Contrast-specific propensity scores

Shasha Han a, b and Donald. B. Rubin c, d

a Beijing International Center for Mathematical Research, Peking University, Beijing, People’s Republic of China; b Department of Health Care Policy, Harvard Medical School, Boston, MA, USA; c Yau Mathematical Center, Tsinghua University, Beijing, People’s Republic of China; d Department of Statistical Science, Fox Business School, Temple University, Philadelphia, PA, USA

ABSTRACT
Basic propensity score methodology is designed to balance the distributions of multivariate pre-treatment covariates when comparing one active treatment with one control treatment. However, practical settings often involve comparing more than two treatments, where more complicated contrasts than the basic treatment-control one, \((1, -1)\), are relevant. Here, we propose the use of contrast-specific propensity scores (CSPS), which allows the creation of treatment groups of units that are balanced with respect to bifurcations of the specified contrasts and the multivariate space spanned by these bifurcations.

1. Introduction
Studies with multiple treatments, whether due to multi-valued treatments (e.g. doses of a drug) or many factors (e.g. several types of drugs), are common. A contrast among \(T\) treatments is formally a vector of \(T\) coefficients that sum to zero. With two treatments, there is only one linearly independent contrast, conventionally written as \((1, -1)\), which compares the two conditions, but with more than two treatments, there are multiple contrasts. Contrasts have long been used in experimental studies (see [1–4]). They are typically more useful when they correspond to meaningful scientific comparisons.

When \(T = 2\), Rosenbaum and Rubin [5] proposed the use of the propensity score to balance multivariate covariates, and it is now widely used in applied statistics (e.g. [6–15]). The propensity score is the conditional probability of assignment to treatment versus control, given pre-treatment covariates. Some extensions of the propensity score to studies with multiple treatments have been suggested. The ‘generalized propensity score’ proposed by Imbens [16] is primarily applicable to methods using inverse probability weighting estimation [17]. The multidimensional propensity score proposed by Lechner [18] and the propensity function proposed by Imai and van Dyk [19] are two suggestions that differ from what we propose, as explained in more detail later.
A contrast-specific propensity score is the conditional probability of assignment to a bifurcation of treatment groups, for example, the treatment groups having positive coefficients in that contrast versus those having negative coefficients in that contrast, given pre-treatment covariates. We propose the use of contrast-specific propensity scores (csp) to create treatment groups with balanced covariate distributions in the multidimensional space spanned by such bifurcations of contrasts.

2. Contrasts among treatments

A $T$ component vector $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_T)$ is a contrast if $\sum_{t=1}^{T} \lambda_t = 0$. A contrast-specific propensity score uses a bifurcation of treatment conditions based on a contrast. For an example, let $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_T)$, with $\lambda_t \neq 0$. Then consider using the sign of contrast $\boldsymbol{\lambda}$, which is $\text{sgn}(\boldsymbol{\lambda}) = (\text{sgn}(\lambda_1), \ldots, \text{sgn}(\lambda_T))$, with $\text{sgn}(\lambda_t) = \lambda_t/|\lambda_t|$ so that $\lambda^+ = (\lambda_1^+, \ldots, \lambda_T^+)$ and $\lambda^- = (\lambda_1^-, \ldots, \lambda_T^-)$ denote the positive and negative components of $\boldsymbol{\lambda}$ respectively, i.e. $\lambda_t^+ = \max(\lambda_t, 0)$ and $\lambda_t^- = \min(\lambda_t, 0)$. The sgn function bifurcates contrast $\boldsymbol{\lambda}$ into $\text{sgn}(\lambda^+)$ versus $\text{sgn}(\lambda^-)$. For example, it bifurcates contrast $(1/2, 1/2, -1)$ into $(1, 0, 0)$ versus $(0, 0, -1)$. The sgn function bifurcates a contrast into components using zero as the boundary for all $T$ components. Non-zero boundaries can also be appropriate.

