM, anthropometric measurement as weight, waist circumference and hip circumference and demographic information in the form of gender, age.

Research on Diabetic Neuropathy Peripheral (NDP) has many done. For example [9] examines the relationship of hyperglycemia, age and long suffering from diabetes mellitus with neuropathy diabetic events using the spearman correlation test, chi-square and multivariate binary logistic, [10] examine about positive correlation level of AUS with NDP using descriptive analysis and coefficient contingency tests and [11] examine about the prevalence and risk factors using NDP nonparametric Kruskal-Wallis tests and logistic regression.

Logistic regression modeling without attention to the possibility of a strong combination between the factors that affect the NDP, can cause the confounding variable which resulted in the conclusion is inaccurate. Therefore, this research using propensity score stratification bootstrap on the binary logistic regression to reduce the bias from confounding variable. The purpose of this research is to examine the estimates of the propensity score based on logistic regression with bootstrap and perform analysis of the bootstrap PSS in obtaining estimates of ATE in case of NDP.

METHODOLOGY

The Data used is d ata secondary form of medical records of patients with diabetes mellitus type-2 in RSUD Kabupaten Pasuruan. This research uses one response variable the NDP status (Y) and 8 variables predictors i.e. the age of patients (X_1), gender (X_2), dyslipidemia (X_3), hypertension (X_4), long DM (X_5), obesity (X_6), serum HbA1C (X_7), and the level of Acid Uric Serum (AUS) (X_8) [10][15].

The steps of data analysis in this research is as follows [5][16].

a. Descriptive statistics on the data to create graphs and tabulation of data based variables.
b. Determine confounding variable confounding variable, hereinafter dinotasikan Z with parameter θ .
c. Calculate the value of the estimates of the propensity score for the original data and the bootstrap samples with MLE method.
d. Divide the subject become 2-5 strata based on the value of the estimates of the propensity score obtained in step (c). In general the 5 strata good enough to reduce the bias by 90 percent. The division of strata performed on the original data and samples of the bootstrap partition.
e. Test whether the propensity score of the treatment group and control for each strata have the same distribution on each kovariat. If not balance, then return to step (d).
f. Calculate the value of the estimates of the average treatment effect (ATE) samples of the bootstrap partition.
g. Calculate the percent reduction in Bias (PBR) from the original data PSS and PSS bootstrap and compare it.
h. Create a relationship model kovariat X and Y.

According to the [17] binary logistic regression model is distric comparison of the likelihood of an event/success (π) and the likelihood of failed events (1-π). Specific form of logistic regression model with *p* variables predictors revealed in the equation (1).

$$\pi(\mathbf{x}) = \frac{\exp\left(\beta_0 + \sum_{m=1}^{p} \beta_m x_m\right)}{1 + \exp\left(\beta_0 + \sum_{m=1}^{p} \beta_m x_m\right)} \quad (1)$$

Similarities (1) may be simplified as follows.

$$g(\mathbf{x}) = \ln\left(\frac{\pi(\mathbf{x})}{1-\pi(\mathbf{x})}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p = \mathbf{x}^T \boldsymbol{\beta} \quad (2)$$

With $\pi(\mathbf{x})$ is the success probability $1-\pi(\mathbf{x})$ is the probability to fail, β_m is the parameters of linier function with a variable predictors $m = 1.2, \ldots, p$.

Rosenbaum and Rubin (1983) defines the propensity score for the observation i (i = 1, ...,n) as the conditional probability of treatment based on specific characteristics kovariat x_i observed where randomness cannot be done. Propensity Score Stratification (PSS) is the procedure of grouping the subject to the class /strata based on the estimation of propensity score. The subject of the sorted according to estimates of the propensity [6]. Cochrane (1968) shows that the 5 strata are enough to reduce 90% from the bias to one kovariat [15].

