European Journal of Therapeutics
pISSN: 2564-7784
eISSN: 2564-7040

Eur J Ther. 2023;29(3):667-676.
https://doi.org/10.58600/eurjther1813

Review

An Introduction to Propensity Score Analysis: Checklist for Clinical Researches

Nazlı Totik Doğan[1], Sevinç Püren Yücel Karakaya[1], Zeliha Nazan Alparslan[1]

[1] Department of Biostatistics, Faculty of Medicine, Çukurova University, Adana, Turkey

Received: 2023-08-22 / **Accepted**: 2023-09-22 / **Publication Online**: 2023-09-23

Correspondence
Nazli Totik Dogan
Addresss: Department of Biostatistics, Faculty of Medicine, Çukurova University, Adana, Turkey
E-mail(s): ntotik@cu.edu.tr; nazlitotik.biostat@gmail.com

ABSTRACT

Background: Propensity score analysis is a widely used method to estimate treatment effect in dealing with the selection bias (i.e. lack of randomization) of observational studies. Although, there are relatively many guidelines in the literature for the adoption of this analysis, no checklists exist.

Objective: In this study, we propose a basic guideline for propensity score analysis, a tutorial that may be used to improve the quality of studies which implement this analysis. Additionally, in line with this guideline, we present an easy-to-use checklist which will assist researchers in the analysis process.

Conclusion: In light of the principles in this guideline/checklist, we propose that minor updates be considered for STROBE.

Keywords: Observational study, propensity score, treatment effect, selection bias, STROBE

INTRODUCTION

The most important characteristic of randomized clinical trials (RCTs), which are considered to be the gold standard for assessing the effect of treatments, is randomization [1,2]. On the other hand, RCTs have limitations due to temporal, financial, ethical, logistical or other reasons. To overcome these limitations, applied clinical researchers tend to plan observational studies that provide useful information for addressing health-related questions [3,4,5,6]. In RCTs, since there are randomly allocated similar treatment groups, the effect of treatments is often directly comparable; but, owing to covariate imbalance, such direct comparisons may not be possible in observational studies. Observational studies are frequently vulnerable to selection bias due to the lack of random treatment allocation, and this vulnerability leads to an imprecise estimate of the treatment effect [4,6,7,8]. To control selection bias, the propensity score(PS), described as "the conditional probability of assignment to a particular treatment given a vector of observed covariates" has first been proposed by Rosenbaum and Rubin in 1983 [9].

Over the past decade, PS analysis has gained popularity among clinical researchers. Examples of such studies include evaluating the effect of pancreatoduodenectomy(minimally invasive vs. open) on short-term outcomes among European centers [10]; the effectiveness of a placebo over another placebo in multiple sclerosis[11]; the effect of community pharmacy-based medication on death and readmission after hospital discharge[10,11,12]. Despite its popularity, there is a lack of complete guideline for PS analysis to help applied clinical researchers in the methodology literature [2,13,14]. Thus, in order to contribute to the methodology, we present basic guideline and checklist for the step-by-step implementation of PS analysis.

Stages of PS Analysis

Background

Randomization is not performed in observational studies. Thus, treatment groups generally not being comparable may result in confounding bias. If this happens, an appropriate adjustment should be made for confounding. As the PS summarizes all covariates into a single score, it decreases the potential for overfitting [15].

The PS is theoretically defined as a participant's probability of receiving the treatment conditional on the covariates at baseline. There are many settings in which the treatment may be binary, multinomial, ordinal, or continuous [16]. Let T_i be a binary treatment indicator variable (where T_i equals 1 if the participant is in the treatment group or 0 if the participant is in the control group) and x_i be a vector of observed covariates. Then, for each participant *i*, PS (e_i) is expressed as follows:

$e_i = \Pr(T_i | X_i)$

Strongly ignorable treatment assignment assumption has been proposed by Rosenbaum and Rubin to obtain unbiased treatment effects using PS. This assumption consists of two conditions:

(1) Conditional independence: $Y(1), Y(0) \perp T | X$

(2) Positivity: $0 < P(T=1 \mid X) < 1$

The first condition means that all confounding variables must be known. Lee and Little (2017) state that this condition cannot be tested empirically [13]. Instead, one must be persuaded that all crucial variables are measured in the study design. The second condition means that the individual had a non-null probability to receive treatment [5,6,8,17,18]. PS analyses generally consist of two phases: the design phase (phase I) and the analysis phase (phase II). In the design phase PSs are estimated using the PS model, and in the analysis phase the estimated PSs are used in the treatment effect model to adjust the effect of treatment. The type of the PS model depends on the nature of the treatment, and the treatment effect model on the outcome variable [5]. The stages of these phases are summarized in Table 1.

