M – 10

Propensity Score Stratification Analysis using Logistic Regression for Observational Studies in Diabetes Mellitus Cases

Ingka Rizkyani Akolo[1], B.W.Otok[2], Santi W. Purnami[2], Rama Hiola[3]
[1]Dept. of Statistics, Institut Teknologi Sepuluh Nopember
[2]Dept. of Statistics, Institut Teknologi Sepuluh Nopember
[3]Dept. of Public Health, Universitas Negeri Gorontalo
inkarizkyani05@gmail.com / molavecha@gmail.com

***Abstract*—** Observational studies are the basis of epidemiological research to draw the conclusions of the effects or a response treatment. In general, a randomized trial is required in order to meet the assumption of independence to minimize the bias effects. However in an observational study, particularly in medical field, randomization not able to implement because conduces in doubtful treatment effects estimation. Propensity score is the conditional probability to get certain treatments involving the observed covariates. This method is used to reduce bias in the estimation of the impact of treatment on observational data for their confounding factors. If treatment is binary, then the logistic regression model is one estimated of propensity score because of easiness in terms of estimation and interpretation. In the analysis of observational studies, propensity score stratification (PSS) has proven to be one of methods to adjust the unbalanced covariate for the purposes of causal inference. The data used in this study is the medical records of patients DM in X hospital about the factors that influence the type of diabetes mellitus. In this study PSS used in diabetes mellitus cases to reduce bias due to confounding factors, so that can be known the factors affect the type of diabetes mellitus with obesity as confounding factors. The results of PSS analysis is known that the variables directly influence the type of DM are obesity, age, gender and variable does not directly influence the type of DM are genetic variable, sport activities and dietary habit of patients DM.

Keywords: *observational studies, confounding, propensity score stratification, diabetes melitus*

I. INTRODUCTION

The attention of non-communicable diseases is increasing currently. From ten leading causes of death, two of them are non-communicable diseases. Diabetes mellitus (DM) is a non-communicable disease with high prevalence. International Diabetes Federation (IDF) stated that people with diabetes mellitus figure reached 382 million people of the world in 2013. It is estimated as 592 million in 2035. In Indonesia, people with diabetes mellitus has reached 8.4 million in 2000 and is estimated to be approximately 21.3 million in 2030. Because of high number of patients, it makes Indonesia ranks fourth after the United States, India and China [1].

According to the results of Indonesia Basic Health Research (RISKESDAS) in 2013, an increase in the prevalence of Indonesia's diabetes mellitus in 2007 was 1.1% to 2.1% in 2013. The results of the analysis of the Diabetes Mellitus prevalence's picture based on a doctor's diagnosis and symptoms increase with age. It began with age ≥ 65 years old of decline. The prevalence of diabetes in women is 1.7% while men have 1.4%. Based on its territory, the prevalence of urban areas (2.0%) is higher than in rural areas 1.0% [2].

Diabetes mellitus (DM) is a chronic metabolic disorder due to the pancreas does not produce enough insulin or the body can not use the insulin that is produced effectively. Insulin is a hormone that regulates blood glucose levels. Diabetes mellitus is classified into type 1 diabetes, which is known as insulin-dependent or childhood-onset diabetes, characterized by a lack of insulin production. Type 2 diabetes, known as non-insulin-dependent or adult-onset diabetes, caused by the body's inability to use insulin effectively which then lead to overweight and lack of physical activity [3].

Increasing the number of people with diabetes are mostly caused by the interaction between the factors of genetic susceptibility and exposure to the environment, such as changes in lifestyle and physical activity often leading to obesity. It is a risk factor for the onset of DM [4]. Therefore, diabetes mellitus type 2 is often also called diabetic lifestyle for causes not only because of heredity, but also environmental factors include age, obesity, insulin resistance, food, physical activity, and unhealthy play roles in the occurrence of diabetes [5].

