CAUCHY – Jurnal Matematika Murni dan Aplikasi
Volume 10 (2) (2025), Pages 941-956
p-ISSN: 2086-0382; e-ISSN: 2477-3344

Application of Propensity Score Matching for Analyzing Factors Contributing to Pre-Diabetes

Oktaviani Aisyah Putri[1], Anwar Fitrianto[1*, 2], and Aam Alamudi[1]

[1] *Study Program Statistics and Data Science, School of Data Science, Mathematics, and Informatics, IPB University, Indonesia*
[2] *Institute of Engineering Mathematics, Universiti Malaysia Perlis, Malaysia*

Abstract

Inappropriate comparisons between control and treatment groups can be caused by overlapping factors, usually called confounders. Propensity score methods help reduce bias from measured confounding by summarizing the distribution of multiple measured confounders into a single score, based on the probability of receiving treatment. This study applies binary logistic regression to estimate propensity scores and identify risk factors that significantly influence complications in fasting blood glucose levels. Nearest Neighbor Matching (NNM) is used with various caliper and score orders to determine the most effective combination in reducing bias. The dataset consists of 692 records of lecturer health examination data collected in 2023 from IPB University. The results show that gender becomes a confounding variable. Both the order of propensity scores and caliper selection affect the outcome of the matching process. The matching condition using propensity scores in a random order with a caliper performed the best, achieving approximately 99.964% bias reduction. The significance of the average treatment effect for treated (ATT), all condition order with caliper indicates that gender have a positive relationship and significantly affects fasting blood glucose levels. Also, based on the matching results with the best combination, it indicates that age, academic position, structural position, education level, and lecturer performance do not influence abnormal fasting blood sugar (FBS). This could imply that other unmeasured variables, such as dietary habits, stress level, genetic predisposition, or physical activity might play a more dominant role in influencing fasting blood sugar levels than institutional roles or professional background factors.

Keywords: binary logistic regression; caliper; fasting blood glucose levels; propensity score matching (PSM); structural positions

1 Introduction

In several case studies, comparisons between control and treatment groups are often inaccurate due to the lack of well defined control groups and the presence of overlapping factors, commonly known as confounding [1]. Confounding refers to bias in estimating the effect of a risk factor on an outcome due to the presence of extraneous variables that are associated with both the treatment and the outcome [2]. Although confounding can often be mitigated through randomized

*Corresponding author. E-mail: anwarstat@gmail.com

Submitted: April 27, 2025 Reviewed: June 25, 2025 Accepted: September 21, 2025
DOI: https://doi.org/10.18860/cauchy.v10i2.32754

sampling, in practice, randomization becomes challenging when many covariates are involved, making it difficult to find individuals with similar or identical characteristics. If left unaddressed, confounding can obscure the true relationship between risk factors and outcomes, leading to either underestimation or overestimation of effects and ultimately, invalid conclusions[3]. Therefore, propensity scores can be used to minimize bias due to confounding variables with a single score based on the conditional probability of treatment[4]. Then, each observational data is paired by adjusting propensity scores based on the same or similar covariates (X_i) between the treatment group ($Z_i = 1$) and control group ($Z_i = 0$)[2]. Among various techniques, propensity score matching (PSM) method was the best method for reducing the influence of bias in comparing treatment and control groups compared to other propensity score techniques in individual studies[5]. For instance, propensity score matching (PSM) using binary logistic regression was able to reduce bias by 57.1 percent with a standard error of 0.103[2]. Additionally, simulations using Nearest Neighbor Matching (NNM) with caliper, followed by binary logistic regression, demonstrated a strong ability to minimize bias[6]. Also, the order of propensity scores can affect the performance of reducible bias. Caliper is the maximum difference limit of propensity scores used to limit treatment individuals to be paired; in this study, the caliper is obtained by multiplying 0.2 with the standard deviation of propensity scores[7]. However, limited research has examined the application of this method in the context of health risks in academic environments, particularly in the analysis of non-communicable disease risk factors among university lecturers.

The quality of higher education institution is partly determined by the performance of the lecturers[8]. In accordance with Law number 12 of 2012, lecturers are responsible for fulfilling the Tri Dharma Higher Education, which includes education, research, and community service. Article 30 paragraph 4 of the Minister of Education and Culture Regulation Number 3 of 2020 states that the workload of lecturers is calculated based on the equivalent of full teaching time. Their academic position directly impacts their teaching load and workload. The higher their academic position are expected to have accumulate higher credit points and often hold leadership positions, which can lead to additional institutional duties outside classroom teaching. For instance, doctoral level lecturers usually teach higher level courses (such as master's or doctoral level) and possibly teach fewer classes compared to master's degree lecturers. These responsibilities may result in longer working hours, inadequate rest, irregular eating patterns, and increased stress, all of which are potential contributors to deteriorating health.

Such lifestyle factors, combined with insufficient participation in regular medical check-ups (MCU), can lead to undetected chronic conditions and decreased job satisfaction [9]. Health is a critical determinant of employee productivity [10], yet awareness of preventive healthcare remains low. For instance, in 2023, only 692 out of 1,312 lecturers at IPB University (approximately 52%) underwent medical check-ups [11]. MCUs play an essential role in the early detection of non-communicable diseases, such as type 2 diabetes mellitus (DM). A precursor to this condition is prediabetes, characterized by elevated blood glucose levels that have not yet reached diabetic thresholds. Fasting blood sugar (FBS) tests are standard for screening this condition, with normal values ranging between 70–110 mg/dL. If undiagnosed or untreated, prediabetes may progress to full-blown diabetes. In 2023, MCU results, abnormal FBS was the second most prevalent health issue, affecting 351 lecturers. Notably, blood sugar levels are also associated with age and gender [12].

