Multinomial Propensity Score for Ternary Exposure for Genetic Study

Cathy S. J. Fann
Academia Sinica
Thai Son Dinh
National Changhua University of Education
Yu-Hsien Chang
National Changhua University of Education
Jia Jyun Sie
National Changhua University of Education
Ie-Bin Lian (maiblian@cc.ncue.edu.tw)
National Changhua University of Education https://orcid.org/0000-0003-3992-6049

Research article

Keywords: propensity adjustment, confounding effect, SNP, contrast coding

Posted Date: August 10th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-50198/v1

Multinomial propensity score for ternary exposure for genetic study

Cathy S. J. Fann[1], Thai Son Dinh[2], Yu-Hsien Chang[3], Jia Jyun Sie[2], Ie-Bin Lian[3*]

1. Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan

2. Department of Mathematics, National Changhua University of Education, Changhua, Taiwan

3. Institute of Statistics and Information Science, National Changhua University of Education, Changhua, Taiwan.

*Correspondence:
Ie-bin Lian
maiblian@cc.ncue.edu.tw

ABSTRACT

Background: Propensity score (PS) is a popular method for reducing multiple confounding effects in observational studies. It is applicable mainly for situations wherein the exposure/treatment of interest is dichotomous and the PS can be estimated through logistic regression. However, multinomial exposures with 3 or more levels are not rare, e.g., when considering genetic variants, such as single nucleotide polymorphisms (SNPs), which have 3 levels (aa/aA/AA), as an exposure. Conventional PS is inapplicable for this situation unless the 3 levels are collapsed into 2 classes first.

Methods: A simulation study was conducted to compare the performance of the proposed multinomial propensity score (MPS) method under various contrast codings and approaches, including regression adjustment and matching.

Results: MPS methods had more reasonable type I error rate than the non-MPS methods, of which the latter could be as high as 30~50%. Compared with MPS-direct adjusted methods, MPS-matched cohort methods have better power but larger type I error rate. Performance of contrast codings depend on the selection of MPS models.

Conclusions: In general, two combinations had relatively better performance in our simulation of ternary exposure: MPS-matched cohort method with Helmert contrast and MPS-direct adjusted regression with treatment contrasts. Compared with the latter, the former had better power but larger type I error rate as a trade-off.

Keywords: propensity adjustment, confounding effect, SNP, contrast coding.

Background

Propensity score (PS) approach is a useful and popular method in epidemiology for mitigating confounding effects in observational studies [1]. It was first proposed by Rosenbaum and Rubin to infer cause and effect from observational studies [2]. A PS is estimated through a conditional probability wherein a particular treatment is assigned given a vector of observed covariates. Adjustment for the PS is sufficient to remove bias due to all observed covariates.

An observational study attempts to estimate the effects of a treatment or exposure (X) on the outcomes (Y) of subjects who are not assigned at random to the treatment or control group [3]. In such a study, the adjustment for confounders (Zs) that are associated with X and Y is crucial. Otherwise, inference can be severely biased. Common ways for adjustment include matching, stratification, and covariate-adjusted regression [4]. However, when potential confounders are numerous, the above approaches may be implausible. After stratifying on many covariates, some strata will have insufficient cases for analysis even for datasets with large samples. For covariate-adjusted regression, adjustment with excessive covariates or by applying a variable selection scheme can result in severe over-dispersion on the inference [5]. In such cases, PS is considered to be an efficient way to reduce the dimensions of

covariates for valid inference. It adjusts for the effects of multiple confounders to estimate treatment effects and balances covariates (such as age, gender, or population principal components) such that the treatment and control groups are comparable [1]. For a binary exposure *X*, the logistic regression model is often used to estimate the true PS by regressing *X* on observed covariates. Several ways to adjust for covariates when estimating the effects of treatment on outcomes by using PS include matching, stratification, direct adjustment in regression, and inverse probability treatment weighing [6]. Ali et al. reported that in studies using PS to control for confounding, matching on the PS is the most common approach (68.9%), followed by PS direct adjustment (20.9%); together, such studies comprise more than 90% of the applications involving PS [7]. Therefore, in this work, we focused on the categories of the MPS-direct adjusted and MPS-matched methods.