Research on the incidence of diabetes mellitus (DM) has been done in large quantities. For example Wicaksono [4] investigated the factors associated with the occurrence of diabetes mellitus (DM) type II using descriptive analysis and logistic regression. Trisnawati et al. [6] studied the risk factors of type 2 DM outpatients using the McNemar test and logistic regression and Indriyani et al. [7] studied the effect of physical exercise to decreased levels of blood sugar of patients with type 2 DM using the t test with the one group pretest-posttest study design.

The above researches mostly used descriptive analysis and logistic regression without considering the possibility of a powerful combination of factors affecting diabetes mellitus (DM). In fact, as explained previously that the combination of these factors led to the existence of confounding variables that lead to obtain inaccurate conclusions.

Some previous studies have tried to discuss confounding factors randomly, but in the case of health sector, it can not be done. But how the confounding variables included in the factors studied. Therefore, we need a method that can handle the effects of bias caused by these confounding factors. One method that can handle confounding is the propensity score method. it was first introduced by Rosenbaum and Rubin in 1983. The propensity score is defined as the conditional probability to receive interventions based on those characteristics before the intervention [8]. This method is a statistical adjustment that can be used to analyze data from non-experimental research design where design giving treatment through randomization to treatment or control group is not possible. Researchers can use the propensity score for statistical balance or equalize the group of research subjects to reduce bias due to the provision of treatment which is not random.

One method of propensity score that is proven to reduce bias due to confounding effects is the propensity score stratification method. This method focuses on the division of classes / strata based on the estimated value of propensity score. The division of classes / strata aims to balance the distribution between treatment and control groups so that estimate of average treatment effect more accurate.

Several studies of the model used to estimate the value of propensity score, they are McCaffrey et al.[9] which used a model of generalized boosted, McCandless et al. [10] used Bayesian, and Littnerova et al.[11] used logistic regression to estimate propensity score. Of all the study, estimated by logistic regression simpler and easier in interpretation, particular to the category data used.

Based on the description above, the aim of research in this study are to get an estimation of average treatment effect and binary logistic regression model based on the propensity score that shows the factors affecting the type of DM in patients treated in X hospital district after being controlled by confounding variables of obese patients' status.

II. THEORY

2.1 Logistic Regression Model

According to Hosmer & Lemeshow [12] binary logistic model is the logarithm of odds ratio of occurrence of success (π) and probability of occurrence of fail ($1 - \pi$) . The specific form of the logistic regression model with p predictor variables expressed in equation (2.1)

$$\pi(\mathbf{x}) = \frac{\exp\left(\beta_0 + \sum_{m=1}^{p} \beta_m x_m\right)}{1 + \exp\left(\beta_0 + \sum_{m=1}^{p} \beta_m x_m\right)} \tag{2.1}$$

Form of simplification of the equation above, then used a logit transformation of the form below.

$$g(\mathbf{x}) = \ln\left(\frac{\pi(\mathbf{x})}{1-\pi(\mathbf{x})}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p = \mathbf{x}^T \boldsymbol{\beta} \tag{2.2}$$

with $\pi(\mathbf{x})$ is the probability of success, $1-\pi(\mathbf{x})$ is probability of fail event, β_m are the parameters of the linear function with the predictor variables $m = 1,2 , \ldots , p$.

2.2 Propensity Score

Propensity score analysis introduced by Rosenbaum and Rubin 1983 in the journal entitled "The central role of the propensity score in observational studies for causal effects". Propensity score analysis is a statistical method that rapidly evolving innovative and useful for evaluating treatment effects when using observational data [13]. Rosenbaum and Rubin [8] define the propensity score for observation i (i = 1, ..., n) as the conditional probability of a specific treatment (Z_i = 1) versus non-treatment (Z_i = 0) based on the characteristics of the covariates $\mathbf{x}_i$ observed.

According to Guo & Fraser [13] the value of propensity score is defined as follows.

$$e(\mathbf{x}_i) = P(Z_i = 1 \mid X_i = x_i) \tag{2.3}$$

According to Littnerova et al.[11] propensity score using a logistic regression model, the response variable is a binary where to treatment and to the control unit with the following model.

$$e(\mathbf{x}_i) = \mathrm{P}(Z_i = 1 \mid X_i = x_i) = \frac{\exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip})} \tag{2.4}$$

with β_0 is a constant, $\beta_1, \beta_2, \cdots, \beta_p$ the regression coefficients and $x_1, x_2, \cdots, x_p$ are covariate variables.

