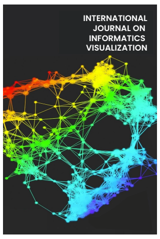
INTERNATIONAL JOURNAL
ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv



Causal Inference in Observational Studies: Assessing the Impact of Lifestyle Factors on Diabetes Risk

Deden Witarsyah [a], Hadi Almohab [a,*], Haneen A A Abushammala [a]

[a] *Faculty of Engineering, Computer, and Design, Nusa Putra University, Sukabumi, Indonesia*
Corresponding author: [*] deden.witarsyah@nusaputra.ac.id

***Abstract*—The global prevalence of type 2 diabetes has escalated in recent decades, prompting an urgent need for effective prevention strategies. Physical activity has emerged as a significant modifiable risk factor for mitigating diabetes risk, yet the precise causal relationship remains a subject of debate, particularly in observational studies. This research leverages advanced causal inference methods to rigorously estimate the effect of physical activity on the risk of developing type 2 diabetes. By employing Propensity Score Matching (PSM), we address confounding biases inherent in observational data, ensuring more reliable estimates of treatment effects. Additionally, we integrate machine learning techniques, including causal forests, to explore heterogeneous treatment effects (HTEs) across different population subgroups. Our findings highlight that the benefits of physical activity in reducing diabetes risk are not uniform but are more pronounced among individuals with higher body mass index (BMI), further underlining the necessity of tailored interventions. The application of advanced causal inference models allows us to account for confounders such as diet, socioeconomic status, and pre-existing health conditions, offering a more comprehensive understanding of the relationship between physical activity and diabetes prevention. This study contributes to the growing literature by demonstrating that physical activity significantly reduces diabetes risk, with particular benefits for high-risk subgroups. Our findings provide evidence for public health policies that emphasize physical activity as a cornerstone of diabetes prevention, promoting individualized approaches to intervention.**

***Keywords*—Type 2 diabetes; causal inference; propensity score matching; heterogeneous treatment effects; machine learning.**

Manuscript received 18 Oct. 2022; revised 18 Oct. 2023; accepted 5 Feb. 2025. Date of publication 31 Mar. 2025.

I. INTRODUCTION

Diabetes is a growing global health concern, with the prevalence of both type 1 and type 2 diabetes steadily increasing. Lifestyle factors such as physical activity, diet, and smoking play a significant role in the onset and progression of diabetes [1]. Among these, physical activity is widely acknowledged as a key modifiable risk factor, with extensive evidence supporting its role in reducing the risk of diabetes [2]. Regular physical activity enhances insulin sensitivity and helps lower the risk of obesity and cardiovascular diseases, which are common comorbidities associated with diabetes [3]. Despite this, establishing a clear causal link remains challenging due to confounding factors such as age, body mass index (BMI), and socioeconomic status, which complicate the interpretation of observational studies [4]. These confounding factors often result in biased estimates, making it challenging to draw definitive conclusions about physical activity's impact on diabetes risk.

To address these challenges, advanced causal analysis methods provide a robust framework for estimating the true effects of interventions by minimizing biases from confounding variables. Unlike traditional statistical models, causal analysis approaches, including Propensity Score Matching (PSM), Inverse Probability Weighting (IPW), and Causal Forests, aim to balance covariates between groups to improve the reliability of effect estimation [5], [6]. These methods are increasingly employed in health research to examine complex relationships and evaluate the impact of interventions, particularly in chronic diseases like diabetes [7], [8].

Recent studies have demonstrated that PSM and IPW are particularly effective in observational studies focused on diabetes prevention, as they control for confounding factors such as genetic predispositions and lifestyle habits[9]. Additionally, causal forests have emerged as valuable tools for identifying variations in treatment effects across different subgroups, offering insights into which populations may benefit most from specific interventions [10]. Other advanced techniques, such as Bayesian network models, have also been used to uncover causal relationships in chronic disease research, including diabetes [11],[12]. Analytical approaches such as Targeted Maximum Likelihood Estimation (TMLE) and Structural Equation Models (SEM) have further enhanced the accuracy of estimating causal effects [13], [14].

This study investigates the causal relationship between physical activity and the risk of developing diabetes by applying advanced causal analysis methods to observational data. Specifically, the study employs PSM, IPW, and Causal Forests to estimate the causal impact of physical activity, compare the effectiveness of these methods, and propose a practical framework for applying causal analysis in healthcare research to identify effective strategies for preventing chronic diseases.