Table 1. Basic Stages of PS Analysis

Phase I	**Stage 1**	PS Model Building ○ Covariate Selection ○ PS Estimation ○ Checking Overlap ○ First Balance Control
Phase II	**Stage 2**	Application of PS Methods ○ Specifying Treatment Effect ○ PS Methods ○ PS Matching ○ PS Weighting ○ PS Stratification ○ PS Covariate Adjustment ○ Second Balance Control
	Stage 3	Treatment Effect Estimation

PS: Propensity score

Main Points:

- Providing detailed tutorial and design guide for PS analysis
- Contribution to the methodology with easy step-by-step PS analysis implementation
- Allowing to control the quality of PS analysis with the checklist
- Standardizing the reporting of future studies by following the checklist
- Emphasizing consideration of minor updates for STROBE.

Stage 1: PS Model Building

Stage 1 consists of covariate selection, PS estimation, checking overlap, first balance control subtitles. These are defined as follows, respectively.

Covariate Selection

The PS model uses treatment status rather than clinical outcome state as the dependent variable. There are many debates in the methodology literature regarding which variables to include in the PS model. Possible sets of variables that can be included in the PS model are baseline covariates, treatment-related covariates, outcome-related covariates, and both treatment and outcome-related covariates (The concept of "related"; Statistical significance in the literature; clinical significance; or statistical significance as determined by hypothesis tests performed on the data obtained in the current study).

Austin et al. (2007) highlight that excluding the confounding variable(s) in the PS model resulted in biased estimation of the treatment effect [7]. The same study also states that neglect to include a confounding variable in the PS model may lead

to biased estimation of the treatment effect. In addition, this study concludes, variables related to treatment but not related to outcome should not be included in the PS model, because they will not have an improving effect on the results obtained from the PS analysis [5,7]. Brookhart et al. (2006) have suggested that variables not related to treatment but related to outcome should always be included in the PS model, since this reduces the variance of the estimated treatment effect without increasing the bias [19]. In general, it is emphasized that, including the prognostically important covariates (that are outcome-related) or confounding variables (that are treatment and outcome-related) in the PS model should be preferred.

PS Estimation

The stage following the selection of the covariates is estimating the PSs for each participant in the sample. The estimated PS is the predicted probability of treatment obtained from the fitted model. In the case of a categorical treatment (e.g. A-drug vs B-drug), the primary parametric method used in PS estimation is logistic regression. The generalized boosted model (GBM), a nonparametric method proposed by McCaffrey (2004), which automatically includes all higher order and interaction terms of covariates and does not require full data, is frequently used for PS estimation in addition to logistic regression. Besides these methods, various parametric (e.g. probit regression, discriminant analysis) and nonparametric (e.g. tree-based methods, neural networks) methods are used in PS estimation in the literature [8,18,20,21].

Checking Overlap

The next stage after PS estimation is the evaluation of the overlap (also referred to as "common support" in some studies) of PS between treatment groups. How much the PSs overlap between treatment groups is a main issue facing investigators using PS. The similarity of PS distribution among the treatment groups can be evaluated with the Q-Q plot, box-whisker plot, etc. A large overlap increases the confidence that the estimated treatment effect will be generalized to the entire population represented by the sample. On the other hand, a low overlap implies that the treatment effect will only be represented by a small subgroup of the population. The lack of overlap in PSs can be an indication of large differences between treatment groups. In such a situation, unbiased treatment effects cannot be obtained by comparing treatment groups [13].

First Balance Control

Balance and bias have been shown to be related in a simulation study conducted by Belitser et al. (2011). In this study; standardized difference, Kolmogorov-Smirnov distance and Lévy distance showed high correlation with bias [22].

Balance in PS analysis is generally assessed by examining the differences in distributions of covariates between treatment groups [23]. Performing the first balance control in Phase 1 is important to determine the existence, and if exists the degree, of imbalance on covariates. The statistical methods frequently used to assess balance in PS analysis are summarized in Table 2. These should be applied separately for each covariate.

Phase 1 is completed with the first balance control. Next phase consists of two stages, and includes application of the PS methods and treatment effect estimation. These are respectively defined in the following sections.