According to Cochran & Rubin (1973) in the Pan & Bai [14] measures the bias is reduced for each covariate can use equation (2.11)

$$PBR = \frac{B_{\text{before PS}} - B_{\text{after PS}}}{B_{\text{before PS}}} \text{x}100\% \tag{2.5}$$

and

$$B = p_1(x_p) - p_0(x_p) \tag{2.6}$$

with *PBR* is Percent Bias Reduction, *B* is an average difference of the treatment group and the control group for each covariate , $p_1(x_p)$ and $p_0(x_p)$ are proportion of covariates for the treatment group and the control group, $B_{\text{before PS}}$ and $B_{\text{after PS}}$ are represents the difference between the average treatment and control group before propensity score and after propensity score.

2.3 Propensity Score Stratification

Propensity Score Stratification (PSS) is a procedure of classifying subjects into classes based on the estimated propensity score . Subjects are sorted by the estimated propensity score (Austin, 2011). Cochrane (1968) showed that the five sub-class is enough to reduce 90 % of bias with a single covariate [15]. Imbens [16] declared the entire bias under unconfounded associated with the propensity score, it indicates that under the normality used 5 strata change is largely biased with all covariates.

According to Yanovitzky, Zanutto, and Hornik [17] general steps of propensity score analysis are described as follows

1. Choose a covariate as a confounder for the estimation of propensity score. The election process can confounder based on theory and empirical evidence about the relationship between variables.
2. Estimated value of propensity score.
3. Divide the strata based on the propensity score.
4. Check the balance of covariates between the treatment group and the non-treatment.
5. Calculate the effect of confounders.

One way to assess the quality of the propensity score stratification by comparing a variety of statistics such as mean, median, variance, t-test statistics, chi-square test or Kolmogorov-Smirnov (KS) test on each covariate [15]. In this study, KS and chi-square used for testing difference distribution between the treatment group and the control group.

2.4 *Diabetes Mellitus*

Diabetes mellitus is metabolic diseases which is a collection of symptoms that arise in a person because increase in blood glucose levels above normal values . The disease is caused by disorders of the metabolism of glucose due to a deficiency of insulin both absolute and relative terms . There are two types of diabetes mellitus. The first type of DM is type 1, that usually acquired since childhood and results from the pancreas failure to produce enough insulin. The second type of DM is type 2, that

usually acquired an adult and condition in which cells fail to respond to insulin. According Poretsky [18] factors that affect type 1 diabetes is a genetic, autoimmune, age, race and ethnicity, gender, and environmental factors such as viral infections, diet / nutrition, stress. In addition, according Gungor, Hannon, Libman, Bacha, & Arslanian [19] factors affecting the type 2 diabetes are genetic, age, gender and environmental factors such as diet, obesity, sports activities.

III. METHODOLOGY

The method used in this study is propensity score stratification (PSS) method to find the factors that influence the type of diabetes mellitus (DM) with obesity status of patients as a confounding factor. The data used is secondary data from medical records of patients (DM) at Hospital X in 2013. The number of respondents are 497 patients. The Patients consist of patients with type 1 of DM (42 patients) and patients with type 2 DM (455 patients). The response variable is the type of DM and predictor variables are genetic, age, gender, dietary habit, sport activities and obesity. The stages of research process can be seen in Figure 1 below.