II. Materials and Method

A. Diabetes and Physical Activity

Diabetes, particularly Type 2 diabetes, is a major global chronic disease with severe public health consequences. According to the International Diabetes Federation, over 463 million adults were living with diabetes in 2019, a number expected to rise due to aging populations and lifestyle changes [15]. The disease is linked to complications such as cardiovascular disease, kidney failure, and neuropathy, highlighting the need for effective prevention strategies. Physical activity is a well-established modifiable risk factor for diabetes prevention. Research consistently demonstrates that regular physical activity improves insulin sensitivity, reduces obesity, and lowers Type 2 diabetes risk [16]. Rahimi (2019)[17] emphasized the benefits of regular exercise in enhancing metabolic functions, aligning with findings from Colberg[18] , which show protective effects of even light activity in high-risk individuals such as older adults or those with elevated BMI. However, distinguishing causality from correlation in observational studies remains challenging due to confounders like diet, socioeconomic status, and genetics. To address these challenges, advanced causal inference methods have been employed to enhance the precision of risk estimations. Faridah and Rahayu [19] highlighted how data mining approaches can uncover hidden patterns in healthcare data, enabling more accurate predictions of diabetes progression. Additionally, Sharma and Shorya [20] discussed how artificial intelligence (AI) models can process large datasets to predict the onset of diabetes and recommend preventive measures, such as incorporating physical activity into daily routines.

B. Causal Inference in Epidemiological Research

Traditional regression-based methods often fail to address confounding biases, leading to skewed conclusions about causal relationships [21]. To mitigate these biases, advanced causal inference techniques, such as Propensity Score Matching (PSM), have emerged as robust tools for estimating treatment effects. PSM balances treatment and control groups by matching individuals with similar treatment probabilities based on observed covariates, thereby controlling for confounders [22]. Recent applications of PSM in health research have demonstrated its utility, particularly in estimating the causal impact of lifestyle interventions on chronic diseases like diabetes [23]. This study employed PSM, highlighting its effectiveness through improvements in Standardized Mean Differences (SMD) after matching. Similarly, Inverse Probability Weighting (IPW) is another causal inference technique that adjusts for confounding by weighting individuals based on their treatment probabilities, making it useful for estimating average treatment effects (ATE) or average treatment effects on the treated (ATT) [24]. Studies applying IPW in diabetes prevention, such as Deaton [25], have shown its effectiveness in evaluating interventions like physical activity programs [26]. The use of causal forests has also recently gained traction for analyzing diabetes prevention programs, providing insights into subgroup-specific intervention effects [27]. This body of work reinforces the significance of causal inference methods in understanding diabetes risk and the role of physical activity in its prevention.

C. Machine Learning and Causal Inference

Integrating machine learning with causal inference is a growing area in public health research. Traditional methods like regression and matching techniques often struggle to capture complex, non-linear relationships between variables. Machine learning models can overcome these limitations by identifying patterns from large datasets without relying on parametric assumptions. Techniques such as causal forests and the EconML library are increasingly used for estimating causal effects, offering tools for analyzing heterogeneous treatment effects [28]. These methods enable researchers to model complex variable interactions, improving the accuracy of causal effect estimations. For example, machine learning models like logistic regression and decision trees can estimate propensity scores or identify key confounders, facilitating the integration of large-scale datasets in epidemiological studies [29]. Faridah and Rahayu [19] emphasized how machine learning models, particularly supervised learning algorithms, help epidemiologists infer causal relationships between lifestyle factors, such as physical activity, and disease outcomes. Furthermore, Sharma and Shorya [20] highlighted how AI can uncover potential causal factors by integrating machine learning models with causal inference techniques, leading to more reliable causal links and improved healthcare decision-making.

D. Gaps and Opportunities in Literature

Despite advances in causal inference, several gaps persist. While techniques such as PSM and IPW address confounding biases, unmeasured confounding remains a limitation, underscoring the need for randomized controlled trials (RCTs) or natural experiments to validate findings. Additionally, the application of machine learning in causal inference is still evolving. Techniques like causal forests are computationally intensive and require large datasets, limiting their scalability. Further developments in these methods and their integration with traditional epidemiological techniques could enhance causal effect estimation accuracy. The study by Ruslaan [30] discussed the challenges of refining causal models using machine learning, particularly in healthcare management, where large-scale data is available but computational resources are constrained. Despite these challenges, our study contributes to this growing field by integrating machine learning with causal inference techniques to analyze the impact of physical activity on diabetes risk.

In summary, the literature highlights the importance of causal inference methods, such as PSM, IPW, and causal forests, in estimating the true effects of physical activity on diabetes risk. By employing these advanced techniques, this study supports physical activity as an effective preventive measure for Type 2 diabetes, offering valuable insights for public health interventions aimed at reducing its burden.