Stage 2: Application of PS Methods

Stage 2 consists of specifying treatment effect, identifying the PS method to implement, and second balance control.

Specifying Treatment Effect

Treatment effect can be defined as the effect of the treatment arm on the dependent variable or response variable of interest. Average treatment effect on treated (ATT) is the average treatment effect for participants who actually received the treatment, while average treatment effect (ATE) is the average treatment effect for entire participants in the treatment and control groups. If the treatment groups are similar, ATE and ATT give nearly the same results. For this reason, in a RCT, the ATT and ATE are equivalent. Researchers should determine the type of treatment effect (ATE or ATT) in accordance with the purpose of the study. For instance, if countries' approach paths to an epidemic alert are examined, it will be more critical to evaluate the average effect on individuals suffering from pandemics rather than on all individuals [8,13,18].

PS Methods

A vital issue when using PS is to choose the type of PS method to implement. Four different methods are in use: the PS matching, stratification (or subclassification), weighting, and covariate adjustment. A brief summary of the first three methods which are suitable to be schematized is given in Figure 1.

Table 2. Summary of statistical methods used for assessing balance in PS analysis

Balance diagnostic	Short Definition and Interpretation
Variance ratio	It is the ratio of covariate variances between treatment groups. Close to 1 indicates good balance in covariate, while less than 0.5 or greater than 2 indicates extreme imbalance.
Standardized difference	It is the most frequently used balance criterion in the literature and is defined as the absolute treatment group difference in means/rates divided by the pooled standard deviation of the covariate. Although there is no universally accepted threshold to be used for standardized difference, a value of < 0.1 indicates a good balance on the covariates. For a continuous covariate: $d = \frac{(\bar{x}_t - \bar{x}_c)}{\sqrt{\frac{s_t^2 + s_c^2}{2}}}$ $\bar{x}_t$ and $\bar{x}_c$ denoted the sample mean of the covariate in treatment and control groups, respectively. s_t^2 and s_c^2 denoted the sample variance of the covariate in treatment and control groups, respectively. For a dichotomous covariate: $d = \frac{(\hat{p}_t - \hat{p}_c)}{\sqrt{\frac{\hat{p}_t(1-\hat{p}_t) + \hat{p}_c(1-\hat{p}_c)}{2}}}$ $\hat{p}_t$ and $\hat{p}_c$ denoted the prevalence of the covariate in treatment and control groups, respectively (Austin et al., 2007; Deb et al., 2016; Austin, 2008; Alam et al., 2019).
Overlapping Coefficient	It measures the amount of overlap of the covariate in the treatment groups. It takes a value between 0 and 1 (0: non-overlap, 1: perfect overlap).
Kolmogorov–Smirnov distance	It is the maximum vertical distance between two cumulative distribution functions of a covariate. It takes a value between 0 and 1 (0: perfect overlap, 1: non-overlap).
Lévy distance	It can be considered as a type of Kolmogorov-Smirnov distance that takes into account both the horizontal and vertical distance between the two cumulative distribution functions. It takes a value between 0 and 1 (0: perfect overlap, 1: non-overlap).

Notes: (1) The statistical methods are defined under the assumption that there are 2 treatment groups. (2) Detailed information on Overlapping Coefficient, Kolmogorov– Smirnov distance and Lévy distance is available in Belitser et al.(2011).