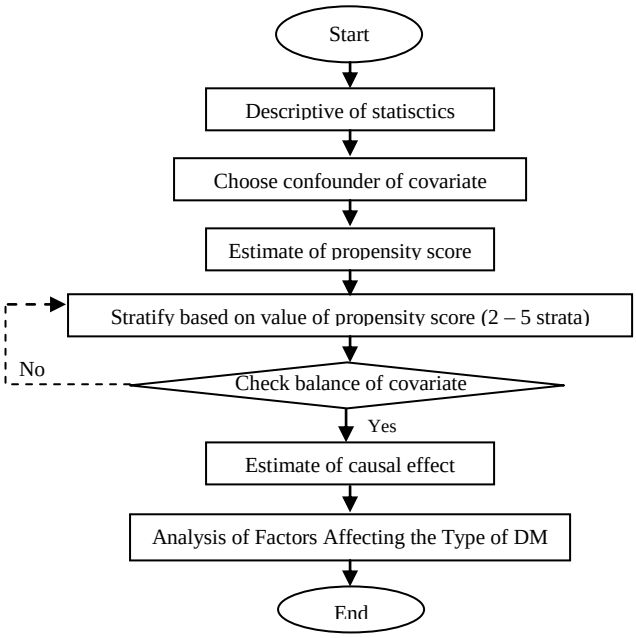


FIGURE 1. STAGES OF RESEARCH PROCESS

IV. RESULT AND DISCUSSION

4.1 Descriptive of Statistics

Descriptive of statistics is an early stage of data exploration to get a general overview of the research data. Characteristics of respondents can be seen from the descriptive of each variables shown in Table 2.

TABLE 1. DESCRIPTIVE ANALYSIS OF COVARIATE

Covariate	Status Obesity		%	Type of DM		%
	Obesity	No Obesity		Type 1	Type 2	
Genetic						
- Have genetic	379	31	82,49	0	410	82,49
- Have not genetic	52	35	27,51	42	45	27,51
Age	431	66	-	42	455	-
Gender						
- Male	192	32	45,07	24	200	45,07
- Female	239	34	54,93	18	255	54,93
Dietary habit						

- Meet	29	63	18,51	25	67	18,51
- No Meet	402	3	81,49	17	388	81,49
Sport Activities						
- Active	29	65	18,91	27	67	18,91
- Less Active	402	1	81,09	15	388	81,09

Based on the table 2 can be shown that to 82.49 % patients have genetics DM, 81.49 % patients have dietary habit (no meet) and 81.09 % patients less active in sports activities. In addition, it was known that the number of female patients (54.93%) are greater than male patients (45.07 %). From table 2 can shown too that the most patients have obesity and type 2 diabetes are genetics diabetes, female gender, dietary habit (no meet) and patients who has less active exercise in sport activities.

4.2 Propensity Score Stratification Analysis

4.2.1 Choose a covariate as a confounder

The first step in the propensity score analysis is to choose covariate as a confounder variable. The determination of confounding variables based on the theory and proven with empirical evidence like the relationship between variables. Testing relationship between variables used chi-square test. Based on research conducted by Betteng, et al.[5] known that obesity has a relationship with genetic factors, dysfunction of the brain, dietary habit is over, less activities of sport, emotional, environmental factors, social factors and lifestyle. Therefore, this relationship will be proven by empirical evidence using chi-square test . Results of testing the correlation between covariates with obesity variables are shown in Table 2.

TABLE 2. TESTING RESULTS CORRELATION BETWEEN COVARIATES WITH OBESITY

Variable	χ^2	**Df**	**P-*value***	**Decision**
$x_4 * x_1$	66,513	1	0,000	Reject H_0
$x_4 * x_2$	2,047	3	0,563	Failed to reject H_0
$x_4 * x_3$	0,358	1	0,549	Failed to reject H_0
$x_4 * x_5$	298,701	1	0,000	Reject H_0
$x_4 * x_6$	314,208	1	0,000	Reject H_0

Based on Table 2 can be shown that genetic, diet and active sports activities has significant influence to obesity variables. Meanwhile age and gender has not significant influence to obesity. Based on those results, so it is a proof that obesity variable is the most variable that associated with other variables. Therefore, obesity variable is selected as confounding variable Z with parameter θ.

4.2.2 Estimating the Propensity Scores

In this study the propensity score estimated by logistic regression. There are five variables will be estimated, their variables are genetic, age, gender, dietary habit and sports activities. The result of parameter is shown in Table 3.