This study contributes scientifically by offering empirical evidence on the effectiveness of PSM variations in minimizing bias due to confounding in health risk studies, and practically by identifying significant risk factors of abnormal fasting blood sugar (FBS) among lectures in higher education institutions after establishing matched pairs.

2 Methods

This section systematically presents the data sources used, details of variables and the analytical approaches applied. Following the explanation of the research background and objectives in the previous section, this section serves as a crucial foundation for ensuring the validity of the results obtained. By detailing the characteristics of the data and the statistical methods employed, readers are expected to gain a clear understanding of the research process and the basis for concluding.

2.1 Data

The data used is secondary data obtained from the IPB Polyclinic and the Directorate of Human Resources at IPB University. These data consist of medical check-up (MCU) results and IPB lecturers' workload information. The number of lecturers who underwent MCU is 692 in 2023, with the following details of variables used:

Table 1: Variable in the study

Code	Variable Name	Description
FBS	Fasting Blood Sugar	As response variables 1 = Abnormal >110 mg/dl, 0 = Normal
Z	Gender	1 = Female, 0 = Male
-	Age	Years
Instructor Assist. Prof. Assoc. Prof. Prof.	Academic Position	4 dummy variables: 1 = Instructor, 0 = Others 1 = Assistant Professor, 0 = Others 1 = Associate Professor, 0 = Others 1 = Professor, 0 = Others
-	Lecturer Performance	Average performance over 1 year
-	Structural Position	1 = Yes, 0 = No
-	Education Level	1 = Doctoral, 0 = Master's

2.2 Analysis Procedure

The analytical procedures in this research were carried out through the following steps (Fig. 1):

a) Perform data pre-processing (categorizing each individual's fasting blood sugar level (FBS) as normal or abnormal).

b) Determine confounding variables.
Confounding variables, denoted as Z, were identified using the Chi-square test with the null hypothesis stating that there is no significant relationship between the variables potentially acting as confounders (Z). The Chi-square test was conducted using the following Eq. 1 [13]:

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{(O_{ij} - \hat{E}_{ij})}{\hat{E}_{ij}}; \hat{E}_{ij} = \frac{O_{i.} \times O_{.j}}{O_{..}} \tag{1}$$

Rejection rule for the null hypothesis if $\chi^2_{calculated} > \chi^2_{\alpha;(r-1);(c-1)}$ or P-value < alpha (α) 5%.

c) Estimate propensity scores using logistic regression with the maximum likelihood estimation method. Propensity scores use a logistic regression model with a binary dependent variable (Z_i = 1 for treatment and Z_i = 0 for control). Risk factors as predictors and confounders as responses because the response variable is not included during matching[14], [15]. The propensity score will be obtained using the following Eq. 2:

$$\pi(x) = \frac{exp(g(x))}{1 + exp(g(x))} \tag{2}$$

$$ln\left(\frac{\pi(x_i)}{(1-\pi(x_i))}\right) = g(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \tag{3}$$

$$e\left(x_i\right) = P(Z_i = 1 |\, X_i = x_i) = \frac{exp\left(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k\right)}{1 + exp\left(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k\right)} \tag{4}$$

where $0 < e(x_i) = P(Z_i = 1|X_i = x_i) < 1$ for each $x \in X$ is the conditional probability value of a group based on the observed covariates.

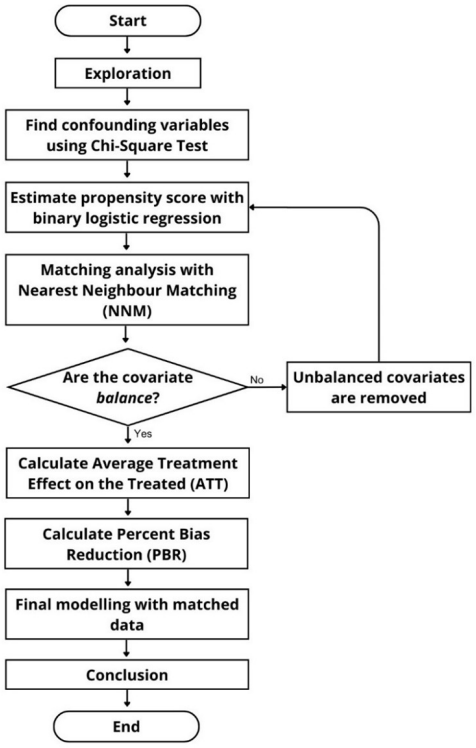


Figure 1: Research flow chart

d) Conduct a matching analysis using the Nearest Neighbor Matching (NNM) algorithm to pair individuals in the treatment group with those in the control group. Unmatched individuals in the treatment group are eliminated and not included in the next stage of analysis. In this step, combinations of caliper and score orders are tested to identify the most effective configuration for reducing bias. Caliper is used to limit the difference in propensity scores between the treatment group and the control group. The smaller caliper, the more likely that more observations will be discarded or not used in the next step. Caliper equal to 0.2 times the standard deviation of its propensity score is considered effective in minimizing estimation bias caused by confounding variables[7].

e) Perform post-matching, testing the balance of variables ($rchi^2$) and estimating Average Treatment of Treated (ATT).
Balance testing uses the t-test if numeric and Z-test if categorical. If there are unbalanced covariates, they will be removed from the model. If the covariates used are balanced, then the Average Treatment of Treated (ATT) estimation can be continued. ATT is used to determine how much influence Z has on the occurrence of Y when the influence of other variables (covariates) has been reduced to confounding variables (Z)[15], [16]. The test statistic for estimation with the following Eq. 5:

$$t_{hit} = \frac{\hat{\theta}}{SE(\hat{\theta})} \tag{5}$$

f) Calculate Percent of Bias Reduction (PBR).
Calculate the percent bias reduction (PBR) to know how much bias can be reduced with the following Eq. 6 [17]:

$$PBR = \frac{|B_{before\ PSM} - B_{after\ PSM}|}{|B_{before\ PSM}|} \times 100\% \tag{6}$$

B is the difference in means between the treatment and control groups for each covariate, written as $B = P_1(x_p) - P_0(x_p)$ while $P_1(x_p)$ and $P_0(x_p)$ were proportions of the covariate for the groups.