Most of the applications of PS have been in binary exposure. However, multinomial exposures with 3 or more levels are not rare, e.g., smoking can have 3 statuses, namely, “never”, “current”, and “ever-smoking”. In genetic studies, specific genetic variants, such as single nucleotide polymorphisms (SNPs), which have 3 levels (aa/aA/AA), can also be considered as an exposure. Conventional PS is inapplicable in this situation unless some levels are collapsed first.

Yoshida et. al. (2019) considered an extended propensity score for exposure X with J+1 levels [8]. They suggested using PS vector $\boldsymbol{e}' = (e_0 \dots e_J)$ with 1 probability of assignment for each level where $e_j = \mathrm{P}(\mathrm{X} = \mathrm{j}|\text{covariate } \mathrm{Z's})$ for $\mathrm{j} \in \{0,1, \dots, \mathrm{J}\}$ and used simulation to study the influence of the proposed ternary PS trimming methods for the inverse probability weighting approach. Using the same PS vectors, Wang et al. [9] proposed an application of stratification on the multiple propensity score to dose-response relationships in drug safety studies, and Wang et al, [10] generalized the pair-patching scheme to a trio-matching via defining a new distance among subjects from 3 treatment groups and discuss the choice of optimal caliper. Another alternative for multiple treatments is the generalized boosted model, a machine learning approach that uses inverse probability weighting regression to capture complex relationships between a treatment assignment and pretreatment covariates [11].

For the application of PS in genetic association studies, Jiang and Zhang (2011) suggested using nonparametric techniques to obtain PS while adjusting for covariates, such as population stratification or environmental factors for SNPs of interest, to identify disease associations [12]. Instead of directly testing epistatic effects from numerous combinations of SNPs, Sengupta Chattopadhyaya et al. (2016) proposed using PS as a dimension-reduction tool to improve the marginal single-point

association result for each SNP by accounting for the loss of heritability [13]. However, to be able to apply conventional PS, which assumes binary X for logistic regression, the 3-level SNP variable is first collapsed to 2 levels for dominant or recessive traits.

In this study, we compare 3 contrast codings for multinomial propensity scores (MPSs) to cope with genetic marker such as SNP, when it is considered as an exposure of interest. Simulation was used to assess the performances of the codings.

Methods

Let y be the outcome variable, x be the risk factor of interest, and $Z = (z_1 \ldots z_p)$ be numerous covariates. Among these covariates, some are confounders of x that are associated with x and y. The purpose of studies is to investigate the association between x and y. We consider the following generalized linear model system among x, y, and Z:

$$f(y) = \alpha x + Z\gamma + \varepsilon,$$

$$g(x) = Z\beta + \tau.$$

For binary y and binary x, the conventional procedure of PS is as follows:

(i) Fitting the logistic regression of x on the zs to obtain the prediction of x as PS:

$$\text{logit}\ (x) = \log(P(x = 1)/P(x = 0) = Z\beta.$$

(ii) Stratifying or matching *x* and *y* on the basis of PS, or, for direct adjustment, fitting the logistic regression of *y* on *x* and PS.

$$\text{logit}\ (y) = \log(P(y=1)/P(y=0) = \alpha x + Z\gamma.$$

In the case of binary *y* and ternary *x* with numerous *Z*s, we considered the following 3 contrast codings to convert *x* into dummy variables before estimating PS.

In ***treatment contrast,*** 1 group needs to be assigned as the baseline reference, and *x* is decomposed into dummy variables T_1 and T_2 as follows:

$$\begin{matrix} x \\ \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} \end{matrix} \rightarrow \begin{matrix} T_1 & T_2 \\ \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \end{matrix}.$$

In the following example, AA (x = 2) was used as the reference group, as follows:

$$x = \begin{pmatrix} aa \\ Aa \\ AA \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} \alpha_0 + \alpha_1 \\ \alpha_0 + \alpha_2 \\ \alpha_0 \end{pmatrix}.$$

Therefore, $\alpha_1 = (aa - AA)$ and $\alpha_2 = (Aa - AA)$. If A is the disease allele, α_1 is expected to be significant in all 3 (dominant, recessive, and additive) traits, whereas α_2 is expected to be significant in the additive and recessive traits.