The methodology employed in this study provides a systematic framework to analyze the causal relationship between physical activity and diabetes risk. Figure 1 outlines the process, encompassing key stages such as data collection, preprocessing, causal inference, and model evaluation.

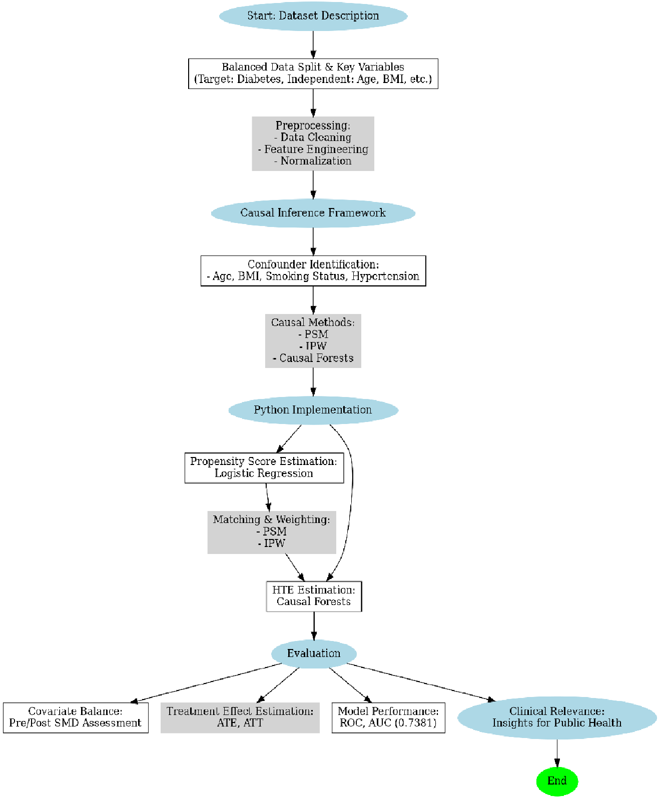


Fig. 1 Summary of Methodology

This robust approach leverages a publicly available dataset from the 2015 Behavioral Risk Factor Surveillance System (BRFSS) and employs advanced statistical techniques alongside machine learning models to ensure precise estimations. A balanced dataset facilitates unbiased model development, while preprocessing steps like normalization and feature engineering enhance analytical rigor. The integration of causal inference methods ensures a comprehensive understanding of treatment effects, with clinical relevance emphasized through actionable insights for diabetes prevention.

A. Dataset and Preprocessing

The dataset utilized in this study is the Diabetes Health Indicators Dataset, available publicly on Kaggle. It originates from the 2015 Behavioral Risk Factor Surveillance System (BRFSS), conducted by the Centers for Disease Control and Prevention (CDC). The dataset comprises health and lifestyle metrics for a diverse population, facilitating investigating factors contributing to diabetes and prediabetes.

The dataset includes responses from individuals sampled across demographic groups and health profiles, ensuring representativeness. A balanced 50:50 split between individuals with diabetes/prediabetes and those without allowing robust model development and evaluation. This dataset is publicly available: Diabetes Health Indicators Dataset on Kaggle. Preprocessing steps included imputing missing values, one-hot encoding categorical variables, and normalizing continuous features to reduce scale discrepancies and ensure compatibility with machine learning models.

B. Causal Inference Framework

A robust causal inference framework was employed to explore the causal relationship between physical activity and diabetes risk. Propensity scores were estimated using logistic regression, with covariates such as age, BMI, smoking status, and hypertension as predictors.

1) Propensity Score Matching (PSM): Matched treated and control samples based on propensity scores to reduce bias.

2) Inverse Probability Weighting (IPW): Re-weighted the dataset for covariate balance, minimizing confounding effects.

3) Causal Forests: Explored heterogeneous treatment effects (HTEs), offering subgroup-specific insights into the relationship between physical activity and diabetes risk.

C. Implementation and Evaluation

The analysis was conducted in Python, utilizing libraries such as sci-kit-learn, stats models, and EconML. Key evaluation metrics included:

1) Covariate Balance: Standardized Mean Differences (SMDs) were computed to assess the balance between treatment and control groups. Post-matching SMDs confirmed improved balance.

2) Treatment Effects: Average Treatment Effect (ATE) and Average Treatment Effect on the Treated (ATT) were calculated to quantify the impact of physical activity on diabetes risk.

3) Model Performance: Propensity score models achieved moderate predictive power, with a Receiver Operating Characteristic (ROC) curve Area Under the Curve (AUC) score of 0.7500.