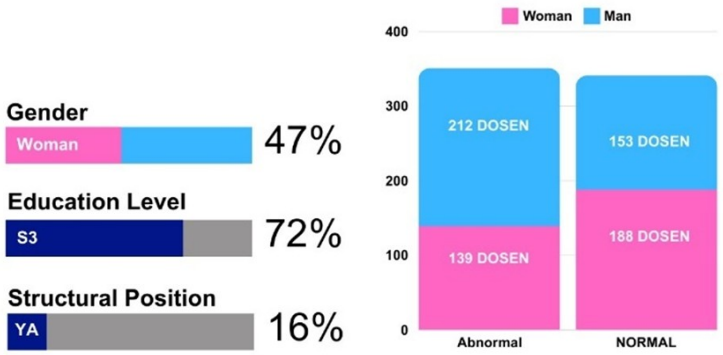
g) Interpreted and draw conclusions

3 Results and Discussion

This section presents the empirical findings and their interpretation. We begin with a brief exploratory overview of the data, summarising lecturer characteristics and fasting blood sugar (FBS) outcomes. We then identify potential confounders of the relationship between structural position (treatment) and FBS (outcome) using Chi–Square dependence tests, and retain gender as the primary confounder. Next, we estimate propensity scores with a binary logistic regression model and implement nearest–neighbour matching using a caliper set to 0.2 times the standard deviation of the propensity score. After matching, we assess covariate balance between treatment and control groups. We then quantify the Average Treatment effect on the Treated (ATT) and compute percentage bias reduction to evaluate how well matching reduces systematic differences. Finally, we report the post–matching outcome model and discuss the substantive implications of the results.

3.1 Data Exploration

The data used consists of lecturers with both civil servant (PNS) and non-civil servant (Non-PNS) employment status. Out of a total of 1,312 lecturers, only 692 lecturers, or approximately 53 percent, underwent medical check-up (MCU). From the total lecturers who underwent MCU, 51 percent had abnormal Fasting Blood Sugar (FBS) test results above 110 ml/dl.

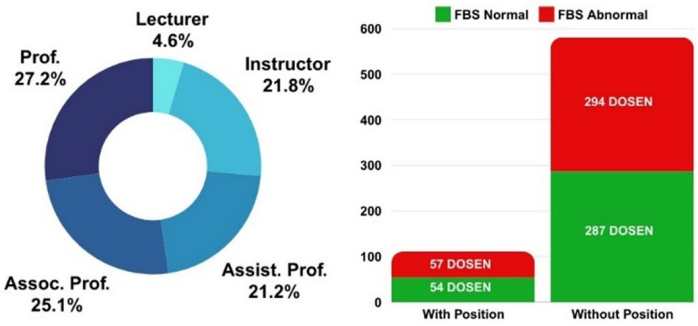


(a) **(b)**

Figure 2: Proportion of lecturers based on (a) gender, education level, and structural position; and (b) gender in relation to blood sugar

Fig. 2a shows that the proportion of MCU lecturers based on education level and structural position is unbalanced. This becomes interesting to see if it can be addressed after matching

analysis. Unlike other variables, the proportion of MCU lecturers based on gender does not much different between male and female proportions, with the proportion of females being 3 percent less than males who underwent MCU. However, Fig. 2b shows that lecturers with abnormal FBS are actually more likely to be male. Meanwhile, in the theory, women have a higher chance of developing diabetes due to higher increases in lipid levels (blood fat) compared to men, increasing the risk factor for diabetes mellitus in women 3-7 times higher than in men, which is 2-3 times. High fat levels can reduce body cell sensitivity and make it difficult for insulin to deliver glucose into cells[18].



(a) **(b)**

Figure 3: Proportion of lecturers based on (a) academic position, and (b) structural position in relation to blood sugar

Fig. 3a shows that the largest proportion of academic position is instructor. However, those the largest proportion abnormal FBS results are Assist. Prof., with 101 out of 174 lecturers, or 58 percent. Additionally, Fig. 3b shows that both lecturers with and without structural positions have 51 percent abnormal FBS results. Among lecturers with positions, 57 out of 111 have abnormal FBS, and among those without positions, 294 out of 581 have abnormal FBS.

Table 2: Characteristics of lecturers based on age and lecturer performance

Variable	Mean	Standard Deviation	Min	Max
Age	51	11	28	70
lecturer performance	29.805	10.551	1.405	7.855

The minimum and maximum observations far from the mean, suggest the possibility of extreme values or outliers. Additionally, the lecturers ages range from 28 to 70 years, with an average age is 51 years. The lecturer performance score is calculated based on the average of 2 semesters, with an average lecturer performance is 29.805. This score falls into the medium category and meets the incentive requirements. The requirement for receiving an incentive is the lecturer performance score at least 12. This means that with a minimum score of 1.405, does not receive an incentive. This could happen if the lecturer is ill or on leave for certain reasons and thus cannot meet the requirements for receiving an incentive. The maximum of lecturer performance score reaching 75.855 proves that there are lecturers who are assumed to have activities far beyond the target, even excessive, which may pose a risk to their health.