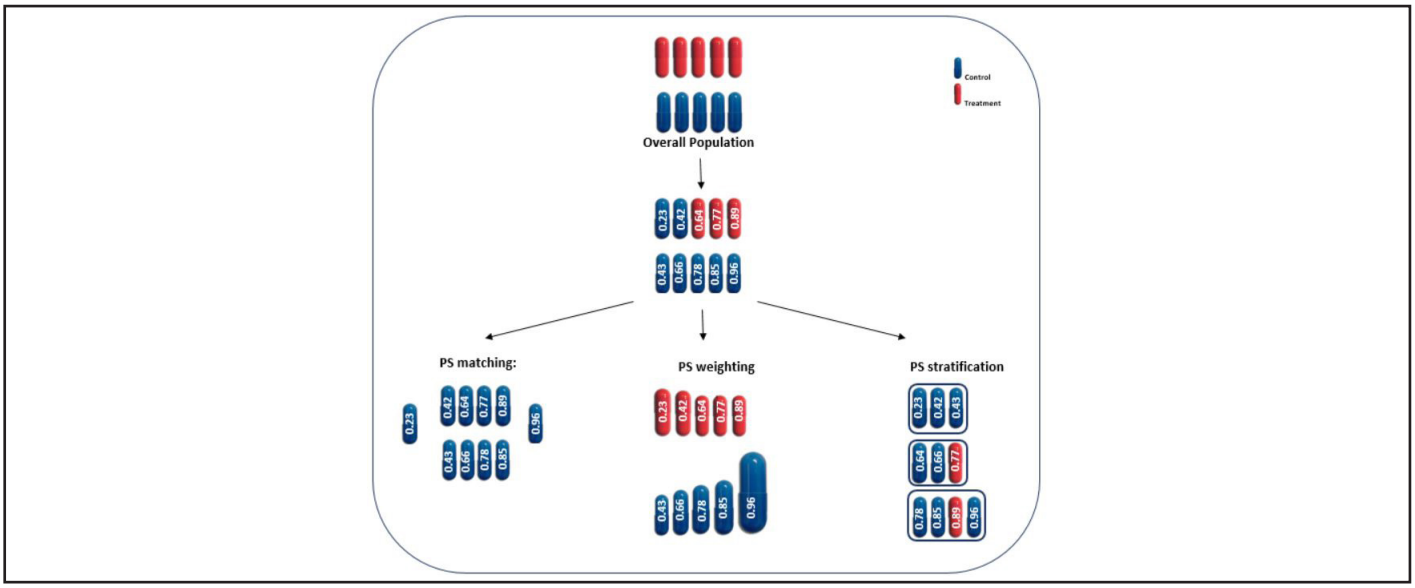

Figure 1. PS methods: matching, weighting and stratification

The overall population is a group of patients, consisting of the treatment group (red) and the control group (blue). Using the PS model, the estimated propensity scores for each patient are given in the figurations. Matching, weighting and stratification were used as PS methods. *PS matching*: In the treatment and control groups, patients with the closest PS values were matched and two patients who could not be matched were excluded. *PS weighting*: The patients were weighted according to their PS and the figurations were scaled according to these weights. *PS Stratification:* Three strata were created considering the closeness of the patients' PS.

PS Matching

PS matching, which makes it possible to obtain ATT estimates, is mainly used to create treatment groups more similar in terms of their distinctive characteristics. In this framework, participants with similar PSs in the treatment groups are matched with the preferred PS matching method. This means a participant in treatment group and participant in control group are matched when the smallest distance between their PSs is obtained. Several PS matching mechanisms are available in the literature. This study will emphasize the most frequently encountered four matching methods in medical literature; namely greedy matching, nearest neighbor matching, optimal matching and full matching will be discussed. In addition to these, there are mechanisms such as kernel matching, genetic matching, difference-in-differences matching etc. to be used as alternatives. For the ease of explanation, it will be assumed that there are 2 treatment groups (treatment and control).

Greedy Matching: The logic of this mechanism is based on matching a participant in the treatment group with the first participant who obtained the closest distance from participants randomly selected in the control group [24].

Nearest Neighbor Matching: This mechanism matches a participant in the treatment group with the participant who obtained the closest distance from many participants in the control group. Greedy and nearest neighbor matching are explained as similar methods under one title in some sources but some authors claim that this may cause confusion among researchers [13].

Optimal Matching: This mechanism is based on minimizing average absolute PS distance called global PS distance in whole matched pairs [24].

Full Matching: The matches are obtained in a similar manner to an optimal matching but the weights are used after matching is performed. In this context, each treated participant is given 1 as weight; and the control participant in each match takes the weight obtained by proportioning the number of participants receiving treatment in the match to the number of control participants in the same match [13].

A disadvantage related to PS matching is that among the discussed matching mechanisms only full matching guarantees all participants to be included in the matching process. It is clear that the exclusion of non-matched participants from the study will result in a decrease in statistical power and loss of both generalizability and precision of treatment effect estimates [15,25].

One point to be considered in the PS matching methods is whether the matching will be made "with replacement" or "without replacement". If a participant in the treatment group is matched with a participant in control group, and that control participant is used again (i.e. matching is done with replacement) a control participant can match with more than one participant in treatment group. Matching with replacement is particularly useful when there are few control participants that can be matched to treated participants. On the other hand, if a participant in the treatment group matches a participant in control group and that control participant is not used again (i.e. matching is done without replacement) precision will be increased but also bias will also increase [16,17,26]. Another point to be considered in PS matching methods is the ratio of treatment and control participants in the matching process. The most common ratio in the literature is 1:1, i.e. matching 1 treatment participant with 1 control participant. Other ratios can also be used (1:M matching) [17].