In ***Helmert contrasts***,

$$\begin{matrix} \text{x} \\ \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} \end{matrix} \rightarrow \begin{matrix} H_1 & H_2 \\ \begin{bmatrix} -1 & -1 \\ 1 & -1 \\ 1 & 2 \end{bmatrix} \end{matrix}.$$

Therefore,

$$x = \begin{pmatrix} aa \\ Aa \\ AA \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 & -1 & -1 \\ 1 & 1 & -1 \\ 1 & 0 & 2 \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} \alpha_0 - \alpha_1 - \alpha_2 \\ \alpha_0 + \alpha_1 - \alpha_2 \\ \alpha_0 + \alpha_2 \end{pmatrix}.$$

$\alpha_1 = (Aa - aa)/2,\ \alpha_2 = [AA - (Aa + aa)/2]/2$. i.e., the significance of α_1, suggests the difference between (aa, Aa), and that of α_2 suggests the difference between AA and the average of Aa and aa. Therefore, if A is the disease allele, α_1 is expected to be significant in the dominant and additive traits, whereas α_2 is expected to be significant for the recessive and additive traits and has weak power for the dominant trait due to the effect size being partially offset by the cancel-out effect between AA and Aa.

Finally, we also consider a custom contrast, which we refer to as SNP contrasts, as shown below:

$$\begin{matrix} \text{x} \\ \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} \end{matrix} \rightarrow \begin{matrix} S_1 & S_2 \\ \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \end{matrix}.$$

Under the contrasts,

$$x = \begin{pmatrix} aa \\ Aa \\ AA \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} \alpha_0 \\ \alpha_0 + \alpha_2 \\ \alpha_0 + \alpha_1 + \alpha_2 \end{pmatrix}.$$

Consequently, $\alpha_1 = (AA - Aa)$ and $\alpha_2 = (Aa - aa)$. Given A as disease allele, a recessive trait is expected to have significant α_1 but nonsignificant α_2 and a dominant trait is expected to have significant α_2 but nonsignificant α_1, whereas an additive trait is expected to have significant α_1 and α_2.

If α_1 is significant but α_2 is not, this situation indicates that (AA) is different from (aa, Aa). This situation is referred to as the recessive trait. If $\alpha_2$is significant but $\alpha_1$is not, this indicates that (AA, Aa) is different from (aa). This situation is referred to as the dominant trait. On the other hand, if $\alpha_1$and α_2 are significant and have the same sign, this situation can be referred to as the additive trait.

For each contrast variable x_1 and x_2, we estimate the corresponding PSs via the stepwise logistic regression of x_i on *Z*s as $PS_i = P(x_i = 1 \mid \text{selected } Z\text{s})$, $i = 1, 2$, and then fit y on (x_1, x_2, PS_1, and PS_2) along with some *Z*s selected through the variable selection procedure. We refer to the above procedure as the MPS method. In direct-adjustment approach, y is regressed on x_1, x_2, PS_1, and PS_2 directly. In case-control matching, since only 2 groups (case and control) needs to be balanced, the conventional matching scheme can be applied. In cohort matching with 3 exposure

levels, instead of the trio-matching used by Wang et al. [10], we adopt a simpler double-pair-matching scheme which can be easily implemented in R package. Regarding the exposure SNP marker =0/1/2 as group 1/2/3 respectively, when treatment contrast is used, a subject from group 3 was matched with one from group 1 based on PS1, and one from group2 based on PS2 separately, that together form a matched trio. Analogously when Helmert contrast is used, after a matched pair between group 1 and 2 based on PS1 being formed, their average PS2 was calculated, and then a subject from group 3 with the closest PS2 to the average was selected to form a trio.

Simulation study was conducted using R package (version 4.0.1) to assess the performance among the combination of 8 models (2 non-PS models, 3 MPS-direct adjustment models and 3 MPS-matching models as shown in Table 1) and 3 types of contrast codings (treatment, SNP and Helmert) via simulation.

Comparing the simulation setups of Yoshida et al. (2019) for the log-linear model and Lian (2003) for the logit model [5,8] reveals that Yoshida's simulation is mainly used for assessing marginal estimands. However, in our case wherein SNPs were considered as exposures, we focused on assessing its conditional effect with other SNPs and phenotypes being adjusted as covariates. Therefore, in our simulation, the exposure–outcome functions were specified, and the later setup by Lian (2003) was adopted.