More generally, let $f(\boldsymbol{\lambda}) = (f(\lambda_1), \ldots, f(\lambda_T))$ be the function for bifurcation, with $f(\lambda_t) = \lambda_t/|\lambda_t|$ if $\lambda_t \leq \ell_t$ or $\lambda_t > u_t$, where $\ell_t$ and $u_t$ are the lower and upper boundaries for component $t$. For example, for the linear contrast with four groups, $(-3, -1, +1, +3)$, the bifurcation function $f$ with the lower boundary $(-1, -1, -1, -1)$ and upper boundary $(1, 1, 1, 1)$ bifurcates the contrast into $(-1, 0, 0, 0)$ versus $(0, 0, 0, 1)$.

Example 2.1 (One active treatment with two control conditions): For one active treatment, the first, with two control conditions, the contrast $(1, -1/2, -1/2)$ compares the active treatment to the average of the two control conditions, and the contrast $(0, 1, -1)$ compares the two control conditions. Contrasts $(1, -1/2, -1/2)$ and $(0, 1, -1)$ are orthogonal. For instance, LaLonde [20] was interested in the contrast of one experimental treatment, a job training programme, versus the average of two controls, using the groups from the Panel Study of Income Dynamics and from the Current Population Survey.

Example 2.2 (Multiple factors each with two levels): Table 1 displays the case of three factors (A,B,C), each with two levels denoted by 0 and 1, and thus $2^3 = 8$ treatments. The contrasts $\lambda_j, j = 1, 2, 3$, define the three main effects of A, B and C, the contrasts $\lambda_j, j = 4, 5, 6$, define the three two-way interaction effects, commonly labelled AB, BC and AC, and the contrast $\lambda_7$ defines the three-way interaction effect, ABC (see, e.g. [3]). The contrast $\lambda_8$ compares the combination of factors A and B versus the main effect of factor A; the contrast $\lambda_9$ compares the combination of factors A and B versus the main effect of factor B; the contrast $\lambda_{10}$ compares the effect of factors A and B both at level ‘1’ versus when they are both at level ‘0’. For example, in a recent study, Kaplan et al. [21] considered the contrasts $(-1, 1, 0, 0)$ and $(0, 0, -1, 1)$.

3. Notation

Consider a study with $N$ units, indexed by $i \in \{1, \ldots, N\}$. The outcome variable $Y$ is measured on each unit after its treatment exposure. Associated with treatment $t$ is the potential
outcome $Y_i(t)$, the value of $Y$ when the unit $i$ is exposed to treatment $t$, which implicitly assumes the stable unit treatment value assumption (SUTVA) [22], within the ‘Rubin Causal Model’ [23] – often also called the ‘potential outcomes approach to causal inference’ [24]. Each unit $i$ is associated with covariates $X_i$, $X_i = (X_{i1}, \ldots, X_{iK}) \in \mathbb{R}^K$, that are measured prior to treatment exposure, and which ideally are balanced across treatment groups, meaning that they have the same distributions under all treatments, which is ensured by randomization. The $\lambda$ contrast of potential outcomes for unit $i$ is $\sum_{t=1}^{T} \lambda_t Y_i(t)$.

Let $w_{it}$ be the indicator for whether unit $i$ is assigned to treatment $t$. Specifically, $w_{it} = 1$ if $W_i = t$ and 0 otherwise, where $W_i = t$ indicates that unit $i$ receives treatment $t$, $t \in \{1, \ldots, T\}$. To illustrate, let $D_i$ be the indicator for whether $w_{it}$ corresponds to a positive or negative $\lambda_t$, i.e. $D_i = 1$ if $\text{sgn}(\lambda_t) = 1$ and $D_i = 0$ if $\text{sgn}(\lambda_t) = -1$. Based on $D_i$, csps are the conditional probability of assignment to the treatment groups with positive coefficients versus negative coefficients of contrast $\lambda$, given pre-treatment covariates $X_i$,

$$c(X_i) = \text{pr}(D_i = 1 \mid X_i, D_{it} = 0, 1).$$

More generally, Equation (1) defines csps where $D_i$ indicates the bifurcation function for the contrast.

Let the conditional probability of assignment to treatment $t$, given covariates $X_i$, be $p_t(X_i) = \text{pr}(W_i = t \mid X_i)$. In Example 2.1, csps for the contrast $(1, -1/2, -1/2)$ is $p_1(X_i)$; In Example 2.2, csps for the contrast $\lambda_1$ is $\sum_{t=5}^{8} p_t(X_i)$, which equals $\text{pr}(A = 1 \mid X_i)$.