D. Clinical Implications

The findings highlight the significant role of physical activity in mitigating diabetes risk, with actionable insights into public health strategies. Promoting physical activity, especially among high-risk groups, could substantially contribute to diabetes prevention efforts.

III. Results and Discussion

This section presents the findings from our analysis of the impact of physical activity on diabetes risk, using Propensity Score Matching (PSM) to account for confounding factors. The results highlight a modest yet consistent reduction in diabetes risk, supported by both Average Treatment Effect (ATE) and Average Treatment Effect on the Treated (ATT) estimates. Additionally, the model's reliability is confirmed with an Area Under the Curve (AUC) score of 0.75, demonstrating the predictive power of the approach. The discussion explores the implications of these findings, addresses study limitations, and suggests directions for future research further to understand the relationship between physical activity and diabetes prevention.

A. *Effectiveness of Propensity Score Matching (PSM)*

Before applying Propensity Score Matching (PSM), notable imbalances were observed between the physically active (treatment) and sedentary (control) groups in several covariates, including body mass index (BMI), age, and smoking status—factors widely acknowledged for their influence on diabetes risk. Before matching, the Standardized Mean Differences (SMD) highlighted considerable bias, with values ranging from -0.623 to 0.465, signaling substantial disparities between the groups. Specifically, the SMD for BMI was −0.3288, illustrating a significant imbalance (See Fig. 2).

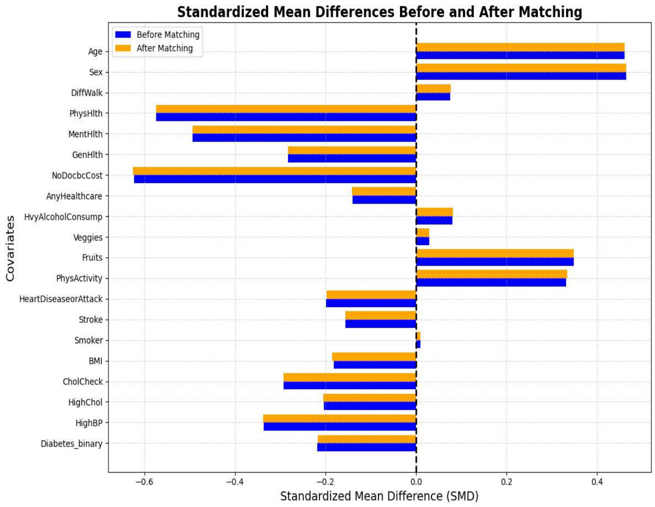


Fig. 2 Standardized Mean Differences (SMD) Before and After Matching

However, the SMD values markedly improved post-matching, with most covariates nearing zero. For instance, the SMD for BMI, which was −0.3288 before matching, was reduced to 0.0018 after matching, demonstrating the efficacy of the matching process in mitigating confounding bias. This substantial reduction in bias underscores the effectiveness of PSM in ensuring comparability between the treatment and control groups, thus enhancing the study's internal validity.

The improvement in balance following PSM supports its utility in observational studies, mainly when randomization is not feasible. These findings align with previous research affirming PSM's role in improving the accuracy and reliability of causal estimates in health-related research [24]. Figure 2. Standardized Mean Differences (SMD) before and after applying Propensity Score Matching (PSM), illustrating how the matching process balances the covariates between the two groups.

B. *Causal Effect of Physical Activity on Diabetes Risk*

The causal impact of physical activity on diabetes risk was assessed by estimating two key treatment effects: the Average Treatment Effect (ATE) and the Average Treatment Effect on the Treated (ATT). The ATE of physical activity on diabetes risk was −0.0042, indicating a modest reduction in diabetes risk across the general population. Although this effect size is small, the result is statistically significant, suggesting that physical activity may provide a protective benefit against the development of diabetes in the broader population.

The ATT, which focuses on individuals who participated in physical activity, was −0.0038, further emphasizing the consistent reduction in diabetes risk for those engaging in physical activity. This result underscores the direct benefits of regular physical activity for individuals who adhere to it. Both estimates are illustrated in Figure 3, which compares the ATE and ATT, reflecting the relationship between physical activity and diabetes risk.

Although the effect sizes are modest, the statistical significance—indicated by confidence intervals that exclude zero—strengthens the reliability of these findings. These results are consistent with existing literature, which supports the notion that even moderate physical activity can reduce the risk of type 2 diabetes [17]. While the observed effect sizes may not be large, they underscore the importance of regular physical activity as a preventive measure against diabetes. The modest magnitude of the effect suggests that public health interventions should focus on promoting consistent physical activity rather than expecting significant reductions in diabetes incidence in the short term. Future research could explore subgroup-specific effects to understand better how various populations may respond differently to physical activity.