The restriction of "maximal acceptable difference" was proposed as a solution to the problem of matching individuals whose PSs are not close in treatment groups. Maximal acceptable difference is also called 'caliper' or 'tolerance' and is expressed by ε. Basically, using the caliper means that the closest match is determined by the $d(i,j) < \varepsilon$ inequality, where and j represent the individuals in the treatment and control groups respectively. The caliper proposed by Cochran and Rubin (1973) is $\varepsilon < 0.25\sigma_{ps}$, σ_{ps} being the standard deviation of PS [3]. Another approach is matching participants using calipers of width equal to 0.2 of the standard deviation of the logit of PS [17,24]. Austin (2011) states

that the usage of this stated approach eliminates about 99% of the bias and minimizes the mean square error of the treatment effect. In some studies, however, the restriction of maximal acceptable difference is considered as a separate matching method (namely caliper or radius matching) [6].

PS Stratification

In the PS stratification method (which is also known as the PS subclassification method), the PSs of the participants are ordered and using these PSs, mutually independent stratas of approximately equal sizes are created. Commonly, it is recommended to use 5 strata, which are formed by quintiles of ordered PSs, because this causes about 90% reduction in bias [4,9,14]. It has also been suggested that it is appropriate to use 10 or 20 strata if the sample size is large [16].

PS Weighting

In the PS weighting method, the treatment and control participants in the sample are weighted with the weights produced from their PSs. Both ATE and ATT estimates can be obtained in PS weighting, although with different mechanisms. In this context, ATT estimation is produced using 'weighting by the odds', and ATE estimation using inverse probability weighting.

Weighting by the Odds: In this mechanism, every participant in the treatment group receives a weight of 1, while participants in the control group receive a weight of their PS, converted to the odds scale ($e_i / (1-e_i)$). With weighting by the odds, the control participant whose PS is closer to the participant in the treatment group receives more weight [27].

***Weighting by Inverse Probability of Treatment Weights (IPTW*)**: The difference of this mechanism from weighting by the odds is that the participants in the treatment group also receive weights based on PS. In IPTW, each participant is weighted by the inverse probability of receiving the treatment. In this case, the participant in the treatment group is weighted with $1/e_i$ and the participant in the control group with $1/(1-e_i)$. A problem that may arise in IPTW is that a participant in the treatment group with a very low PS receives a very large weight, or similarly, a participant in the control group with a PS close to 1 receives a very large weight. It is not recommended to simply exclude these participants from the study, instead some approaches have been proposed to deal with extreme weights. One of the approaches, the stabilization procedure, uses the standardized weights $\hat{e}/e_i$ for the treatment group and $(1-e)/1-e_i$ for the control group where $\hat{e}=1 / \sum_{i=1}^{n} e_i$). Another approach is the trimming procedure that restricts all participant weights to a predetermined range. In the literature, the trimming is often applied to the extreme 1% or 5% of the weights [15,28]. Detailed information on trimming procedures is availabled in the study by Yoshida L. et al(2019) [29].

PS Covariate Adjustment

In this PS method, PSs produced in the Stage 1 are included in the treatment effect model as explanatory variables. In other words, PS is used as a control variable in estimating the treatment effect. An important assumption in this method is that the nature of the relationship between PS and the outcome is modeled properly [8,17,18].

Second Balance Control

Once the PS methods are implemented, the next stage is mainly evaluating whether the PS model has been adequately specified. The adequacy of the PS model, in other words, the success of the PS model can be evaluated by comparing the balance between treatment groups [8,18]. The statistical methods in Table 2 can also be used in this stage, with slight differences. For PS matching, the calculations are made considering the matching samples whereas for PS weighting, the weighted sample should be taken into consideration. For stratification, the calculations are made separately for each strata [8,18]. Besides the five statistical methods summarized in Table 2, statistical significance tests, Hosmer-Lemeshow goodness of fit test and c-statistic are also used for balance assessment [30]. The quality of a PS model must be evaluated on the basis of how well individual characteristics are balanced between the two treatment groups. For this very reason, a number of studies declare that goodness of fit tests such as the Hosmer-Lemeshow test or discriminant measures such as c-statistics are not suitable for balance assessment [5,17,19,31]. A detailed information on why statistical significance tests should not be recommended for balance assessment can be found in Austin (2011) [8,18].