According to Littnerova et al. [4] propensity score using logistic regression model with response variable is a binary form with the model as follows.

$$e(\mathbf{x}_i) = \mathrm{P}(Z_i = 1 | X_i = x_i) = \frac{\exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip})} \quad (3)$$

With β_0 is constant, $\beta_1, \beta_2, \cdots, \beta_p$ regression coefficient and $x_1, x_2, \cdots, x_p$ is covariate variable.

According to Cochran & Rubin (1973) in the pan & Bai [16] measure large biases reduced for each kovariat can use common (4)

$$PBR = \frac{B_{\text{before PS}} - B_{\text{after PS}}}{B_{\text{before PS}}} \text{x}100\% \quad (4)$$

and

$$B = p_1(x_p) - p_0(x_p) \quad (5)$$

$$B_{after\ PSS} = \sum_{k=1}^{K} B_k \quad (6)$$

With the PBR is a Percent Bias Reduction, B is the difference in the average from the treatment groups and control for every covariate, $p_1(x_p)$ and $p_0(x_p)$ a proportion of covariate for treatment groups and control, $B_{\text{before PS}}$ and $B_{\text{after PS}}$ each is a difference in the average treatment groups and control before done propensity score and after done propensity score.

According to Tu & Zhuo [7] steps *propensity score stratification analysis bootstrap* explained as follows.

- Select covariate as confounder for the estimation of propensity score. Confounder selection process can be based on the theory and based on empirical evidence about the relationship between the variables.
- The estimation of the value of the propensity score.
- Share strata based on the propensity score.
- Check the balance kovariat on confounders between the group treatment and non-treatment.
- Assume the N_{tk} number of class treatment groups subject to- k, the N_{ck} number of class control group subject to- k, $Y_{tk1}, \ldots, Y_{tkN_{tk}}$ class treatment group response to- k, $Y_{ck1}, \ldots, Y_{ckN_{ck}}$ class treatment group response to- k, $k = 1, \ldots, K$.
- Determine the average response from treatment groups $\hat{Y}_{tk}$ and control groups $\hat{Y}_{ck}$ in k Starting with kangaroos and the number of the subject of treatment groups n_t and controls n_c with common (7).

$$\hat{Y}_{tk} = \sum_{i=1}^{N_{tk}} \frac{Y_{tki}}{N_{tk}};\ \hat{Y}_{ck} = \sum_{i=1}^{N_{ck}} \frac{Y_{cki}}{N_{ck}};\quad n_t = \sum_{k=1}^{K} N_{tk};\ n_c = \sum_{k=1}^{K} N_{ck} \quad (7)$$

- Estimates of the Average Treatment Effect (ATE) $\hat{\theta}$ obtained as follows.

$$\hat{\theta} = \sum_{k=1}^{K} \frac{N_{tk} + N_{ck}}{n_t + n_c} \left(\hat{Y}_{tk} - \hat{Y}_{ck}\right) \quad (8)$$

- Calculate the standard error ATE with equation (9).

$$\hat{s}(\hat{\theta}) = \sqrt{\sum_{k=1}^{K} \left(\frac{N_{tk} + N_{ck}}{n_t + n_c}\right) \left(\frac{s_{tk}^2}{n_{tk}} + \frac{s_{ck}^2}{n_{ck}}\right)} \quad (9)$$

- Determine the sample *bootstrap partition* $(\mathbf{X}_i^{(b)}, Z_i^{(b)})$ where i =1, ...,n taken from $(\mathbf{X}_i, Z_i)$ with the return.
- Mengestimasi return value of the *propensity score* for each resample subject
- Repeat steps 1 to 5 until the obtained estimates of ATE $\hat{\theta}^b$ and standard error.
- Perform the steps 6 to 8 as many as B (200 transactional replication)
- Determine the standard error of the *bootstrap estimates* for the distribution of the sampling $\hat{\theta}$ using common 10.