4. Basic results

The assignment with respect to the sgn bifurcation of contrast $\lambda$ is unconfounded, given covariates $X_i$, if $D_i \perp \perp Y_i(1), \ldots, Y_i(T) \mid X_i$. The condition is weaker than the strong unconfoundedness condition $W_i \perp \perp Y_i(1), \ldots, Y_i(T) \mid X_i$, originally defined in Rubin [25] and as stated in Imbens and Rubin [24].

As with the basic propensity score, the key advantage of csps is their balancing property. For one contrast, the standard propensity [5] simply follows,

**Property 4.1:** Creating balance on $c(X_i)$, the csps of contrast $\lambda$, balances the bifurcation of that contrast, i.e. $D_i \perp \perp X_i \mid (c(X_i), Di = 0, 1)$. 

---

### Table 1. Three factorial treatments, each with two levels.

| Treatments Indexings | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|----------------------|---|---|---|---|---|---|---|---|
| Factors A            | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| Factors B            | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 |
| Factors C            | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 |

| Contrasts | $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $\lambda_4$ | $\lambda_5$ | $\lambda_6$ | $\lambda_7$ | $\lambda_8$ | $\lambda_9$ | $\lambda_{10}$ |
|-----------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
|           | $-1$        | $-1$        | $-1$        | $-1$        | $1$         | $1$         | $1$         | $-1$        | $0$         | $-1$        |
|           | $-1$        | $1$         | $-1$        | $1$         | $1$         | $1$         | $1$         | $1$         | $-1$        | $1$         |
|           | $-1$        | $-1$        | $1$         | $1$         | $-1$        | $1$         | $1$         | $1$         | $1$         | $1$         |
|           | $1$         | $1$         | $1$         | $-1$        | $1$         | $-1$        | $1$         | $1$         | $1$         | $1$         |
|           | $-1$        | $1$         | $1$         | $-1$        | $1$         | $-1$        | $1$         | $1$         | $1$         | $1$         |
|           | $0$         | $0$         | $0$         | $0$         | $-1$        | $1$         | $1$         | $1$         | $1$         | $-1$        |
|           | $0$         | $-1$        | $0$         | $-1$        | $0$         | $1$         | $0$         | $1$         | $0$         | $1$         |
|           | $-1$        | $0$         | $-1$        | $0$         | $0$         | $1$         | $0$         | $1$         | $0$         | $1$         |
For $J$ contrasts, let $c_j(X_i)$ be the csps of the contrast $\lambda_j$ and let $D_{ij}$ be the bifurcation indicators with respect to contrast $\lambda_j$, $j = 1, \ldots, J$.

**Property 4.2:** Balance on $c_j(X_i)$, or any one-to-one function of $c_j(X_i)$, $j = 1, \ldots, J$, balances the associated bifurcations of contrasts $\lambda_j$, $j = 1, \ldots, J$. That is, $(D_{ij}; j = 1, \ldots, J) \perp \perp X_i \mid (c_j(X_i), D_{ij} = 0, 1; j = 1, \ldots, J)$, and therefore balances the subspace spanned by these bifurcations of contrasts $\lambda_1, \ldots, \lambda_J$.

Because $T-1$ linearly independent vectors span the full space, we have that balance on csps of contrasts with $T-1$ linearly independent vectors, e.g. sgn bifurcations of orthogonal contrasts, balances all linear contrasts among $T$ treatments, and all bifurcations among $T$ treatments.

Property 4.2 has an interesting implication in practice. The $J$ csps, once created, could be treated as covariates using a usual propensity score analysis, which creates a chain of balance. That is, balance on the propensity scores with the $J$ csps used as covariates balances the $J$ csps, and hence by Property 4.2, balances the subspace spanned by these bifurcations of the $J$ contrasts, and therefore, any linear combination of these bifurcated contrasts.