It should be noted that, the difference between the treatment groups in terms of the covariate of interest (i.e. imbalance) at this stage may be due to the incorrect specification of the PS model or the use of an inappropriate PS method [17]

Graphical methods such as Q-Q plots, cumulative distribution functions, side-by-side boxplots and density plots can also be used in balance assessment for continuous covariates [8,18]. In

the mostly encountered Q-Q plot, the distribution of a covariate in the treatment group is plotted against the distribution in the comparison group and the covariate is considered balanced if a 45 degree straight line is obtained [32].

Stage 3: Treatment Effect Estimation

After the second balance control has been conducted, the next step is the treatment effect estimation. At this stage, general or generalized linear models can be used for any PS method. The type of outcome variable determines which modeling approach will be used. For instance, the linear regression model can be used in the case of a continuous outcome variable (e.g. hemoglobin), and the logistic regression model in the case of a binary outcome variable (e.g. mortality).

Different PS methods require different treatment effect estimation models. In the PS matching method, treatment effect model is fitted to the matched sample. In the PS stratification method, treatment effect model is fitted separately within each strata and the treatment effects obtained from the fitted models for each strata are pooled to calculate an overall treatment effect estimate [15,17]. In the PS weighting method the calculated weights have to be included in the treatment effect model as a weighting variable.

At the end of Stage 3, it is recommended that researchers perform a sensitivity analysis in which they explore to what extent the estimated treatment effects are robust to hidden bias [16].

Table 3. Quality Checklist of PS Analysis

	Item	✓ or X
Preparation for PS Analysis		
	1. Point out the scientific background	☐
	2. Indicate key components of study design	☐
	3. Clearly state the objective(s) of the study	☐
	4. Describe data sources and measurement methods for all variable of interest	☐
PS Model Building		
	5. Determine the appropriate set of variables to use in the PS model	☐
	6. Decide the PS estimation method (parametric and nonparametric) and explain why select to this method	☐
	7. Evaluate the overlap and state the method(s) used for checking overlap	☐
	8. Present the degree of first balance and state balance diagnostic(s) used	☐
Application of PS Methods		
	9. Specify the type of treatment effect	☐
	10. Explain which PS method is used a) If weighting or matching is used, state which strategy were chosen b) PS weighting method - state how to deal with extreme weights, if any c) PS matching method - state the ratio of treatment and control, indicate if there are any excluded participant(s), specify whether matching was done with or without replacement, report caliper if used d) PS stratification method - state how many strata are used	☐
	11. Present the degree of second balance and state balance diagnostic(s) used	☐
	12. State which model approach is used for treatment effect estimation	☐
	13. Perform the Sensitivity analysis	☐
	14. Report and interpretation of treatment effect	☐

PS: Propensity Score

Stages of PS analysis can be applied with many statistical software such as R, Stata, SAS. A detailed implementation of the PS stages in STATA and SAS are available in Lunt(2014) and Lanehart(2012), respectively [33,34]. Which program is preferred depends on which software the researcher is comfortable with.

RESULTS AND DISCUSSION

In recent decades, the use of PS analysis in clinical studies has increased dramatically. It can be said that the PS analysis performed do not fit a standard pattern in many of these studies. With the aim of standardizing the steps in PS analysis, a basic guideline has been presented. In addition, a checklist in parallel with the guideline that can be easily used by researchers to standardize PS analysis is included in the Table 3. From the planning stage of the PS analysis, this checklist offers researchers the opportunity to control their step-by-step implementation as well as making it possible for them to make an assessment on the quality of their PS analysis. We believe that the quality of future PS analysis will increase if such checklists are followed. Since the use of PS analysis methods are being more and more frequent, we propose that the integration of the crucial steps into STROBE (the Strengthening the Reporting of Observational Studies in Epidemiology) Statement may also be taken into consideration. To be concise, items in the "Participants, Bias and Statistical Methods" sections of STROBE can be updated to question, and explain the efforts to overcome selection bias by using PS analysis methods [35].

Author Contributions: Conception: TD, N; YK, SP - Design: TD, N; YK, SP - Supervision: TD, N; YK, SP; A,ZN - Fundings:- Materials: - Data Collection and/or Processing: - Analysis and/or Interpretation: TD, N; YK, SP; A,ZN - Literature: TD, N; YK, SP - Review: TD, N; YK, SP - Writing: TD, N; YK, SP - Critical Review: TD, N; YK, SP; A,ZN.