$$\bar{\hat{\theta}} = B^{-1}\sum_{b=1}^{B}\hat{\theta}^b$$

$$se_B = \sqrt{\frac{\sum_{b=1}^{B}(\hat{\theta}^b - \bar{\hat{\theta}})^2}{(B-1)}} \quad ;b = 1,2,\ldots,B \tag{10}$$

- Test statistics Z bootstrap

$$Z^* = \frac{\hat{\beta}_m^* - \hat{\beta}_m}{SE\left(\hat{\beta}_m^*\right)} \tag{11}$$

- Calculate p-value bootstrap using the following equation.

$$p - value = \frac{\left|\left\{Z^* \geq Z_0\right\}\right| + 1}{B+1} \tag{12}$$

With

$$\left|\left\{Z^* \geq Z_0\right\}\right| = \sum_{b=1}^{B} I\left\{Z^* \geq Z_0\right\} \tag{13}$$

Z_0 Used to test the significance of the parameters before *bootstrap* done, $I\{A\}$ is the indicator of the Genesis A, where $I\{A\} = 1$ if a true and $I\{A\} = 0$ if A one, $A = Z^* \geq Z_0$ and B is the number of replication. Testing the balance on this research done with test-t-test and z. The tests t used to test the difference between the two groups for continuous data, while the test z used for category data.

RESULTS AND DISCUSSION

Descriptive analysis is the early stages of the exploration of the data that is done to get an overview of the research data. Patient characteristics can be seen from the descriptive on each of the variables.

Table 1. Patient Characteristics Based on Status of NDP

Covariate (X)	Status NDP		Total (%)
	Not	Yes	
Gender (X_2)			
- The Male	12	33	54,87
- Women	17	20	45,13
Dyslipidemia (X_3)			
- Not	21	35	68.29
- Yes	8	18	31,71
Obesity (X_6)			
- Not	18	20	46,34
- Yes	11	33	53,66
HbA1C (X_7)			
- Low	18	17	42,68
- High	11	36	57,32
Acid Uric Serum (AUS) (X_8)			
- Low	19	14	40,24
- High	10	39	59,76
Total (%)	35,37	64,63	

Based on Table 1 it is known that diabetics affected by the NDP more than those who are not NDP. This can be seen from the percentage of total patients no NDP that more than 50%. According to covariate (X) known that most patients with diabetes mellitus who treated in X is man does not have dyslipidemia, have obesity, serum HbA1C high and have the level of wear high. This can be seen from the total presentation more than 50%.

The first step in the analysis of the PSS bootstrap is to determine the variables confounding. Confounding variable is determined based on the theory and empirical evidence in the form of the relationship between the variables. The test statistics used to examine the relationship between the variables is test chi-square. Based on previous research [15] known variables related with the level of wear is the age, gender, dyslipidemia, hypertension, obesity and HbA1C. while according to research Darsana [10] It is known that the level of Acid Uric Serum (AUS) correlates positively with NDP, this shows that the level of wear is potentially variable as confounding variables. To prove it, done test dependencies between the variables. Test results the dependencies between the variables are displayed in the following table.

Table 2. Test results dependencies Covariate X

Variables	Chi-Square	df	P- *value*	Decision
X1*X8	1.900	3	0.594	Fail to reject H 0
X2*X8	4.666	1	0.031	Reject H 0
X3*X8	0.001	1	0.981	Fail to reject H 0
X4*X8	12.114	1	0.001	Reject H 0
X5*X8	6.362	2	0.042	Reject H 0
X6*X8	0.381	1	0.537	Fail to reject H 0
X7*X8	0.122	1	0.727	Fail to reject H 0
X8*Y	28.676	1	0.000	Reject H 0

Table 2 provides information that the level of AUS (X_8) has a relationship with the variables gender (X_2), hypertension (X_4), long suffering from diabetes mellitus (X_5), and the status of the NDP (Y). This shows that the level of AUS (X_8) related with covariate (X) and is a risk factor of the status of the NDP. So the level of AUS selected as confounding variables (Z) with parameter, θ to know how much influence the level of AUS against the status of NDP.