Simulation setup

The data of sample size *m* = 100, such that the *m* by 1 vectors binary outcome **y**; ternary exposure of interest **x**; and 20 covariates $z_1 - z_{20}$, were generated by using the following linear equation system:

$$\text{logit}\,(w_{ij}) = \beta \cdot (z_{i1} + \cdots + z_{i\ 10}), j = 1, 2,$$

$$x_i = f(w_{i1}, w_{i2}),$$

$$\text{logit}\,(y_i) = \alpha \cdot x_i + \gamma \cdot (z_{i6} + \cdots + z_{i\ 15}),$$

where $w_{ij} \sim \text{Bernoulli}(p_{ij})$, with

$$p_{ij} = \frac{e^{\beta \times \sum_{i=1}^{10} z_{ij}}}{1 + e^{\beta \times \sum_{i=1}^{10} z_{ij}}}, i = 1 \sim m, j - 1,2.$$

Function *f* specifies the type of the genetic trait of *x*: $x = w_1 + w_2$ for an additive trait, $x = w_1 * w_2$ for a recessive trait, and $x = w_1 + w_2 - w_1 w_2$ for a dominant trait. If a SNP takes a value as the number of *A*, i.e., SNP = 0 for (a, a), = 1 for (A, a) or (a, A), and = 2 for (A, A), then *x* = SNP for an additive trait. For a recessive trait, *x* = 1 if SNP = 2, and =0 otherwise. For a dominant trait, *x* = 1 if SNP > 0 and =0 otherwise.

According to the above linear system, $z_1 - z_5$ are associated *x* only; $z_6 - z_{10}$ are confounders that are associated with *x* and *y*; $z_{11} - z_{15}$ are associated with *y* only; and $z_{16} - z_{20}$ are unassociated with neither *x* nor *y*. The effect sizes re denoted by (α, β, γ) as

shown in Fig 1. The parameter setup used in this study is given by Table 1. The z_is were generated from N(0,1), whereas *w* and *y* were generated from Bernoulli distribution as shown in Fig 2. The following models are the methods compared in this work. The corresponding fitting procedures are listed in Table 2.

--- Insert Tables 1–2 and Figs 1–2 here ---

Note that Model 1 is an unadjusted regression model, and Model 2 is a commonly used covariate-adjusted regression model. Models 3-5 are MPS- and covariate-direct adjusted logistic regression (LR) models, with PS being forced-in and covariates being selected by using a stepwise procedure with levels 0, 0.05, and 0.15. In terms of results, Model 5 adjusts for the most covariates, whereas Model 3 adjusts for PS only. Models 6-8 are MPS-matched methods, wherein Model 6 is a matched case–control condition logistic regression (CLR), and Model 7 is a matched cohort CLR. For Model 8, the Cochran–Armitage Trend (CAT) test with 3 weighting schemes, namely, additive, recessive, and dominant, were used [14]. CAT, which uses the Chi-square statistics of 1 degree of freedom, is considered to be more powerful than the regular Peason's Chi-square of 2 degrees of freedom [15]. Here, Model 6 was used as a reference for comparison with Models 7 and 8 because matched case–control is inefficient or may actually introduce additional cofounders and bias and is not recommended [16-17].

Assessment of performance

A total of 12 parameter setups (α, β, γ) and 3 genetic traits (additive, recessive, and dominant) on exposure *X* were assumed in the simulation. For each of the 12 × 3 scenarios, n = 1000 replicates were generated. For the regression method (Models 1–7), a test on ternary exposure comprised 2 nominal T-statistics $\hat{\alpha}_1/\mathrm{SE}(\hat{\alpha}_1)$ and $\hat{\alpha}_2/\mathrm{SE}(\hat{\alpha}_2)$, where the estimate of $\hat{\alpha}_i$ and its nominal standard error $\mathrm{SE}(\hat{\alpha}_i)$ corresponded to each contrast. Here, exposure was considered as significant if at least 1 of the contrasts tested significant at the level of 0.05. In this case, the expected type I error that at least 1 of the 2 contrast variables is significant is $1 - (0.95)^2 = 9.75\%$. Tables 3–10 list the proportions of significance and the empirical standard deviations of nominal T-statistics (denoted as STD_T) for each of the 8 models and 3 contrast codings listed in Table 1 based on 1000 replicated samples corresponding to 1 of the parameter/trait scenario listed in Table 2. Notably, the proportion of significance is regarded as the type I error rate when the true α is 0 (no x–y association) and is regarded as power when $\alpha \neq 0$. As a nominal Wald's statistics on α, the T-statistic is supposed to follow an asymptotic standard normal distribution with variance 1; however, the variable selection procedure and the confounding effect can distort the assumption [5,18]. With STD_T > 1, the T-statistic has a heavier tail than the standard normal and therefore the

null hypothesis is likely to be rejected; this situation results in a high type I error rate. This scenario is known as over-dispersion. By contrast, STD_T < 1 is known as under-dispersion, which is likely to produce a conservative result.