Conflict of Interest: The authors declare that they have no conflicts of interest.

Funding: None.

REFERENCES

[1] Stürmer T, Joshi M, Glynn RJ, Avorn J, Rothman KJ, Schneeweiss S. (2006) A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. Journal of clinical epidemiology. 9(5), 437-e1. https://doi.org/10.1016/j.jclinepi.2005.07.004

[2] Yao XI, Wang X, Speicher PJ. Hwang ES, Cheng P, Harpole DH, Pang HH (2017) Reporting and guidelines in propensity score analysis: a systematic review of cancer and cancer surgical studies. JNCI: Journal of the National Cancer Institute. 109(8), djw323. https://doi.org/10.1093/jnci/djw323

[3] Olmos A, Govindasamy P. (2015) Propensity scores: a practical introduction using R. Journal of MultiDisciplinary Evaluation. 11(25), 68-88. https://doi.org/10.56645/jmde.v11i25.431

[4] Pattanayak CW, Rubin DB, Zell ER (2011) Propensity score methods for creating covariate balance in observational studies. Revista Española de Cardiología (English Edition). 64(10), 897-903. https://doi.org/10.1016/j.rec.2011.06.008

[5] Heinze G, Jüni P (2011) An overview of the objectives of and the approaches to propensity score analyses. European heart journal. 32(14), 1704-1708. https://doi.org/10.1093/eurheartj/ehr031

[6] Luo Z, Gardiner JC, Bradley CJ (2010) Applying propensity score methods in medical research: pitfalls and prospects. Medical Care Research and Review. 67(5), 528-554 https://doi.org/10.1177/1077558710361486

[7] Austin PC, Grootendorst P, Anderson GM. (2007) A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study. Statistics in medicine. 26(4), 734-753. https://doi.org/10.1002/sim.2580

[8] Austin PC (2011) An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate behavioral research. 46(3), 399-424. https://doi.org/10.1080/00273171.2011.568786

[9] Rosenbaum PR, Rubin DB. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika. 70(1), 41-55. https://doi.org/10.1093/biomet/70.1.41

[10] Klompmaker S, van Hilst J, Wellner UF, Busch OR, Coratti A, D'Hondt M, Lips DJ (2020) Outcomes after minimally-invasive versus open pancreatoduodenectomy: a pan-European propensity score matched study. Annals of surgery. 271(2), 356-363. https://doi.org/10.1097/SLA.0000000000002850 Signori A, Pellegrini F, Bovis F, Carmisciano L, De Moor C, Sormani MP (2020) Comparison of Placebos and Propensity Score Adjustment in Multiple Sclerosis Nonrandomized Studies. JAMA neurology. https://doi.org/10.1001/jamaneurol.2020.0678.

[11] Lapointe-Shaw L, Bell CM, Austin PC, Abrahamyan L, Ivers NM, Li P, Dolovich L (2020) Community pharmacy medication review, death and re-admission after hospital discharge: a propensity score-matched cohort study. BMJ Quality & Safety. 29(1), 41-51. https://doi.org/10.1136/bmjqs-2019-009545

[12] Lee J, Little TD (2017) A practical guide to propensity score analysis for applied clinical research. Behaviour research and therapy. 98, 76-90. https://doi.org/10.1016/j.brat.2017.01.005

[13] Valojerdi AE, Janani L (2018) A brief guide to propensity score analysis. Medical journal of the Islamic Republic of Iran. 32, 122. https://doi.org/10.14196/mjiri.32.122

[14] Elze MC, Gregson J, Baber U, Williamson E, Sartori S, Mehran R, Pocock SJ (2017) Comparison of propensity score methods and covariate adjustment: evaluation in 4 cardiovascular studies. Journal of the American College of Cardiology. 69(3), 345-357. https://doi.org/10.1016/j.jacc.2016.10.060

[15] Stuart EA (2010) Matching methods for causal inference: A review and a look forward. Statistical science: a review journal of the Institute of Mathematical Statistics. 25(1), 1. https://doi.org/10.1214/09-STS313

[16] Deb S, Austin PC, Tu JV, Ko DT, Mazer CD, Kiss A, Fremes SE (2016) A review of propensity-score methods and their use in cardiovascular research. Canadian Journal of Cardiology. 32(2), 259-265. https://doi.org/10.1016/j.cjca.2015.05.015