After confounder determined, the next step is the estimation of the value of the propensity score. Basically, the same propensity value with logistic regression model. Therefore, the propensity value can be known if the parameters from the logistic regression model is obtained. The method used to estimation binary logistic regression model parameter is the bootstrap MLE method. The results of the estimation of the parameters are displayed in the following table.

Table 3. The estimation of the parameters with the Bootstrap MLE

Covariate	The parameters (β^*)	SE	p-value	OR
Intercept	0 .189	2.631	0.995	1.208
X_{1}	-0 .025	0.047	0.572	0.975
$X_{2.1}$	-0 .942	0.766	0.139	0.390
$X_{3.1}$	-0 .588	0.718	0.388	0.555
$X_{4.1}$	1 .423	0.763	0.065 *	4.150
X_{5}	0 .162	0.107	0.035 **	1.176
$X_{6.1}$	0 .142	0.650	0.721	1.153
$X_{7.1}$	-0 .024	0.656	0.975	0.976

Based on Table 3 it is known that a significant effect on the level of AUS (z) is one variable have hypertension ($x_{4.1}$) with p-value = 0.065 and long suffering from diabetes mellitus (x_5) with p-value = 0.035. From the Table 2 can also formed the logistics regression model or the value of the *propensity score* $e(x_i)_{boot}$ as follows.

$$e(x_i)_{boot} = \frac{\exp(0.189 - 0.025X_1 - 0.942X_{2.1} - 0.588X_{3.1} + 1.423X_{4.1} + 0.162X_5 + 0.142X_{6.1} - 0.024X_{7.1})}{1 + \exp(0.189 - 0.025X_1 - 0.942X_{2.1} - 0.588X_{3.1} + 1.423X_{4.1} + 0.162X_5 + 0.142X_{6.1} - 0.024X_{7.1})} \qquad (14)$$

After obtained the value of propensity, next is grouping the subject to the class/strata based on the value of the propensity. At this stage is also the subject is divided into 2 to 5 strata to search for strata prove that covariate already balance on all strata. To PSS bootstrap, resampling is done as much as 200 times. This is based on the [6] stated that the replication of the 200 times is enough to estimate the standard error. Testing using p-value bootstrap according similarities (12), to covariate category data using the p value of the test z while for covariate continuous data using the p value of the test t for continuous data variable. To rank the significance α = 10% obtained the test result balance for the bootstrap samples is shown in the following table.

Table 4 testing the Balance for the Bootstrap Samples

# Strata	Strata ke-k	p-value category data					p-value continue data	
		X_2	X_3	X_4	X_6	X_7	X_1	X_5
Before strata		0.03	0.98	0.00	0.54	0.73	0.56	0.00
2	1	0.96	0.86	0.39	0.86	0.38	0.55	0.37
	2	0.51	0.88	0.61	0.66	0.63	0.63	0.68
3	1	1.00	1.00	0.44	0.91	0.84	0.22	0.10
	2	0.72	0.69	0.31	0.93	0.93	0.64	0.90
	3	0.45	0.96	NA	0.77	0.68	0.12	0.55
4	1	0.25	0.16	NA	0.75	0.40	**0.00**	0.46
	2	0.76	0.46	0.78	0.45	0.89	0.59	0.69
	3	0.55	0.96	0.64	0.73	0.21	0.94	0.90
	4	0.74	0.89	NA	0.68	0.55	0.18	0.94
5	1	0.68	0.11	NA	0.69	0.44	0.14	0.49
	2	0.20	0.55	0.32	1.00	0.71	0.92	0.86
	3	0.45	0.42	0.32	0.30	0.61	0.73	0.68
	4	0.15	0.55	0.55	0.86	0.18	0.84	0.88
	5	0.49	0.22	NA	0.76	0.23	0.72	0.79