### 5. Illustrations

We start with a simple example with 24 units with treatment indicator $W_i$, $W_i = 1, 2, 3$. Suppose that one contrast is under investigation, e.g. $\lambda_1 = (1/2, 1/2, -1)$. Using the sgn bifurcations of the contrast, we bifurcate the multivariate indicators of treatments into bivariate indicators, a collection of $D_{i1} = 0$ and $D_{i1} = 1$ (See Table 2). The standard propensity score in Rosenbaum and Rubin [5] then directly applies.

For creating balance on csps of $J$ contrasts as well as for assessing balance for the contrasts and their linear combinations, the algorithm in Supplementary materials provides an illustrative routine. We refer to this routine as ‘the Algorithm.’

#### 5.1. An artificial example

Table 2 displays an artificial dataset with 24 units, three treatments and four covariates. We implement the Algorithm using two contrasts $\lambda_1 = (1/2, 1/2, -1)$ and $\lambda_2 = (1, -1, 0)$. We show the balance results for a third contrast $\lambda_3 = \lambda_1 - 1/2\lambda_2 = (0, 1, -1)$, which is a linear combination of $\lambda_1$ and $\lambda_2$. We briefly discuss the resulting balance for the four covariates and an alternative approach where only the bifurcation of $\lambda_2$ is used for balancing.

The third column of Table 2 shows the treatment group indicators with respect to the two contrasts $\lambda_1$ and $\lambda_2$. After the contrasts $c_1$ and $c_2$ are created, they are treated as covariates and a usual propensity score analysis is used to estimate the probability of $D_{i3} = 1$ versus $D_{i3} = 0$ given $(c_1(X_i), c_2(X_i))$. The right panel shows the estimated propensity scores and the subclass labels created by this algorithm. Clearly, within each subclass, the difference in means between treatment groups with respect to contrast $\lambda_3$ is exactly balanced because balancing on $(c_1, c_2)$ balances the subspace spanned by these bifurcations of the two contrasts $\lambda_1$ and $\lambda_2$.

However, balancing on the csps of bifurcations does not balance the subspace that is orthogonal to these bifurcations. To illustrate this, we balance on $c_2$ and evaluate
Table 2. A simple artificial example showing that balance on $c_1, c_2$ balances $\lambda_3$.

| $(X_{1i}, X_{2i}, X_{3i}, X_{4i})$ | $W_i$ | $D_{1i}$ | $D_{2i}$ | $c_1$ | $c_2$ | $D_{3i}$ | Propensity score | Subclass labels | Covariate in bifurcated groups |
|----------------------------------|-------|---------|---------|------|------|--------|---------------|----------------|---------------------|
| $(1,1,1,1)$                      | 1     | 1       | 1       | –    | –    | –      | $\frac{1}{5}$ | 1              | $(1,1,1,1)$        | $D_3 = 1$ |
|                                 | 2     | 1       | 0       | 1    | –    | –      | –             | –              | $(1,1,1,1)$        | $D_3 = 0$ |
|                                 | 3     | 0       | –       | $\frac{1}{3}$ | $\frac{1}{2}$ | 0 | $\frac{1}{5}$ | 1              | $(1,1,1,1)$        | $D_3 = 1$ |
|                                 | 3     | 0       | –       | 0    | –    | –      | –             | –              | $(1,1,1,1)$        | $D_3 = 0$ |
| $(1,0,1,0)$                      | 1     | 1       | 1       | –    | –    | –      | –             | –              | $(1,0,1,0)$        | $D_3 = 1$ |
|                                 | 1     | 1       | 1       | –    | –    | –      | –             | –              | $(1,0,1,0)$        | $D_3 = 0$ |
|                                 | 2     | 1       | 0       | $\frac{2}{3}$ | $\frac{2}{4}$ | $\frac{1}{4}$ | $\frac{1}{5}$ | 2              | $(1,0,1,0)$        | $D_3 = 1$ |
|                                 | 3     | 0       | –       | 0    | –    | –      | –             | –              | $(1,0,1,0)$        | $D_3 = 0$ |
| $(0,1,1,0)$                      | 1     | 1       | 1       | –    | –    | –      | –             | –              | $(0,1,1,0)$        | $D_3 = 1$ |
|                                 | 1     | 1       | 1       | –    | –    | –      | –             | –              | $(0,1,1,0)$        | $D_3 = 0$ |
|                                 | 2     | 1       | 0       | $\frac{5}{6}$ | $\frac{2}{5}$ | 1 | $\frac{3}{4}$ | 3              | $(0,1,1,0)$        | $D_3 = 1$ |
|                                 | 2     | 1       | 0       | 1    | –    | –      | –             | –              | $(0,1,1,0)$        | $D_3 = 0$ |
|                                 | 3     | 0       | –       | 0    | –    | –      | –             | –              | $(0,1,1,0)$        | $D_3 = 0$ |
| $(0,0,0,0)$                      | 1     | 1       | 1       | –    | –    | –      | –             | –              | $(0,0,0,0)$        | $D_3 = 1$ |
|                                 | 1     | 1       | 1       | –    | –    | –      | –             | –              | $(0,0,0,0)$        | $D_3 = 0$ |
|                                 | 2     | 1       | 0       | $\frac{2}{3}$ | $\frac{1}{2}$ | 1 | $\frac{1}{2}$ | 4              | $(0,0,0,0)$        | $D_3 = 1$ |
|                                 | 3     | 0       | –       | 0    | –    | –      | –             | –              | $(0,0,0,0)$        | $D_3 = 0$ |