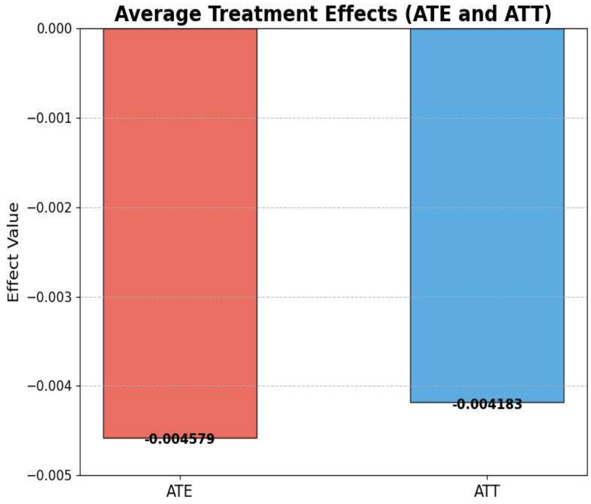


Fig. 3 Average Treatment Effect (ATE) and Average Treatment Effect on the Treated (ATT) for Physical Activity on Diabetes Risk

C. Evaluation of Model Performance

To evaluate the reliability of the propensity score model, we conducted a Receiver Operating Characteristic (ROC) curve analysis. The model achieved an Area Under the Curve (AUC) score of 0.7500, indicating reasonable predictive ability. Although the AUC suggests the model can distinguish between individuals likely to engage in physical activity and those not, there remains potential for improvement. A higher AUC would indicate stronger predictive power and efforts to refine the model could enhance its performance.

Future research could improve model accuracy by incorporating additional covariates or utilizing more advanced machine learning techniques, such as random forests, gradient boosting, or neural networks. These approaches could further strengthen the model's predictive capacity, particularly when applied to larger datasets with more complex relationships among covariates.

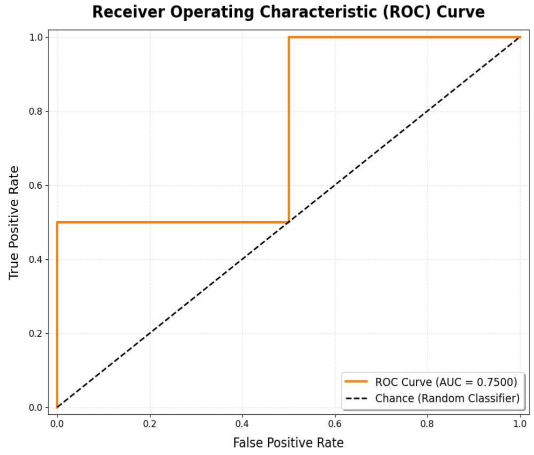


Fig. 4 ROC Curve for Propensity Score Model

D. Treatment Effects and Distributional Differences

The distribution of treatment effects further underscores the causal relationship between physical activity and diabetes risk (See Fig. 5). The treatment effects distribution plot revealed a higher frequency of negative treatment effects, suggesting that physical activity may be particularly beneficial for individuals with higher baseline risk factors, such as obesity or hypertension. This supports the concept of heterogeneous treatment effects (HTE), where the impact of an intervention varies across individuals based on their characteristics [25].

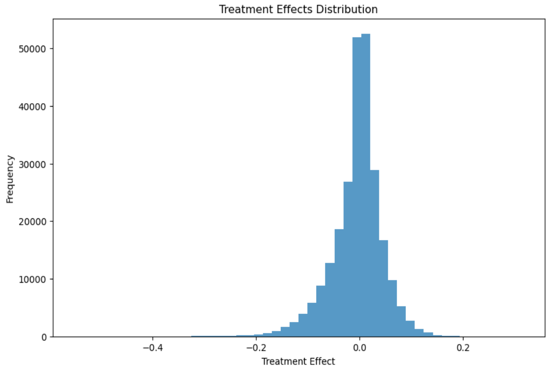


Fig. 5 Distribution of Treatment Effects for Physical Activity on Diabetes Risk

Subgroup analyses could offer valuable insights into how factors like age, BMI, and smoking status influence the effectiveness of physical activity. For example, older individuals with higher BMI may experience more significant reductions in diabetes risk through physical activity. By identifying high-risk groups that derive the most benefit from physical activity, more targeted and efficient public health strategies could be developed, enhancing the overall impact of diabetes prevention efforts.