TABLE 3. PARAMETER ESTIMATION FOR THE RELATIONSHIP OBESITY (Z) WITH COVARIAT (X)

Covariate	**Parameter** (β)	**SE**	**p-value**	**OR**	**OR (95% CI)**
Intercept	3.8357	1.4479	0.0081	33.9019	1.4069 - 16.8948
Genetic	2.3211	0.6562	0.0004**	10.7902	2.9338 - 39.6853
Age	0.0118	0.0192	0.5397	1.0174	0.9706 - 1.0665
Gender	0.1722	0.4500	0.7020	0.9835	0.3872 - 2.4980
Dietary habit	-1.8721	1.3741	0.1731*	0.1682	0.0114 - 2.49269
Sport Activities	-5.2426	1.6029	0.0011**	0.0057	0.0002 - 0.1328

(*) significant at α = 20%, (*) significant at α = 0,1%,

Based on Table 3 can be shown that the variables have significant influence to obesity at significance level (α = 0.1 %) are variable genetic with *p*-value = 0.000 and sport activities with *p*-value = 0.0011, while dietary habit variable is significance at α = 20%. It is indicates that the status of obesity patients DM was determined by genetic factors, dietary habit, and sports activities of patient DM.

From the estimation parameters are shown in Table 3, it can be obtained the value of propensity score below.

$$e(\mathbf{x}_i) = \frac{\exp\left(3,84 + 2,32\ Gen(1) + 0,01\ Age + 0,17\ Gndr(1) - 1,87\ DH(1) - 5,24\ SA(1)\right)}{1 + \exp\left(3,84 + 2,32\ Gen(1) + 0,01\ Age + 0,17\ Gndr(1) - 1,87\ DH(1) - 5,24\ SA(1)\right)} \tag{2.7}$$

Equation (2.7) illustrates that each age of patients DM is increase one year, so the odds of obesity will increase by 1,017 times. The probability of someone who have genetic DM become obesity is 10.79 times greater than someone who does not have a genetic history of diabetes, the probability of a women having obesity is 0.984 times greater than a men , the probability of someone a healthy diet having obesity is 0,168 times than someone whose diets are not healthy and active sports person's probabilitys having obesity is 0.006 times that of someone who rarely exercise.

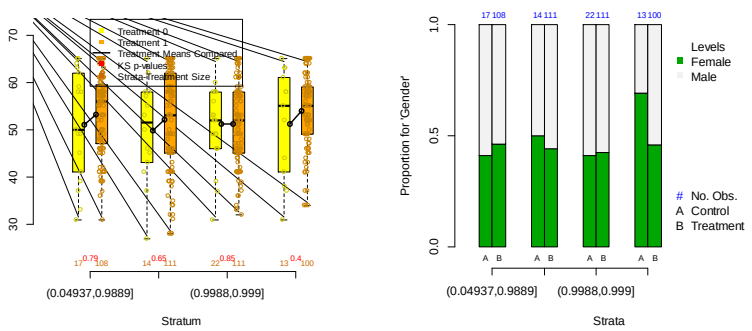
4.2.3 Stratify and Balance the Propensity Scores

After estimating the propensity scores, the next step is subclassified them into different strata. The formation of this stratum aims to balance the treatment and control groups so that estimates of treatment effect is not biased. The number of balanced propensity score strata depends on the number of observations in the data set. Table 4 shows the test of covariate balance after stratification based on the quintiles of the propensity score. Five of the covariates were included in the final propensity score model used for stratification. The initial imbalances were measured by chi-square test for categorical data (genetic, gender, dietary habit and sport activities) and Kolmogorov-Smirnov test for continuous data (age) comparing the obesity and no obesity groups.