Table 4 shows that before *stratification* there are three covariate that does not balance. The covariate is X_2, X_4, and X_5. This is shown by the value of the p-value less than α = 10%. After done *stratification* obtained the results for strata as much as 2, 3

and 5 strata can be seen that all covariate already balance, this is shown by the value of the p-value on each strata for strata 2, 3 and 5 are more than α = 10%. While for strata 4 It is known that there are yet covariate balance, namely covariate X_1 on strata 1 with the value of p the value of the lowest infection rate was from 0.00 or p-value < α. To the value of the NA (Not Available) is on covariate X_4 because there is one of the group treatment or controls that do not have a member, causing the value of the z score undefined values and p-value not out. The results of the test balance in table 4 also shows that a good level and can be used on the next steps is strata 2, 3 and 5 because this gives covariate strata that balance in accordance with the conditions of the propensity score stratification.

The last step from PSS bootstrapping is the estimates of the *Average Treatment Effect* (ATE). The estimation of the value of ATE and standard error is done by using the equation (8) and similarities (9). The results of the estimation of the value of the *Average Treatment Effect* (ATE) for each strata and standard error ATE shown in table 5 below

Table 5 Results of the estimation of ATE and standard error for the Bootstrap Samples

# Strata	Strata to the k	The value of ATE Strata to the k	ATE	SE (ATE)	*P-value*
2	1	0.2804	0.4793	0.0127	0.005
	2	0.1936			
3	1	0.1704	0.4336	0.0123	0.005
	2	0.1446			
	3	0.1186			
4	1	0.1019	0.4237	0.0156	0.005
	2	0.1565			
	3	0.1072			
	4	0.0482			
5	1	0.0921	0.4174	0.0112	0.005
	2	0.1306			
	3	0.0805			
	4	0.0927			
	5	0.0255			

The estimation of treatment effects (ATE) is very important in the propensity, because basically the purpose of propensity is getting estimates of ATE that unbiased and more accurate even though there is a *confounding variable* in the design of the research. Based on Table 5 it is known that that each of the strata provides value ATE and standard error is different where strata provide estimates of the value of the largest error standard is a group of 4 strata with estimates of ATE of 0.4237 and standard error 0,0156. This is caused by the existence of covariate that does not balance on 4 strata namely covariate X_1 strata 1 which gives the value of p-value = 0.010 < α. While the strata that provide estimates of the value of the smallest error standard is a group of 5 strata with estimates of the value of ATE of 0,4174 and standard error 0.0112. Table 5 also shows that the level of variable AUS (Z) is an influential variable significantly on Genesis Diabetic Neuropathy Peripheral (NDP) in patients with diabetes mellitus. This can be seen from the p-value < α = 10 on all strata.

To assess whether the PSS method is good or not can be seen from how big bias is capable of reduced by the methods of PS. Great bias can be known by using the formulation of the formula developed by Cochran in 1968. In the journal that was written by [1] explained that according to Cochran (1968) 5 strata are able to reduce the bias nearly 90%. Therefore, in this discussion will be proven whether it is true that the 5 strata are the best strata which is able to reduce the nearly 90 percent bias or not. Percent reduction of bias will be calculated using the equation (4) until the similarities (6). The results of the calculation of the percent reduction in the bias for each covariate PSS bootstrap shown in table 6 below.

Table 6. Percent Reduction Bias bootstrap PSS

Covariate	PBR to PSS bootstrap			
	2 Strata	3 Strata	4 Strata	5 Strata
X_1	43.55	4.03	98.66	34.16
X_2	77.83	80.06	82.39	91.39
X_3	99.78	92.64	97.00	89.25
X_4	78.57	77.37	83.77	93.87
X_5	38.13	42.57	79.21	62.98
X_6	99.14	95.99	97.20	86.14
X_7	83.40	96.69	83.75	98.27
Overall	46.97	44.90	59.45	89.83

Based on the table 6 it is known that the overall percent reduction Bias (PBR) for 5 strata is greater than the other strata namely 89,83%, while PBR smallest present on the group 3 strata namely 44,90%. In addition, Table 5 also shows that the bootstrap PSS with 5 strata give standard error is smaller than the other. Securities reduction shown bias with a small error default values which means that the variant also small. This shows that the theory of [5] stated that the 5 strata are able to reduce the nearly 90 percent biases and better than other strata proved the truth.