Results

As mentioned in the previous session, a ternary exposure is decomposed into contrast variables T_1, T_2 for treatment contrast; S_1, S_2 for SNP contrast; and H_1, H_2 for Helmets contrast. Therefore, 2 estimated coefficients were obtained. Table 2 depicts the assessment of the performance of different contrasts and regression procedures under the null hypothesis (unassociated *x*–*y* with $\alpha = 0$) and strong confounding ($\beta = 0.5$, $\gamma = 0.5$). Here, type I error is referred to as the case when at least 1 contrast variable is significant at the level of 0.05. The results are shown in Tables 3–10.

--- Insert Table 3–10 here ---

Type I error comparison

Tables 3–4 list the type I error and std of T-statistics for Models 1–8 and 3 contrast codings across 12 simulated scenarios: 2 confounding (high and moderate), 2 *x*–*y* association (high and moderate), and 3 genetic traits (additive, dominant, and recessive). For each model, the contrast with the smallest type I error rate among the 3

genetic traits is bolded.

As expected, Model 1 had unacceptably large type I error rates that ranged from 30%–50% for the high-confounding scenario and from 15%–17% for the low-confounding scenario. After adjusting for the selected covariates, Model 2 reduced the type I error from Model 1. However, the type I error of Model 2 remained considerably larger than that of other MPS-adjusted/matched methods. The nominal standard deviation of the test statistics (STD_T) for Model 2 was also severely inflated and ranged from 1.25–1.5-fold in the high-confounding scenario and from 1.15–1.2 in the low-confounding scenario.

The averaged type I error rates among the MPS-adjusted models followed the order of

Model 6 < 0.0975 ≒ Model 3 ≒ Model 4 < Model 7 < Model 5 < Model 8 < Model 2 < Model 1,

where 0.0975 is the expected error rate. Among the MPS-direct adjusted models, Model 5 had the largest type I error possibly due to over-adjusting by selecting too many covariates. By neglecting Model 6, the MPS-matched method provided larger type I error than MPS-direct adjusted methods.

Among the 3 contrast codings, averaged SNP contrast had the smallest type I error, followed by Helmert contrast, and treatment contrast had the largest type I error. This situation was roughly true across all 3 genetic traits and 8 methods. The tendency shown in Table 4 for the low-confounding scenario was similar to that shown in Table 3, but with a smaller size of type I error.

Given that Models 1 and 2 had considerably larger type I error rates than other models, only MPS-Models 3–8 were compared for power.

Power comparison

Tables 5–8 list the power and std of T-statistics for Models 3–8 and 3 contrast codings across 12 simulated scenarios: 2 confounding (high and moderate), 2 *x*–*y* associations (high and moderate), and 3 genetic traits. The contrast with the largest power among the 3 genetic traits is bolded. Model 6 (MPS-matched case–control) had the worst power, which coincided with the conclusion that matched case–control is inefficient or may actually introduce additional cofounding and bias [16-17] and is therefore not recommended.

As expected, the powers of all models increased as the *x*–*y* association increased. In general, the MPS-matched cohort (Models 7 and 8) had better power than the

MPS-direct adjusted cohort (Models 3–5). In general, the power followed the order of

Model 8 > Model 7 > Model 3≒Model 4≒Model 5 > Model 6.

Among the MPS-direct adjusted models, treatment contrast had better power than other contrast codings for additive and recessive genetic traits. However, when the trait was dominant, Helmert contrast had better power than the other codings. On the other hand, among the MPS-matched cohort models, CAT (Model 8) outperformed CLR (Model 7) for dominant traits, but both models were equally good for additive and recessive traits. Helmert contrast had the best power among the 3 contrasts across all 3 traits.

The power of Model 8 with Helmert contrast for additive traits, for example, was as high as 93.9% under the scenario of high *x*–*y* association and high confounding and as low as 54.1% under the scenario of low *x*–*y* association and low confounding. The power of Model 3 with treatment contrast for an additive trait was as high as 79.2% in the scenario of high *x*–*y* association and low confounding and as low as 36.4% in the scenario of low *x*–*y* association and high confounding.