TABLE 4. TEST OF STRATA BALANCE

Strata	n	Chi-Square Tests for Balance				KS-Test for Balance
		Genetic	Gender	Dietary Habit	Sport Activities	Age
1	126	0,058	0,800	0,954	0,525	0,790
2	125	0,052	0,780	0,525	1,000	0,650
3	133	0,055	1,000	1,000	0,475	0,850
4	113	0,062	0,150	1,000	1,000	0,400

Based on Table 4 can be shown that after testing using chi-square test for categorical data, their covariates such as genetic, gender, dietary habit and sports activities shows that obesity and no obesity have a balance at all strata. Similarly, for the covariates of age which was tested by Kolmogorov-Smirnov (KS) test. Covariate testing balance is supported by Figure 2. Figure 2 represents a picture which shows a balance between the obesity and no obesity for categorical data (gender) and continuous data (age). So that the analysis can be continued to the next step, the step is estimate average treatment effect or average effect of obesity on the type of DM. Pattern of balance can be seen in Figure 2 below.



(A) (B)

FIGURE 2. COVARIATE OF PROPENSITY SCORE IN BALANCE STRATA (A) AGE, (B) GENDER

4.2.4 Estimating the Causal Effect

Propensity score is an ideal method to see the effect of treatment on observational studies. This method can reduce bias effect because differences distribution of covariate between treatment and control groups. Therefore, before the estimated treatment effects, covariates between the treatment and control groups should be balanced. Because in the previous step has been obtained strata with covariates were balanced, then the next step is estimation of the treatment effect. In this case estimate of the effect of obesity on the type of DM. The estimation results for before and after stratification shown in Table 5.

TABLE 5. RESULT OF ESTIMATION AVERAGE TREATMENT EFFECT (ATE)

ODD RATIO FOR ATE					
BEFORE STRATIFICATION			AFTER STRATIFICATION		
UNADJUSTED	SE UNADJUSTED	95% CI STRATA	ADJUSTED	SE ADJUSTED	95% CI STRATA
16,859	0,3583	8,353 – 34,027	7,065	0,516	2,570 – 19,424

Table 5 shows the result for estimated effect of obesity on the type of DM before and after stratification. From table 5 obtained an average yield effects of obesity on the type of DM before stratification (unadjusted) is 16.859 with a standard error of 0.3585 and after stratification (adjusted) the effect of obesity is 7.065 with the standard error of 0.516. Propensity method also provides estimates of 95% confidence interval between 2.570 and 19.424. This confidence interval shown the difference average between the treatment group and the control of obesity is significant, or in other words, obesity significantly influence the type of DM with the effect is 7.065.

4.3 Analysis of Factors Affecting the Type of DM

After the estimation of treatment effects (obesity) was known then the next step is to determine the relationship of covariates with type of DM.

TABLE 6. PARAMETER ESTIMATION FOR THE RELATIONSHIP BETWEEN TYPE OF DM (Y) WITH COVARIAT (X)

Covariate	Parameter (β^*)	SE	p-value	OR	OR (95% CI)
Intercept	2.8368	1.4367	0.0483	17.0611	1.0211 – 285,0692
Genetic (1)	21.2478	1375.7492	0.9877	1689671554	-
Age	-0.0448	0.0277	0.1059*	0.9562	0,9056 – 1,0095
Gender(1)	0.7892	0.5213	0.1301*	2.2016	0,7925 – 6,1162
Dietary habit(1)	-0.5024	1.3191	0.7033	0.6051	0,0456 – 8,0288
Sport Activities(1)	-1.3926	1.2916	0.2810	0.2484	0,0198 – 3,1234

(*) significant at α = 20%

Based on Table 6 can be shown that the variables significantly influence to the type of DM at significance level α = 20 % are variable age with *p*-value = 0.106 and gender with *p*-value = 0.1301. Based on the table 6 known that the type of DM patients was influenced by the age and gender of patients DM, or age and gender variable are variables that directly influence the type of DM patients.

From the estimation parameters are shown in Table 6, can be obtained logistic regression model covariates significant relationship between the type of DM as below.

$$\pi(\mathbf{x}_i) = \frac{\exp\left(2,837 - 0,045\,Age + 0,789\,Gender\,(1)\right)}{1 + \exp\left(2,837 - 0,045\,Age + 0,789\,Gender\,(1)\right)} \tag{2.8}$$

Equation (2.8) illustrates that any increase 1 year of age patients DM, the odds for type of DM decreased by 0.956 times and the probability for women having type 2 of DM is 2,202 times greater than men.