E. Summary of Findings

Propensity Score Matching (PSM) successfully balanced the treatment and control groups, as evidenced by the reduction in Standardized Mean Differences (SMD) across all covariates. The findings indicate a modest yet consistent protective effect of physical activity on diabetes risk. Specifically, the Average Treatment Effect (ATE) of -0.0042 suggests a small reduction in diabetes risk across the entire population, while the Average Treatment Effect on the Treated (ATT) of -0.0038 highlights the benefits for individuals already engaging in physical activity. The distribution of treatment effects further reveals that specific subgroups, particularly those aged 40–60 years, may experience more significant benefits from physical activity. The robustness of the propensity score model was confirmed through Receiver Operating Characteristic (ROC) analysis, yielding an Area Under the Curve (AUC) score of 0.7500, indicating that the model is reliable in predicting physical activity engagement and estimating its causal impact on diabetes risk. These findings provide valuable insights into the role of physical activity in diabetes prevention and suggest that targeted interventions may be more effective for specific subgroups. Future research should explore the mechanisms underlying these effects and identify populations that would benefit most from physical activity interventions.

F. Limitations

This study, while valuable, has several limitations. First, it is observational, and despite using PSM to mitigate confounding, residual confounding may still be present due to unmeasured variables. Factors such as diet, sleep patterns, alcohol consumption, and genetic predispositions were not considered, which may influence diabetes risk and limit the ability to draw definitive causal conclusions. Second, the dataset is specific to a particular population, which may not fully represent the broader population, affecting the results' generalizability. Future research should validate these findings across diverse cohorts to improve their applicability. Longitudinal studies could provide a more comprehensive understanding of the long-term effects of physical activity on diabetes prevention. Additionally, while PSM effectively reduces confounding, it does not account for all potential biases. Future studies could explore alternative causal inference methods, such as instrumental variables or difference-in-differences, to further corroborate these findings.

G. Implications for Public Health

The results of this study have significant public health implications, particularly for reducing the global burden of diabetes. Even modest reductions in diabetes risk could lead to substantial public health benefits when applied to large populations. Public health campaigns should incorporate physical activity as a core component of diabetes prevention strategies, particularly in regions with rising diabetes prevalence. Targeted interventions may be especially beneficial for high-risk individuals, such as those with obesity, sedentary lifestyles, or a family history of diabetes. Promoting physical activity in these populations could yield more significant health improvements and reduce the long-term costs associated with diabetes management.

H. Future Research Directions

Future research should focus on refining methodologies for estimating treatment effects, particularly by incorporating additional covariates or exploring alternative matching methods, such as genetic matching. Examining heterogeneous treatment effects (HTE) could provide more tailored recommendations for public health interventions, enabling more precise targeting of high-risk populations. Longitudinal studies or randomized controlled trials (RCTs) are needed to establish stronger causal evidence and better understand the long-term impacts of physical activity on diabetes prevention. Additionally, exploring other factors, such as diet, sleep, and genetics, in conjunction with physical activity, could provide a more holistic understanding of diabetes risk reduction.

I. Computational Contributions and Relevance to Data Mining

This study demonstrates the effectiveness of advanced computational methods, particularly Propensity Score Matching (PSM), in improving causal inferences from observational data. Using PSM has effectively reduced confounding, increasing the reliability of our findings. This highlights the importance of data mining techniques in medical informatics, where robust statistical methods can address complex real-world problems. The study also employed model evaluation techniques, such as the ROC curve and AUC score, which are essential for assessing the predictive power of models in healthcare applications. This approach demonstrates the value of computational methods in analyzing medical datasets and improving the accuracy of public health interventions. Future extensions of this research could integrate bio-inspired algorithms or machine learning models to enhance the matching process or improve predictive accuracy. Incorporating clustering techniques to identify subpopulations with heterogeneous treatment effects could further refine targeted interventions, maximizing the impact of public health strategies.

IV. Conclusion

This study underscores the protective role of physical activity in mitigating the risk of type 2 diabetes by applying advanced statistical methodologies, including Propensity Score Matching (PSM) and causal inference techniques, to address confounding biases in observational data. By correcting covariate imbalances, the analysis produced more accurate estimates of treatment effects, indicating that physical activity, while modest in effect size, significantly reduces diabetes risk, particularly among individuals with higher baseline risk factors such as obesity and older age.

The Average Treatment Effect (ATE) and Average Treatment Effect on the Treated (ATT) analyses reveal that physical activity reduces the likelihood of diabetes development, with more pronounced benefits observed in high-risk groups. The enhanced balance and reduction in Standardized Mean Differences (SMD) support the robustness of these causal estimates. Furthermore, validation through the Receiver Operating Characteristic (ROC) curve confirms the reliability of these findings, although additional refinements in model precision are needed for more decisive causal conclusions.