When confounding is absent

As a reference, Tables 9 and 10 show the results obtained with MPS when

confounding variables were absent. As shown in Table 9, some *z*s were associated with *y*, but none were associated with *x*. As illustrated in Table 10, some *z*s were associated with *x*, but none were associated with *y*. In these cases, the MPS-matched cohort had lower power than the MPS-direct adjusted regression because matching might be unnecessary to reduce the sample size due to the drop in unmatched observations.

Discussion

Our results showed that in the presence of high confounding effect from multiple confounders, MPS methods had much more reasonable type I error rate than the non-MPS methods. The conventional covariate-adjusted method could have high type I error (~30%) due to the sever over-dispersion of the nominal T-statistics. This was caused by the covariate selection procedure, which normally results the under-estimating the standard deviation for the target coefficient [5]. Compared with MPS-direct adjusted methods, MPS-matched methods have better power but larger type I error rate. Among the MPS-matched methods, MPS-matched cohort CAT has slightly better power than MPS-matched cohort CLR, again, with larger type I error as a trade off. When MPS-direct adjusted regression is used, the selection of excessive other covariates into the model is not recommended. MPS-matched cohort may have lower power than MPS-direct adjusted regression because matching may be

unnecessary to reduce the sample size due to the reduction in unmatched observations.

As for the contrast codings, when MPS-direct adjusted models are used, treatment contrast has better power for additive and recessive genetic traits, whereas Helmert contrast has better power for dominant traits. When the MPS-matched cohort method is used, Helmert Contrast had better power for all 3 genetic traits.

Conclusions

To summarize, for ternary exposure, when a strong confounding effect due to observed covariates is believed to exist, then the MPS-matched cohort method with Helmert contrast is recommended. Otherwise, MPS-direct adjusted regression with treatment and Helmert contrasts are worth trying. This work limited on discussing MPS-direct adjustment and MPS-matching separately. Nguyen et al. (2017) had proposed an approach of combining double-adjustment and matching in propensity score analysis [19]. Our results may suggest a future study on such combination in MPS.

Declarations

Ethics approval and consent to participate

This study involved only computer simulation data, no IRB approval or consent to participate is needed.

Consent for publication

The manuscript contains no individual person's data in any form.

Availability of data and materials

R code for simulation in this study is available upon request.

Competing interests

All the authors declare no competing interest and author.

TSD: no competing interest

YHC: no competing interest

JJS: no competing interest

CSJF: no competing interest

IBL: no competing interest

Funding

The research is self-funded, except that publication fee will be supported by the affiliated institute of corresponding as an encouragement of research.

Authors' contributions

TSD: interpreted the simulation results, literature review

YHC: design and conducting simulation on R, literature review

JJS: design and conducting simulation on R, literature review

CSJF: providing motivation of study, substantively revising the manuscript

IBL: major contributor in writing the manuscript.

All authors read and approved the final manuscript, and had agreed both to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work.

Acknowledgements

We would like to thank Ms. Wan-Tzu Chang and Mr. Ren-hao Liao in Data Science Research Center in NCUE for data management and programming initiation.

Reference

1. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behav Res. 2011;46(3):399-424. doi:10.1080/00273171.2011.568786

2. Rosenbaum PR. and Rubin DB. The Central Role of the Propensity Score in Observational Studies for Causal Effects," Biometrika, 1983;70, 41-55.

3. Jepsen P, Johnsen SP, Gillman MW, Sørensen HT. Interpretation of observational studies. Heart. 2004;90(8):956-960. doi:10.1136/hrt.2003.017269

4. Rosenbaum PR. Observational Studies. New York: Springer-Verlag; 2002.

5. Lian IB., Reducing Over-dispersion by Generalized Degree of Freedom and Propensity Score, Comput. Statist. Data Anal. 2003; 43(2): 197-214.

6. Pirracchio R, Petersen ML, van der Laan M. Improving propensity score estimators' robustness to model misspecification using super learner. Am J Epidemiol. 2015;181(2):108-119. doi:10.1093/aje/kwu253

7. Ali MS, Groenwold RH, Belitser SV, et al. Reporting of covariate selection and balance assessment in propensity score analysis is suboptimal: a systematic review. J Clin Epidemiol. 2015;68(2):112-121. doi:10.1016/j.jclinepi.2014.08.011z

8. Yoshida K, Solomon DH, Haneuse S, et al. Multinomial Extension of Propensity Score Trimming Methods: A Simulation Study. Am J Epidemiol. 2019;188(3):609-616. doi:10.1093/aje/kwy263

9. Wang JX, Donnan PT, Steinke D, MacDonald TM (2001) The multiple propensity score for analysis of dose-response relationships in drug safety studies. Pharmacoepidemiol Drug Saf 10:105–11.