3.2 Selection of Confounder Variable

Before performing matching analysis, a Chi-Square test needs to be conducted to determine the categorical variables that confound the influence of structural position on fasting blood sugar (FBS).

Table 3: Determining confounding variables of structural position against other covariates

Variables	Chi-Square	P-value	Decision
Gender	7.222	0.007*	Reject H0
Instructor	0.466	0.495	Fail to Reject H0
Assist. Prof.	0.006	0.936	Fail to Reject H0
Assoc. Prof.	3.284	0.070	Fail to Reject H0
Prof.	0.277	0.598	Fail to Reject H0
Education Level	14.926	0.000*	Reject H0

*significant at 5% level

Table 4: Determining confounding variables of each covariate against response Fasting Blood Sugar Level (FBS)

Variables	Chi-Square	P-value	Decision
Gender	16.121	0.000*	Reject H0
Instructor	4.955	0.026*	Reject H0
Assist. Prof.	0.021	0.883	Fail to Reject H0
Assoc. Prof.	4.604	0.032*	Reject H0
Prof.	2.177	0.140	Fail to Reject H0
Structural Position	0.002	0.967	Fail to Reject H0
Education Level	9.681	0.002*	Reject H0

*significant at 5% level

Table 3 presents the dependency test between structural position and other covariates with a significance level of 5 percent. The requirement to be a confounding variable is that the covariate has a significant relationship with other covariates and also the response variable. Variables that have a significant relationship with structural position are gender and education level. Table 4 shows that gender, academic position of assist. Prof, academic position of Assoc. Prof., and education level have a significant relationship with the FBS variable because they have p-value < 0.05. However, in this study, the variable chosen as the confounding variable is gender (Z) because according to Ariani et al. (2022) [18], abnormal blood sugar is influenced by gender, as the risk factor for diabetes in women is 3-7 times higher compared to men, which is 2-3 times.

3.3 Estimation of Propensity Score with Logistics Regression Model

The propensity score value is often known as the logistic regression model symbolized by $\phi(x_i)$ as in Eq. 2, while the propensity score value is symbolized by $e(x_i)$, so the propensity score value will be obtained if the binary logistic regression model parameters have been obtained. The results of parameter estimation of propensity score values with binary logistic regression can be seen in Table 5.

Table 5: Parameter estimation of binary logistic regression model between covariates (X) and gender (Z)

Covariate	Parameter(β)	SE	p-value	Odds Ratio
Intercept	1.245	0.610	0.041	3.474
Age	-0.027	0.011	0.015*	0.974
Instructor	0.004	0.406	0.991	1.005
Assist. Prof.	0.173	0.438	0.692	1.189
Assoc. Prof.	0.224	0.475	0.638	1.251
Prof.	-0.105	0.514	0.837	0.900
Structural Position	-0.736	0.233	0.002*	0.479
Education Level	-0.153	0.228	0.503	0.858
Lecturer Performance	0.005	0.008	0.501	1.005

*significant at 5% level

Table 5 shows that the age variable and structural position have a negative and significant

relationship with gender (Z) with p-value of 0.015 and 0.002 respectively. Additionally, it is known that probability female lecturer holding a position is lower than the male lecturer. Next, the propensity score $e(x_i)$ is obtained using the following Eq. 7:

$$e\left(x_i\right) = P(Z_i = 1 | \, X_i = x_i) = \frac{exp\left(\hat{g}\left(x_i\right)\right)}{1 + exp\left(\hat{g}\left(x_i\right)\right)} \tag{7}$$

where,

$$\begin{aligned} \hat{g}\left(x_i\right) = & 1.245 - 0.027 \; Age \; + 0.004 \; Instructor + 0.173 \; Assist. \; Prof. + \\ & 0.224 \; Assoc. \; Prof. - 0.105 \; Prof. - 0.736 \; Structural \; Position - \\ & 0.153 \; Education \; Level + 0.005 \; Lecturer \; Performance \end{aligned} \tag{8}$$

Based on the logit model above, the probability calculation function can be obtained as follows.

$$\pi\left(x\right) = \frac{e^{\left(\begin{array}{l}(1.245 - 0.027Age + 0.004Instructor + 0.173Assist.Prof. + \\ \quad 0.224Assoc.Prof. - 0.105Prof. - 0.736StructuralPosition - \\ \quad 0.153EducationLevel + 0.005LecturerPerformance)\end{array}\right)}}{1 + e^{\left(\begin{array}{l}(1.245 - 0.027Age + 0.004Instructor + 0.173Assist.Prof. + \\ \quad 0.224Assoc.Prof. - 0.105Prof. - 0.736StructuralPosition - \\ \quad 0.153EducationLevel + 0.005LecturerPerformance)\end{array}\right)}} \tag{9}$$

The propensity score values range from 0.199 to 0.678. Then, the propensity score values are given 3 different treatments: observations sorted based on scores from smallest to largest, scores from largest to smallest, and random scores. The values for each individual propensity score are as follows. Randomization was done by taking 3 times in Microsoft Excel.