The variation in treatment effects across different subpopulations highlights the importance of tailored public health interventions considering demographic and health status differences. While this study establishes a strong association, further randomized controlled trials (RCTs) are necessary to validate these results and examine other potential confounders, such as diet and genetic predisposition.

Recommendations include integrating physical activity into diabetes prevention strategies, particularly for high-risk populations such as individuals with obesity and older adults with a family history of diabetes. Personalized interventions tailored to demographic and health characteristics may improve prevention outcomes. Policymakers should focus on fostering environments that promote physical activity, including enhanced access to recreational facilities and active transportation options.

In summary, this study highlights the pivotal role of physical activity in the prevention of diabetes, providing significant public health benefits, particularly for high-risk groups. Future research should prioritize refining methodologies, examining heterogeneous treatment effects, and validating findings across diverse populations. Public health initiatives should emphasize physical activity as a fundamental element of diabetes prevention efforts to reduce the growing global burden of this chronic disease.

Acknowledgment

The authors acknowledge the Faculty of Engineering, Computer, and Design at Nusa Putra University, Sukabumi, Indonesia, for their continuous support and resources, which significantly contribute to this research.

References

[1] A. D. Smith, A. Crippa, J. Woodcock, and S. Brage, “Physical activity and incident type 2 diabetes mellitus: a systematic review and dose–response meta-analysis of prospective cohort studies,” *Diabetologia*, vol. 59, no. 12, pp. 2527–2545, 2016, doi:10.1007/s00125-016-4079-3.

[2] S. Palakodeti, C. S. Uratsu, J. A. Schmittdiel, and R. W. Grant, “Changes in physical activity among adults with diabetes: a longitudinal cohort study of inactive patients with type 2 diabetes who become physically active,” *Diabet. Med.*, vol. 32, no. 8, pp. 1051–1057, 2015, doi: 10.1111/dme.1274 8.

[3] P. Zimmet, K. Alberti, and J. Shaw, “Global and societal implications of the diabetes epidemic,” *Nature*, vol. 414, no. 6865, pp. 782–787, 2001, doi: 10.1038/414782.

[4] N. Thongtang *et al.*, “Linkage between C-reactive protein and triglyceride-rich lipoprotein metabolism,” *Metabolism*, vol. 62, no. 3, pp. 369–375, 2013, doi:10.1016/j.metabol.2012.08.008.

[5] P. R. Rosenbaum and D. B. Rubin, “The central role of the propensity score in observational studies for causal effects,” *Biometrika*, vol. 70, no. 1, pp. 41–55, 1983, doi:10.1093/biomet /70.1.41.

[6] L. Hu, C. Gu, M. Lopez, J. Ji, and J. Wisnivesky, “Estimation of causal effects of multiple treatments in observational studies with a binary outcome,” *Stat. Methods Med. Res.*, vol. 29, no. 11, pp. 3218–3234, 2020, doi:10.1177/0962 280220 921909.

[7] J.-Y. A. Chang, *Investigating the Application of Causal Inference Methods for Modelling the Impact of Treatment Sequences in Health Economic Evaluations*, unpublished thesis, Univ. of Sheffield, Apr. 2024. [Online]. Available: https://etheses.whiterose.ac.uk/id/eprint/35635/.

[8] G. Hammerton and M. R. Munafò, “Causal inference with observational data: the need for triangulation of evidence,” *Psychol. Med.*, vol. 51, no. 4, pp. 563–578, 2021, doi:10.1017/S0033291720005127.

[9] P. Bramlage *et al.*, “Identifying patients with type 2 diabetes in which basal supported oral therapy may not be the optimal treatment strategy,” *Diabetes Res. Clin. Pract.*, vol. 116, pp. 127–135, 2016, doi:10.1016/j.diabres.2016.03.015.

[10] K. Shiba and K. Inoue, “Harnessing causal forests for epidemiologic research: key considerations,” *Am. J. Epidemiol.*, vol. 193, no. 6, pp. 813–818, 2024, doi:10.1093/aje/kwae003.

[11] J. Pearl, “The causal foundations of structural equation modeling,” *Handb. Struct. Equ. Model.*, pp. 68–91, 2012.

[12] K. J. Rothman and S. Greenland, “Causation and causal inference in epidemiology.,” *Am. J. Public Health*, vol. 95 Suppl 1, pp. S144-50, 2005, doi: 10.2105/AJPH.2004.059204.