10. Wang Y, Cai H, Li C, Jiang Z, Wang L, et al. (2013) Optimal caliper width for propensity Score Matching of Three Treatment Groups: A Monte Carlo study. PLoS ONE 8(12): e81045. doi:10.1371/journal.pone.0081045

11. McCaffrey, D. F., Ridgeway, G., & Morral, A. R. (2004). Propensity Score Estimation With Boosted Regression for Evaluating Causal Effects in Observational Studies. Psychological Methods. 9(4);403-425. https://doi.org/10.1037/1082-989X.9.4.403

12. Jiang Y, Zhang H. Propensity score-based nonparametric test revealing genetic variants underlying bipolar disorder. Genet Epidemiol. 2011;35(2):125 - 132. doi:10.1002/gepi.20558

13. Sengupta Chattopadhyay A, Lin YC, Hsieh AR, Chang CC, Lian IeB, Fann CS. Using propensity score adjustment method in genetic association studies. Comput Biol Chem. 2016;62:1-11. doi:10.1016/j.compbiolchem.2016.02.017

14. Dizier MH, Demenais F, Mathieu F. Gain of power of the general regression model compared to Cochran-Armitage Trend tests: simulation study and application to bipolar disorder. BMC Genet. 2017;18(1):24. Published 2017 Mar 10. doi:10.1186/s12863-017-0486-6

15. Zhou Z, Ku HC, Huang Z, Xing G, Xing C. Differentiating the Cochran-Armitage Trend Test and Pearson's χ2 Test: Location and Dispersion. Ann Hum Genet. 2017;81(5):184-189. doi:10.1111/ahg.12202

16. Pearce N, Analysis of matched case-control studies. BMJ. 2016 Feb 25;352:i969. doi: 10.1136/bmj.i969. PMID: 26916049

17. De Graaf MA, Jager KJ, Zoccali C, Dekker FW. Matching, an appealing method to avoid confounding?. *Nephron Clin Pract*. 2011;118(4):c315-c318. doi:10.1159/000323136

18. Lian, I. (1995), Asymptotics of Forced-In Variables From Screening Processes", Commun. in Statistics - Theory and Method, 24(1) 131-151.

19. Nguyen TL, Collins GS, Spence J, et al. Double-adjustment in propensity score matching analysis: choosing a threshold for considering residual imbalance. BMC Med Res Methodol. 2017;17(1):78. Published 2017 Apr 28. doi:10.1186/s12874-017-0338-0

Figures

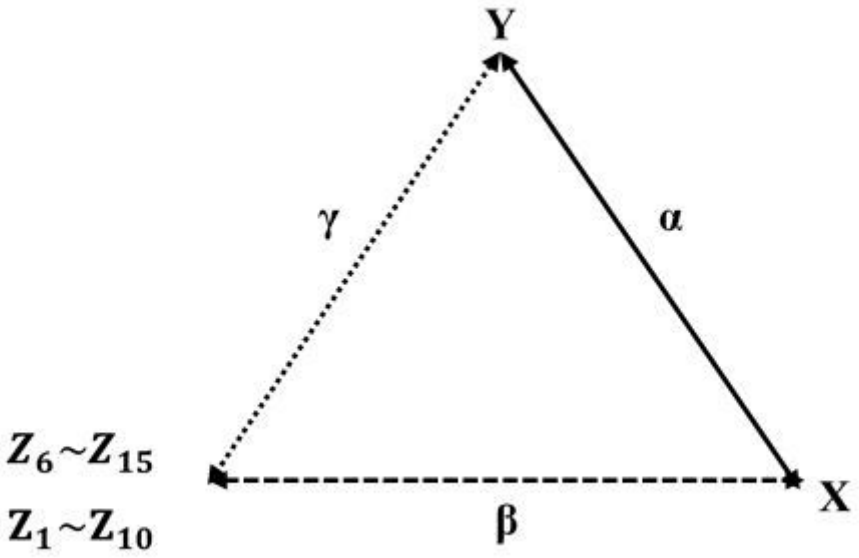

Figure 1

z1– z5 are associated with x only; z6–z10 are confounders that are associated with x and y; z11–z15 are associated with y only; and z16– z20 are not associated with either x or y. Effect sizes are denoted by (α, β, γ)