Table 6: Propensity score estimates for each observation using binary logistic regression, ordered from smallest to largest

Propensity score estimates for each observation									
0.199	0.218	0.219	0.221	0.221	0.226	0.239	0.240	0.241	0.242
0.244	0.251	0.252	0.269	0.270	0.272	0.272	0.276	0.278	0.284
0.285	0.294	0.295	0.298	0.299	0.300	0.303	0.303	0.308	0.308
0.309	0.312	0.316	0.318	0.318	0.319	0.322	0.323	0.325	0.326
0.326	0.327	0.327	0.329	0.329	0.331	0.333	0.334	0.335	0.336
0.336	0.337	0.337	0.339	0.340	0.341	0.341	0.342	0.343	0.343
0.344	0.346	0.347	0.347	0.347	0.347	0.348	0.348	0.349	0.350
0.350	0.351	0.351	0.352	0.352	0.352	0.352	0.354	0.354	0.355
0.355	0.356	0.357	0.357	0.357	0.358	0.359	0.359	0.360	0.361
0.361	0.361	0.362	0.362	0.363	0.363	0.363	0.363	0.363	0.363
...	...	...	...	...	...	...	...	...	etc

3.4 Matched Analysis

The goodness of propensity score matching is seen from how much bias can be reduced. Therefore, matching analysis needs to be done first to determine the bias after matching. The matching method used is nearest neighbor matching and the caliper used is the result of 0.2 times the standard deviation of the propensity score, which is 0.022. Observations that do not get their matched (pair) will not be used for further analysis.

Table 7: Propensity score estimates for each observation using binary logistic regression, ordered from largest to smallest

Propensity score estimates for each observation									
0.678	0.671	0.667	0.663	0.663	0.663	0.661	0.660	0.656	0.656
0.656	0.655	0.653	0.652	0.650	0.649	0.648	0.647	0.644	0.643
0.642	0.639	0.639	0.639	0.638	0.635	0.633	0.633	0.633	0.632
0.631	0.629	0.629	0.628	0.626	0.625	0.625	0.625	0.624	0.624
0.623	0.623	0.622	0.622	0.622	0.621	0.621	0.620	0.620	0.620
0.620	0.620	0.619	0.619	0.619	0.618	0.617	0.617	0.617	0.616
0.615	0.615	0.614	0.613	0.613	0.613	0.613	0.613	0.612	0.611
0.611	0.611	0.610	0.609	0.609	0.609	0.608	0.608	0.608	0.608
0.608	0.607	0.607	0.607	0.607	0.607	0.607	0.606	0.606	0.602
0.602	0.601	0.600	0.598	0.598	0.598	0.598	0.598	0.598	0.597
...	...	...	...	...	...	...	...	...	etc

Table 8: Propensity score estimates for each observation using binary logistic regression, random order

Propensity score estimates for each observation									
0.392	0.285	0.389	0.617	0.507	0.507	0.452	0.528	0.432	0.431
0.445	0.369	0.550	0.504	0.419	0.494	0.565	0.460	0.560	0.382
0.465	0.272	0.638	0.270	0.389	0.555	0.628	0.368	0.626	0.447
0.374	0.446	0.424	0.424	0.357	0.453	0.660	0.508	0.611	0.395
0.379	0.358	0.388	0.612	0.409	0.571	0.609	0.359	0.399	0.469
0.369	0.499	0.399	0.347	0.564	0.471	0.489	0.371	0.384	0.639
0.574	0.375	0.459	0.362	0.464	0.583	0.394	0.465	0.557	0.610
0.501	0.569	0.341	0.380	0.545	0.533	0.388	0.447	0.419	0.422
0.333	0.562	0.378	0.481	0.425	0.361	0.607	0.361	0.434	0.389
0.398	0.295	0.421	0.443	0.542	0.532	0.434	0.545	0.416	0.560
0.392	0.285	0.389	0.617	0.507	0.507	0.452	0.528	0.432	0.431
...	...	...	...	...	...	...	...	...	etc

Table 9: Number of pairs formed

	Before matching	**After matching**	**Not Matched**
	Propensity score order from smallest to largest		
Treatment, without caliper	327	327	0
Control, without caliper	365	327	38
Treatment, with caliper	327	233	94
Control, with caliper	365	233	132
	Propensity score order from largest to smallest		
Treatment, without caliper	327	327	0
Control, without caliper	365	327	38
Treatment, with caliper	327	233	94
Control, with caliper	365	233	132
	Propensity score for random order		
Treatment, without caliper	327	327	0
Control, without caliper	365	327	37
Treatment, with caliper	327	235	92
Control, with caliper	365	235	130

Table 9 shows that different orders without caliper produce the same number of pairs, but when using caliper, the number of pairs obtained is different. In the post-matching condition with scores ordered from smallest to largest and with caliper, the characteristics of matched lecturer pairs, if viewed based on abnormal FBS, there are 235 lecturers and normal FBS there are 231 lecturers. Meanwhile, if viewed based on structural positions, there are 65 lecturers holding positions and 401 lecturers not holding positions. Table 9 also shows the number of matched groups or groups that are continued to the next analysis and unmatched groups or groups that are eliminated and not continued to the next analysis. The results of individuals or

observations that have been paired are as follows.

Table 10: Individuals which obtained matches with propensity scores sorted from smallest to largest

Number	Without Caliper		Number	With Caliper	
	Treatment	Control		Treatment	Control
1	10	8	1	10	9
2	101	105	2	101	105
3	102	100	3	102	100
4	104	103	4	104	103
5	109	108	5	109	108
6	110	111	6	110	111
7	113	114	7	113	114
8	117	115	8	117	115
9	118	116	9	118	116
10	12	9	10	122	123
...	...	...	...	...	etc

In the first ten pairs formed, it can be seen that there are observations that get different control matches. However, many have already obtained the same matches between using a caliper and not using a caliper. This proves that ordering propensity scores from smallest to largest can facilitate finding suitable matches.

Table 11: Individuals which obtained matches with propensity scores sorted from largest to smallest

Number	Without Caliper		Number	With Caliper	
	Treatment	Control		Treatment	Control
1	692	691	1	593	586
2	683	674	2	583	580
3	593	532	3	579	578
4	592	525	4	576	577
5	591	523	5	573	571
6	590	522	6	572	570
7	589	519	7	568	566
8	588	518	8	567	564
9	585	516	9	565	559
10	584	512	10	680	679
...	...	...	...	...	etc

None of the first ten pairs formed obtained the same match between using a caliper and without a caliper. This may occur because it makes it difficult to find suitable matches.