After the obtained estimates of the effects of the level of A (Z) against NDP status (Y) using analysis of the propensity score. So the next step is to know the direct relationship covariate X against NDP status (Y). Because the status of the NDP is a binary category data, then the analysis that is used is a binary logistic regression analysis. The estimation of parameters

using the same method with the logistics regression parameter estimates used in the previous stages using the Maximum Likelihood method (MLE) with resampling (bootstrap).

Table 7. The estimation of the parameters with the Bootstrap MLE

Covariate	Parameters (B^*)	SE	P-value	OR
Intercept	-1 .792	3.345	0.378	0.166
X_1	-0 .006	0.059	0.890	0.994
$X_{2.1}$	-0 .726	0.929	0.194	0.483
$X_{3.1}$	-0 .566	0.992	0.368	0.568
$X_{4.1}$	1 .126	0.925	0.094*	3.083
X_5	0 .243	0.111	0.055*	1.275
$X_{6.1}$	0 .962	0.791	0.099*	2.617
$X_{7.1}$	1 .417	0.761	0.055*	4.129

Based on the table 7 it is known that the variables affect the significant impact on the status of NDP (Y) on equal significance α = 10% is a variable that DIABETICS have hypertension (x4.1) with p-value = 0.094, long patients suffer from diabetes mellitus (x5) with p-value = 0.055, diabetics have obesity (x6.1) and diabetics long have the level of HbA1C high (x7.1) with p-value = 0,055. Based on the table 7 can also formed the logistics regression model X covariate relationship with the status of NDP (Y) as follows.

$$\pi(x_i) = \frac{\exp(1.126x_{4.1} + 0.243x_5 + 0.962x_{6.1} - 1.418x_{7.1})}{1 + \exp(1.126x_{4.1} + 0.243x_5 + 0.962x_{6.1} - 1.418x_{7.1})}$$

Similarities (20) provides that opportunity diabetics that hypertension affected by Diabetic Neuropathy Peripheral (NDP) is 3,083 times greater than diabetics who do not hypertension. Every increase of 1 years long patients suffer DM, then the opportunity for patients affected by NDP will rise by 1.275 with other variables constant assumptions. The opportunity diabetics that obesity affected NDP is 2,617 times greater than diabetics who do not obesity, and opportunities diabetics that HbA1C high affected NDP is 4.129 times greater than diabetics that HbA1C low.

From the Table 7 also known that the status of the NDP patients influenced by the status of hypertension patients, long patients suffer DM, the status of obesity patients and serum HbA1C patients with diabetes mellitus. This means that by the status of hypertension patients, long patients suffer DM, the status of obesity patients and serum HbA1C diabetics is the variables that affect directly to the status of NDP patients.

CONCLUSION

Propensity score is a good method used to see the effect of treatment on the study of observation especially on the data involves *confounding variable* in it. The results of the estimation of the parameters with the MLE method does not close the form and done iteration Newton-Raphson, next iteration Newton-Raphson resampling results as much as B times and obtained bootstrap parameters. While the results of the bootstrap PSS analysis show that the 5 strata provide standard errors that are smaller and the largest bias reduction (89.83%) compared to other strata. In addition, from logistic regression model it is known that the status of hypertension patients, long patients suffer DM is the variables that impact directly and indirectly on the status of NDP diabetics, while the status of obesity patients and serum HbA1C diabetics is the variables that affect only directly to the status of NDP patients.