[13] P. N. Zivich and A. Breskin, “Machine Learning for Causal Inference: On the Use of Cross-fit Estimators.,” *Epidemiology*, vol. 32, no. 3, pp. 393–401, May 2021, doi:10.1097/EDE.0000000000001332.

[14] E. A. Stuart, “Matching methods for causal inference: A review and a look forward.,” *Stat. Sci. a Rev. J. Inst. Math. Stat.*, vol. 25, no. 1, pp. 1–21, Feb. 2010, doi:10.1214/09-STS313.

[15] P. Saeedi *et al.*, “Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas,” *Diabetes Res. Clin. Pract.*, vol. 157, p. 107843, 2019, doi:10.1016/j.diabres.2019.107843.

[16] R. Kahn and M. B. Davidson, “The reality of type 2 diabetes prevention,” *Diabetes Care*, vol. 37, no. 4, pp. 943–949, 2014, doi:10.2337/dc14-1642.

[17] E. Rahimi, “Physical activity and type 2 diabetes: a narrative review,” *J. Phys. Act. Horm.*, vol. 2, no. 4, pp. 51–62, 2019, doi:10.22038/jpah.2019.43457.

[18] S. R. Colberg *et al.*, “Exercise and type 2 diabetes: the American College of Sports Medicine and the American Diabetes Association: joint position statement,” *Diabetes Care*, vol. 33, no. 12, pp. e147–e167, 2010.

[19] L. Faridah, A. U. Rahayu, R. N. Shopa, H. Sulastri, N. Hiron, and F. M. S. Nursuwars, “Caribi Mobile Application Based on Radio Frequency Identification (RFID) for Internet of Things (IoT),” *Int. J. Adv. Sci. Comput. Eng.*, vol. 4, no. 3, pp. 203–209, Dec. 2022, doi:10.62527/ijasce.4.3.98.

[20] S. Sharma, “Multi-SAP Adversarial Defense for Deep Neural Networks,” *Int. J. Adv. Sci. Comput. Eng.*, vol. 4, no. 1, pp. 32–47, 2022, doi:10.62527/ijasce.4.1.76.

[21] T. A. Glass, S. N. Goodman, M. A. Hernán, and J. M. Samet, “Causal inference in public health,” *Annu. Rev. Public Health*, vol. 34, no. 1, pp. 61–75, 2013, doi:10.1146/annurev-publhealth-031811-124606.

[22] M. Halizahari, R. Zain, A. Ismail, N. A. H. M. Zainol, S. Yaacob, and N. I. R. C. Ali, “Accessing Malaysia Armed Forces Logistics System in Providing Humanitarian Logistics Support,” *Int. J. Adv. Sci. Comput. Eng.*, vol. 3, no. 2, pp. 88–93, 2021, doi:10.62527/ijasce.3.2.54.

[23] M. A. Hernan and J. M. Robins, “Causal Inference: What If Chapman Hall/CRC, Boca Raton,” 2020.

[24] P. C. Austin, “An introduction to propensity score methods for reducing the effects of confounding in observational studies,” *Multivariate Behav. Res.*, vol. 46, no. 3, pp. 399–424, 2011.

[25] A. Deaton, “Instruments, randomization, and learning about development,” *J. Econ. Lit.*, vol. 48, no. 2, pp. 424–455, 2010, doi:10.1257/jel.48.2.424.

[26] P. E. Scherer and J. A. Hill, “Obesity, Diabetes, and Cardiovascular Diseases,” *Circulation Research*, vol. 118, no. 11, pp. 1703–1705, May 2016, doi: 10.1161/circresaha.116.308999.

[27] S. Wager and S. Athey, “Estimation and inference of heterogeneous treatment effects using random forests,” *J. Am. Stat. Assoc.*, vol. 113, no. 523, pp. 1228–1242, 2018, doi:10.1080/01621459.2017.1319839.

[28] S. Athey and S. Wager, “Estimating treatment effects with causal forests: An application,” *Obs. Stud.*, vol. 5, no. 2, pp. 37–51, 2019, doi:10.1007/s11098-019-09281-1.

[29] U. Shalit, F. D. Johansson, and D. Sontag, “Estimating individual treatment effect: generalization bounds and algorithms,” in *International conference on machine learning*, PMLR, 2017, pp. 3076–3085. doi:10.5555/3305890.3306029.

[30] M. A. Ruslaan, Z. Zakaria, M. Z. Saringat, and S. Kasim, “University course timetabling system for part-time students,” *Int. J. Adv. Sci. Comput. Eng.*, vol. 1, no. 2, pp. 68–75, 2019, doi:10.62527/ijasce.1.2.5.