Table 12: Individuals who obtained matches with random order propensity score

Number	Without Caliper		Number	With Caliper	
	Treatment	Control		Treatment	Control
1	339	338	1	339	338
2	292	268	2	292	287
3	130	127	3	130	127
4	270	272	4	270	272
5	360	365	5	360	365
6	92	91	6	92	91
7	320	321	7	320	321
8	273	143	8	261	257
9	261	257	9	59	58
10	59	58	10	250	249
...	...	...	...	...	etc

Table 12 proves that random selection of individuals can obtain suitable matches. There are 7

out of 10 pairs that get the same match between using a caliper and without a caliper.

3.5 Covariate Balance Test

Before conducting the ATT significance test, it is necessary to perform a covariate balance test on the confounding variables between the treatment group and the control group. The null hypothesis used to test the difference in means for numeric variables is that there is no difference in the mean of the covariate (X) between the treatment group and the control group. Additionally, the null hypothesis used to test the difference in proportions for categorical variables is that there is no difference in the proportion of the covariate (X) between the treatment group and the control group.

Table 13 would help assess whether the matching process has successfully balanced the covariates between the treatment and control groups. A p-value greater than the chosen significance level (typically 0.05) would indicate that there is no significant difference between the groups for that covariate, suggesting successful balance.

Table 13: Results of covariate balance test

Covariate	Balance	p-value		
		Smallest to largest	Largest to smallest	Random
Age	Before matching	0.000	0.000	0.000
	After, without caliper	0.007	0.007	0.007
	After, with caliper	0.772	0.772	0.697
Instructor	Before matching	0.05	0.05	0.05
	After, without caliper	0.162	0.162	0.162
	After, with caliper	0.736	0.736	0.162
Assist. Prof.	Before matching	0.476	0.476	0.476
	After, without caliper	0.931	0.931	0.794
	After, with caliper	0.919	0.919	1
Assoc. Prof	Before matching	0.697	0.697	0.697
	After, without caliper	0.474	0.474	0.421
	After, with caliper	0.834	0.834	0.328
Prof.	Before matching	0.002	0.002	0.002
	After, without caliper	0.157	0.157	0.131
	After, with caliper	0.905	0.905	0.306
Lecturer Performance	Before matching	0.824	0.824	0.824
	After, without caliper	0.960	0.960	0.930
	After, with caliper	0.410	0.410	0.478
Structural Position	Before matching	0.005	0.005	0.005
	After, without caliper	0.142	0.142	0.173
	After, with caliper	0.894	0.894	1
Education Level	Before matching	0.003	0.003	0.003
	After, without caliper	0.026	0.026	0.033
	After, with caliper	0.835	0.835	0.026

3.6 Estimated Average Treatment of Treated (ATT)

Before testing the significance of the ATT, it is necessary to perform a covariate balance test on the confounding variables between the treatment group and the control group. The null hypothesis used to test the mean difference in numerical variables is that there is no difference in the mean of the covariate (X) between the treatment group and the control group. Additionally, the null hypothesis used to test the difference in proportions for categorical variables is that there is no difference in the proportion of the covariate (X) between the treatment group and the control group. Table 13 shows that the balance for the variables age, employment status, structural position, and last education can be addressed when using a caliper during matching. Only three conditions with caliper are used for further analysis. The next step is to test the significance of the Average Treatment of Treated (ATT) using Eq. 5, with results shown in Table 14 below.

Table 14: Estimated Average Treatment of Treated (ATT)

	ATT (θ)	**SE (ATT)**	t_{hit}	**p-value**
Caliper, score smallest to largest	-0.148	0.047	-3.171	0.002*
Caliper, score largest to smallest	-0.121	0.047	-2.593	0.010*
Caliper, score random	-0.172	0.048	-3.577	0.000*

*significant at 5% level

Table 14 shows that in all the three combinations of propensity scores with caliper, the gender variable (Z) significantly affects FBS, as evidenced by p-value < 0.05.

3.7 Percentage Bias Reduction (PBR)

This section, we calculate the percent bias reduction using Eq. 6, with results as follows.

Table 15: Bias before and after matching

	Before Matching	**After Matching**	**PBR**
	Propensity score order from smallest to largest		
Without caliper	0.042	0.025	40.476%
With caliper	0.042	0.001	97.619%
	Propensity score order from largest to smallest		
Without caliper	0.022	0.025	-13.636%
With caliper	0.022	0.001	95.455%
	Propensity score random order		
Without caliper	0.042	0.025	40.476%
With caliper	0.042	0.000	99.964%

Table 15 shows that the bias in the random propensity score order with caliper can reduce 99.964 percent of the bias, much better compared to without caliper, which reduces 40.476 percent. It is also proven that caliper reduces bias well for different matching order conditions. Interestingly, the bias before matching in the largest to smallest propensity score order without caliper is smaller than the bias after matching. This proves that matching will seek the closest and earliest values. In this condition, the bias increases by 13.636 percent.

3.8 Final Model

Using the result of matched data with a combination of caliper and random order propensity score, the following estimate parameter is obtained.