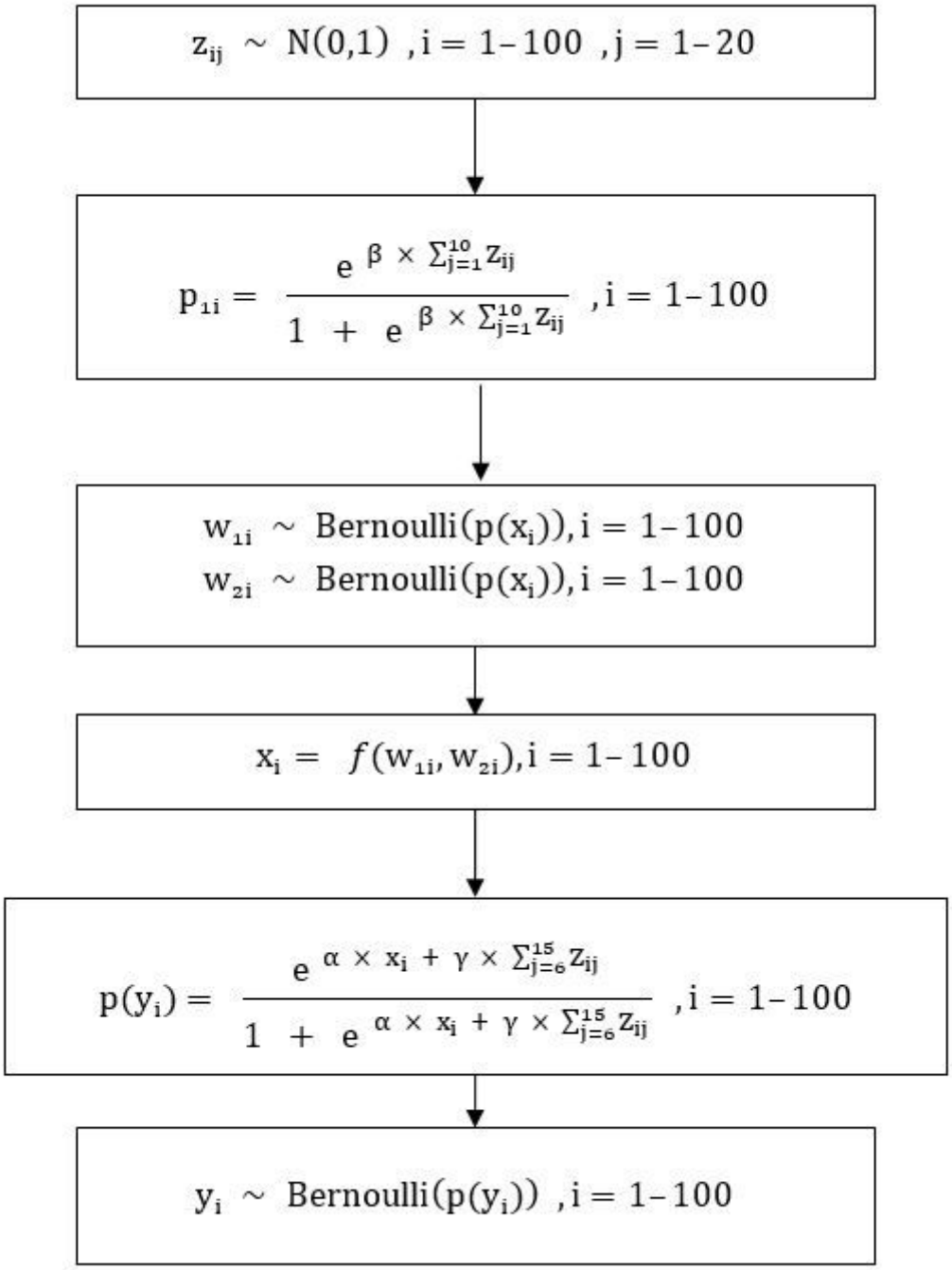

Figure 2

Flow chart of generalizing x, y, and zs.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TABLE2000727.pdf