[17] Austin PC (2011) A tutorial and case study in propensity score analysis: an application to estimating the effect of in-hospital smoking cessation counseling on mortality. Multivariate behavioral research. 46(1), 119-151. https://doi.org/10.1080/00273171.2011.540480

[18] Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Stürmer T (2006) Variable selection for propensity score models. American journal of epidemiology. 163(12), 1149-1156. https://doi.org/10.1093/aje/kwj149

[19] de Vries BBP, van Smeden M, Groenwold RH (2018) Propensity Score Estimation Using Classification and Regression Trees in the Presence of Missing Covariate Data. Epidemiologic Methods, 7(1). https://doi.org/10.1515/em-2017-0020

[20] Zhou J (2015) Comparison of approaches for handling missingness in covariates for propensity score models. PhD Thesis, The Pennsylvania State University. https://doi.org/10.1186/s12874-020-01053-4

[21] Belitser SV, Martens EP, Pestman WR, Groenwold RH, De Boer A, Klungel OH (2011) Measuring balance and model selection in propensity score methods. Pharmacoepidemiology and drug safety. 20(11), 1115-1129. https://doi.org/10.1002/pds.2188

[22] Weitzen S, Lapane KL, Toledano AY, Hume AL, Mor V (2004) Principles for modeling propensity scores in medical research: a systematic literature review. Pharmacoepidemiology and drug safety. 13(12), 841-853. https://doi.org/10.1002/pds.969.

[23] Austin, P. C (2011) Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. Pharmaceutical statistics. 10(2), 150-161. https://doi.org/10.1002/pst.433.

[24] Campbell MJ (2017) What is propensity score modelling? https://doi.org/10.1136/emermed-2016-206542

[25] Dehejia RH, Wahba S (2002) Propensity score-matching methods for nonexperimental causal studies. Review of Economics and statistics, 84(1), 151-161 https://doi.org/10.1162/003465302317331982

[26] Harder VS, Stuart EA, Anthony JC (2010) Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research. Psychological methods. 15(3), 234. https://doi.org/10.1037/a0019623

[27] Austin PC, Stuart EA (2015) Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. Statistics in medicine, 34(28), 3661-3679. https://doi.org/10.1002/sim.6607

[28] Yoshida K, Solomon DH, Haneuse S, Kim SC, Patorno E, Tedeschi SK, Glynn RJ (2019) Multinomial extension of propensity score trimming methods: a simulation study. American Journal of Epidemiology, 188(3), 609-616. https://doi.org/10.1093/aje/kwy263

[29] Ali MS, Groenwold RH, Belitser SV, Pestman WR, Hoes AW, Roes KC, Klungel OH (2015) Reporting of covariate selection and balance assessment in propensity score analysis is suboptimal: a systematic review. Journal of clinical epidemiology, ; 68(2), 122-131. https://doi.org/10.1016/j.jclinepi.2014.08.011

[30] Kuss O, Blettner M, Börgermann J (2016) Propensity Score: An Alternative Method of Analyzing Treatment Effects: Part 23 of a Series on Evaluation of Scientific Publications. Deutsches Ärzteblatt International, 113(35-36), 597. https://doi.org/10.3238/arztebl.2016.0597

[31] Alam S, Moodie EE, Stephens DA (2019) Should a propensity score model be super? The utility of ensemble procedures for causal adjustment. Statistics in medicine. 38(9), 1690-1702. https://doi.org/10.1002/sim.8075

[32] Lunt M (2014) Propensity analysis in Stata revision: 1.1. Documento disponible en: http://personalpages.manchester.ac.uk/staff/mark.lunt/propensity_guide.pdf

[33] Lanehart RE, de Gil PR, Kim ES, Bellara AP, Kromrey JD, Lee RS (2012) Propensity score analysis and assessment of propensity score approaches using SAS procedures. In Proceedings of the SAS Global Forum Conference (pp. 22-25). Cary, North Carolina: SAS Institute Inc.

[34] von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al (2007) The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. PLoS Med. 4(10):e296. https://doi.org/10.1371/journal.pmed.0040296

How to Cite;

Totik Doğan N, Yücel Karakaya SP, Alparslan ZN (2023) An Introduction to Propensity Score Analysis: Checklist for Clinical Researches. 29(3):667-676. Eur J Ther. https://doi.org/10.58600/eurjther1813