Table 16: Estimated parameters of the binary logistic regression model between covariate (X) and fasting blood sugar (FBS)

Kovariat	**Parameter (β)**	**SE**	**p-value**	**Odd Ratio**
Intercept	-1.204	0.758	0.112	0.300
Age	0.012	0.014	0.405	1.012
Academ.Staff.	0.463	0.515	0.368	1.589
Senior.Lect.	0.454	0.553	0.412	1.575
Assoc.Prof.	0.562	0.587	0.339	1.753
Prof.	0.529	0.636	0.405	1.698
Structural Position	-0.191	0.303	0.529	0.826
Last Education	0.246	0.276	0.372	1.279
Lecturer Performance	-0.001	0.010	0.881	0.999

*significant at 5% level

$$\begin{aligned}\widehat{FBS} =& \breve{}1.204 + 0.012Age + 0.463Academ.Staff. + 0.454Senior.Lect.+ \\ & 0.562Assoc.Prof. + 0.529Prof. - 0.191StructuralPosition+ \\ & 0.246EducationLevel\breve{}0.001LecturerPerformance\end{aligned} \tag{10}$$

Based on the logit model above, the probability calculation function can be obtained as follows.

$$\pi(x) = \frac{e^{\left(\begin{array}{l}(1.204 + 0.012Age + 0.463Academ.Staff. + 0.454Senior.Lect.+ \\ 0.562Assoc.Prof. + 0.529Prof. - 0.191StructuralPosition+ \\ 0.246EducationLevel - 0.001LecturerPerformance)\end{array}\right)}}{1 + e^{\left(\begin{array}{l}(1.204 + 0.012Age + 0.463Academ.Staff. + 0.454Senior.Lect.+ \\ 0.562Assoc.Prof. + 0.529Prof. - 0.191StructuralPosition+ \\ 0.246EducationLevel - 0.001LecturerPerformance)\end{array}\right)}} \tag{11}$$

From Eq. 10, based on the matching results with the best combination, it was found that the variables of age, academic position, structural position, education level, and lecturer performance do not have an influence on abnormal fasting blood sugar (FBS). This could imply that other unmeasured variables, such as dietary habits, stress level, genetic predisposition, or physical activity might play a more dominant role in influencing fasting blood sugar levels than institutional roles or professional background factors.

4 Conclusion

Research results show that the variable of gender is a confounding variable. This confounding variable subsequently becomes a response variable, and a binary logistic regression model indicates that structural position has a positive and significant relationship with gender. Moreover, the order of propensity score values appears to affect the results of the matching analysis, as the matching process essentially seeks the closest and earliest values, although the differences are not substantial. Additionally, the use of a caliper can effectively reduce bias, making it sufficient to perform matching with a caliper without needing to consider the order of propensity scores. The matching condition using propensity scores in a random order with a caliper performed the best, achieving approximately 99.964% bias reduction. The caliper also balances the covariates for gender (Z). The results of the average treatment effect of the treated with propensity scores from smallest to largest, largest to smallest, and random, both with a caliper, show that the gender variable significantly affects FBS. Based on the matching results with the best combination, it was found that the variables of age, academic position (Academic Staff, Senior Lecturer, Associate Professor, and Professor), structural position, education level, and lecturer performance do not have significant relationships with abnormal fasting blood sugar (FBS). This could imply that other unmeasured variables, such as dietary habits, stress level, genetic predisposition, or physical activity might play a more dominant role in influencing fasting blood sugar levels than institutional roles or professional background factors.

Acknowledgments

We would like to thank all parties involved in the completion of this article, especially Director of Human Resource IPB University and the Head of IPB Health Unit for providing the data used in this research.

Authors Contribution

Oktaviani Aisyah Putri derived the models and collected and analyzed the data. Anwar Fitrianto designed the model, main conceptual frameworks and ideas. Aam Alamudi helped in developing the model and data analysis. All authors' have contributed in writing the manuscript.

Data and Code Availability

The data for this article are confidential but may be obtained with Data Use Agreements with the IPB University. Researchers interested in access to the data may contact the Health Clinic of the IPB University.

Competing Interest Statement

We, all authors, have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. We declare no conflict of interest.

References

[1] J. Kahlert, S. B. Gribsholt, H. Gammelager, O. M. Dekkers, and G. Luta, "Control of confounding in the analysis phase - an overview for clinicians," eng, *Clinical Epidemiology*, vol. 9, pp. 195–204, 2017. DOI: 10.2147/CLEP.S129886.

[2] N. Faulina, K. Nisa, D. Aziz, and E. Setiawan, "Analisis propensity score matching pada kejadian diabetes melitus yang memuat faktor confounding," id, *Jurnal Siger Matematika*, vol. 2, no. 2, pp. 43–51, Sep. 2021, Number: 2. DOI: 10.23960/jsm.v2i2.2810. Accessed: Jul. 24, 2025. Available online.

[3] M. Webster-Clark et al., "Using propensity scores to estimate effects of treatment initiation decisions: State of the science," en, *Statistics in Medicine*, vol. 40, no. 7, pp. 1718–1735, 2021, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/sim.8866. DOI: 10.1002/sim.8866. Accessed: Jul. 24, 2025. Available online.

[4] F. Cottone, A. Anota, F. Bonnetain, G. S. Collins, and F. Efficace, "Propensity score methods and regression adjustment for analysis of nonrandomized studies with health-related quality of life outcomes," en, *Pharmacoepidemiology and Drug Safety*, vol. 28, no. 5, pp. 690–699, 2019, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/pds.4756. DOI: 10.1002/pds.4756. Accessed: Jul. 24, 2025. Available online.

[5] P. C. Austin and T. Schuster, "The performance of different propensity score methods for estimating absolute effects of treatments on survival outcomes: A simulation study," eng, *Statistical Methods in Medical Research*, vol. 25, no. 5, pp. 2214–2237, Oct. 2016. DOI: 10.1177/0962280213519716.