REFERENCES

[1] Rosenbaum, P.R., & Rubin, D.B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. Journal Biometrika, vol.70, No.1, pp. 41-55.

[2] Littnerova, S., Jarkovsky, J., Parenica, J., Pavlik, T., Spinar, J., & Dusek, L. (2013). Why to use Propensity Score in Observational Studies? Case Study Based on Data from the Czech Clinical Database AHEAD 2006-09, cor et Vasa, 55(4), pp. 383-390.

[3] D'Agostino, R.B. (1998) Tutorial in Biostatistics Propensity Score Method for Bias Reduction in the Comparison of a Treatment to a Non-Randomized Control Group. 17, pp. 2265-2281.

[4] Austin, (2011). A Tutorial and Case Study in Propensity Score Analysis: AN Application to Estimating the Effect of In-Hospital Smoking Cessation Counseling on Mortality. Multivariate Behavioral Research, 46: pp:119–151.

[5] Tu, W., & Zhou, X.H. (2002). A Bootstrap Confidence Interval Procedure for the Treatment Effect Using Propensity Score Subclassification. Health Services & Outcomes Research Methodology, pp.135–147.

[6] Efron, B. & Tibshirani, Robert J. (1993). An introduction to the Bootstrap. New York: Chapman & Hall.

[7] Wild, S., Riglic, G., & Green, A. (2004). Global Prevalence of Diabetes: Estimates for the year 2000 and Projection 2030. Diabetes Care vol 27.

[8] Bansal, D., Gudala, K., Muthyala, H., Esam, H.P, Nayakallu, R. & Bhansali, A. (2014). Prevalence and Risk Factors of Development of Peripheral Diabetic Neuropathy in Type 2 DM in a Tertiary Care Setting. Journal Diabetes Invest, Vol. 5: pp. 714–721.

[9] Qilsi F.R.M. & Ardiansyah M. (2012). Hubungan Antara Hiperglikemia, Usia, dan Lama Menderita Pasien Diabetes dengan Angka Kejadian Neuropati Diabetika. FKIK Yogyakarta.

[10] Darsana, I.N. (2014). Korelasi Positif Kadar Asam Urat Serum Tinggi dengan Neuropati Diabetik Perifer pada Penderita DM Tipe-2 di RSUP Sanglah Denpasar. TESIS UNUD.

[11] Bruce, S.G., & Young, T.K. (2008). Prevalence and Risk Factors for Neuropathy in a Canadian First Nation Community. Diabetes Care vol. 31, pp. 1837-1841.

[12] Sadr, SM., Namayandeh, SM., Moadares, MM., & Rafiei, M. (2009). Serum Uric Acid Levels and Its Association with Cardiovascular Risk Factors. Iranian J Publ Health, Vol. 38, No.1, 2009, pp.53-59.

[13] Tesfaye, S. 2004. Epidemiology and Etiology of Diabetic Peripheral Neuropathies. Ad Stud Med, 4: pp.1-8.

[14] Veves, A., & Malik, R.A. (2007). Diabetic Neuropathy Clinical Management Second Edition. Totowa: Humana Press.

[15] Akolo I.R., B.W.Otok, Santi W. Purnami, Rama Hiola. (2016). Propensity Score Stratification Analysis using Logistic Regression for Observational Studies in Diabetes Mellitus Cases. Proceeding of 3RD International Conference on Research, Implementation and Education of Mathematics and Science Yogyakarta, 16 – 17 MAY 2016, M 59 – M66

[16] Pane R., Bambang W.O., Ismaini Z., I Nyoman B. (2016). Bootstrap Inference Longitudinal Semiparametric Regression Model. AIP Conference Proceedings 1707, 080003 (2016); doi: 10.1063/1.4940860

[17] Hosmer, D.W, & Lemeshow, S. (2000). Applied Logistic Regression. New York: John Wiley and Sons, Inc.