V. CONCLUSSION

Propensity score is a good method to see the effect of treatment on observational studies, particularly data with different background covariates. The different of covariate can make inaccurate conclusions. Propensity score stratification can balance the covariates between the treatment and control groups so that can reduce bias due to confounding effects. Analysis of propensity score stratification shown that the

variables influence obesity are genetic variable, sports activities and dietary habit of patients and the effect of obesity on the type of DM after stratification is amount 7.065 with a standard error of 0.516. In addition, the variables that directly influence the type of DM patients are obesity, age, gender and variable that does not directly affect the type of DM patients are genetic variable, sport activities and dietary habit of patients DM with the obesity as confounding factors if modeled by logistic regression.

REFERENCES

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

[2] Kemenkes. (2013). *Riset Kesehatan Dasar 2013.* Badan Penelitian dan Pengembangan Kesehatan Kemenkes RI

[3] ----------. (2013). InfoDATIN Pusat Data dan Informasi Kemenkes RI: Situasi dan Analisis Diabetes

[4] Wicaksono, R.P. (2011). Faktor-Faktor yang Berhubungan dengan Kejadian DM tipe-II. Karya Ilmiah Kedokteran UNDIP

[5] Betteng, R., Pangemanan., D., & Mayulu, N. (2014) *Analisis Faktor Resiko Penyebab Terjadinya Diabetes Mellitus Tipe 2 pada Wanita Usia Produktif di Puskesmas Wawonasa.* Jurnal e-Biomedik Vol 2 No 2

[6] Trisnawati, S., Widarsa, T., & Suastika, K. (2013). *Faktor Resiko DM Tipe-2 Pasien Rawat Jalan di Puskesmas Wilayah Kec. Denpasar Selatan.* Public Health and Preventive Medicine Archive, Vol 1, No 1

[7] Indriyani, P., Supriyatno, H., & Santoso, A. (2007). *Pengaruh Latihan Fisik; Senam Aerobik terhadap Penurunan Kadar Gula Darah pada Penderita DM Tipe-2 di Wilayah Puskesmas Bukateja Purbalngga.* Jurnal Media Ners vol 1, No.2 pp 49-99

[8] 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.

[9] McCaffrey, D.F., Ridgeway, G., & Moral, A.R. (2004). *Propensity Score Estimation with Boosted Regression for Evaluating Causal Effect in Observational Studies.* Pschological Method, 9(4), pp. 403.

[10] McCandless, L.C., Gustafson, P., & Austin, P.C. (2009). *Bayesian propensity score analysis for observational data. Statistics in Medicine*, 28, pp 94-112.

[11] 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.

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

[13] Guo, S. & Fraser, M. W. (2010). *Propensity score analysis: Statistical methods and applications*. Thousand Oaks, CA: Sage Publications

[14] Pan, W., & Bai, H. (2015). Propensity Score Analysis: Fundamental and Developments. New York: Gulford Press

[15] Mingxiang, L. (2012). *Using the Propensity Score Method to Estimate Causal Effects: A review an Practical Guide.* Organisational Research Methods 00(0) pp. 1-39

[16] Imbens, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. The Review of Economics and Statistics, 86, pp. 4–29

[17] Yanovitzky, I., Zanutto, E., & Hornik, R. (2005). *Estimating Causal Effects of Public Health Education Campaigns using Propensity Score Methodology*. Journal Elsevier Evaluation and Program Planning 28 (2005) pp. 209–220.

[18] Poretsky, L. (2010). Principle of Diabetes Mellitus, Second Edition. New York: Springer

[19] Gungor, N., Hannon, T., Libman, I., Bacha, F., & Arslanian, S. (2005). *Type 2 Diabetes Mellitus in Youth: The Complete Picture to Date.* Pediatric Clinics of North America