[6] H. Olmuş, E. Beşpınar, and E. Nazman, "Performance evaluation of some propensity score matching methods by using binary logistic regression model," *Communications in Statistics - Simulation and Computation*, vol. 51, no. 4, pp. 1647–1660, Apr. 2022, Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/03610918.2019.1679181. DOI: 10.1080/03610918.2019.1679181. Accessed: Jul. 24, 2025. Available online.

[7] M. Ohmori et al., "Propensity score-matched analysis of endoscopic resection for recurrent colorectal neoplasms: A pilot study," en, *Journal of Gastroenterology and Hepatology*, vol. 36, no. 9, pp. 2568–2574, 2021, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jgh.15519. DOI: 10.1111/jgh.15519. Accessed: Jul. 24, 2025. Available online.

[8] S. T. Simanjuntak and C. T. Wahyanti, "Analisis kinerja dosen melaksanakan Tri Dharma Perguruan Tinggi pada masa pandemi covid-19: Studi kasus di Universitas Kristen Satya Wacana," en, *JENIUS (Jurnal Ilmiah Manajemen Sumber Daya Manusia)*, vol. 5, no. 1, pp. 86–101, Sep. 2021, Number: 1. DOI: 10.32493/JJSDM.v5i1.13144. Accessed: Jul. 24, 2025. Available online.

[9] W. Kusgiyanto, S. Suroto, and E. Ekawati, "Analisis hubungan beban kerja fisik, masa kerja, usia, dan jenis kelamin terhadap tingkat kelelahan kerja pada pekerja bagian pembuatan kulit lumpia di Kelurahan Kranggan Kecamatan Semarang Tengah," en-US, *Jurnal Kesehatan Masyarakat*, vol. 5, no. 5, pp. 413–423, Oct. 2017, Number: 5 Publisher: Fakultas Kesehatan Masyarakat Universitas Diponegoro. DOI: 10.14710/jkm.v5i5.18963. Accessed: Jul. 24, 2025. Available online.

[10] O. Mungkasa, "Bekerja dari rumah (working from home/WFH): Menuju tatanan baru era pandemi COVID 19," en, *Jurnal Perencanaan Pembangunan: The Indonesian Journal of Development Planning*, vol. 4, no. 2, pp. 126–150, Jun. 2020, Number: 2. DOI: 10.36574/jpp.v4i2.119. Accessed: Jul. 24, 2025. Available online.

[11] I. A. S. P. Wati, A. K. Sari, and D. Lubis, "Perilaku deteksi dini penyakit tidak menular pada dosen di Fakultas Kedokteran Universitas Udayana," id, *Archive of Community Health*, vol. 4, no. 1, pp. 86–96, Aug. 2019, Number: 1. DOI: 10.24843/ACH.2017.v04.i01.p11. Accessed: Jul. 24, 2025. Available online.

[12] A. Rudi and H. Kwureh, "Faktor risiko yang mempengaruhi kadar gula darah puasa pada pengguna layanan laboratorium," *Jurnal Ilmiah Kesehatan : Wawasan Kesehatan*, vol. 3, no. 2, pp. 33–39, Dec. 2017. DOI: 10.31227/osf.io/d3kes_v1.

[13] S. H. Waego H. Nugroho, "Building multiclass classification model of logistic regression and decision tree using the chi-square test for variable selection method," en-US, *Journal of Hunan University Natural Sciences*, vol. 49, no. 4, May 2022, Number: 4. Accessed: Jul. 24, 2025. Available online.

[14] R. Febrianti, Y. Widyaningsih, and S. Soemartojo, "The parameter estimation of logistic regression with maximum likelihood method and score function modification," en, *Journal of Physics: Conference Series*, vol. 1725, no. 1, p. 012 014, Jan. 2021, Publisher: IOP Publishing. DOI: 10.1088/1742-6596/1725/1/012014. Accessed: Jul. 24, 2025. Available online.

[15] Z. Kang et al., "Psoriasis increases the risk of Sjögren's syndrome: Evidence from a propensity score-matched cohort study and transcriptomic analysis," *BMC Medicine*, vol. 23, no. 1, p. 26, Jan. 2025. DOI: 10.1186/s12916-025-03856-y. Accessed: Jul. 24, 2025. Available online.

[16] R. D. Sururin Nufus, B. Susetyo, B. Sartono, and Efriwati, "Exploring the potential impact of ginger consumption on the duration of COVID-19 recovery: A propensity score matching using random forest," en, *IOP Conference Series: Earth and Environmental Science*, vol. 1359, no. 1, p. 012 139, Jun. 2024, Publisher: IOP Publishing. DOI: 10.1088/1755-1315/1359/1/012139. Accessed: Jul. 24, 2025. Available online.

[17] S. Hasanah, B. W. Otok, and A. Adeni, "Perbandingan metode propensity score matching-support vector machine dan propensity score matching-regresi logistik biner pada kasus HIV/AIDS," en, *Sainmatika: Jurnal Ilmiah Matematika dan Ilmu Pengetahuan Alam*, vol. 18, no. 1, pp. 93–107, Jun. 2021, Number: 1. DOI: 10.31851/sainmatika.v17i3.4925. Accessed: Jul. 24, 2025. Available online.

[18] N. Ariani, R. Alfian, and E. Prihandiwati, "Tingkat perilaku pengobatan, kepatuhan minum obat, dan kadar gula darah pasien diabetes mellitus rawat jalan di RSUD Brigjend. H. Hasan Basry Kandangan," en-US, *Jurnal Ilmiah Manuntung: Sains Farmasi Dan Kesehatan*, vol. 8, no. 1, pp. 156–162, May 2022. DOI: 10.51352/jim.v8i1.523. Accessed: Jul. 24, 2025. Available online.