the resulting balance with respect to contrast $\lambda_1$. In the subclass with $c_2 = 1/2$, the covariate means for treatment groups with $D_{1i} = 1$ and $D_{1i} = 0$ are $(1/3, 2/3, -1, 0)$ and $(2/3, 2/3, 2/3, 2/3)$ respectively, which shows that the covariate difference between treatment groups is not $(0, 0, 0, 0)$ at the same level of $c_2$.

5.2. Simulation studies

In § 5.1, we were able to stratify the units into subclasses such that each subclass has only one level of $(c_1, c_2)$, but this is usually not the case when there are many values of $(c_1, c_2)$. We now consider such a case with $T = 3$ and $N = 800$ units and covariates $X_{ik} \sim \text{Norm}(0, 1)$, $k = 1, 2, 3$. The assignment mechanisms we consider are two different multinomial logistic models, with $\text{pr}(W_i = t \mid X_i) = \exp(\beta'_t X_i)/\sum_{t=1}^{3} \exp(\beta'_t X_i)$. Mechanism I has $\beta_1 = \beta_2 = \beta_3 = (0, 0, 0)$, i.e. complete randomization, and Mechanism II has $\beta_1 = (0, 0, 0)$, $\beta_2 = (0.75, 0.25, 0.5)$, $\beta_3 = (0.25, 0.75, 0.5)$. We consider four contrasts: $\lambda_1 = (1/3, 2/3, -1, 0)$, $\lambda_2 = (1, -1, 0)$, $\lambda_3 = (1, 0, -1)$, $\lambda_4 = (0, 1, -1)$. We implement the Algorithm, balancing on the sgn bifurcation $c_3$ps of $\lambda_1$ and $\lambda_2$.

We repeat the simulations 100 times and report the results in Table 3. For assignment mechanism I, a completely randomized experiment, a small increase in balance is observed after applying the Algorithm, illustrating that using the estimated $c_3$ps can
reduce random imbalance. We find that, for assignment mechanism II, after implementing the Algorithm using simple subclassification, the differences in covariate means are substantially diminished, to less than 0.1 on average.

6. Discussion

csps methodology focuses on creating, at the design stage, treatment groups with balanced covariates. Once balanced groups are created and outcomes are measured, treatment effects can be estimated using more sophisticated methods than the simple comparison of means. For example, recent work suggests that weighting estimators using estimated propensity scores are generally worse than using imputation-based estimators, which use models to impute missing potential outcomes (e.g. Gutman and Rubin [26, 27]), advice which can be traced back to Rubin [28] and Cochran and Rubin [29].

Acknowledgments

We thank the referees for helpful comments and Rui Dong and Ke Zhu for helpful discussions.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Shasha Han is currently a postdoctoral researcher in Beijing International Center for Mathematical Research, Peking University; and Department of Health Care Policy, Harvard Medical School. Her research focuses on forge a scientific basis of health interventions through tackling problems arising in evidence-based decisions in the presence of complexity and uncertainty, from perspectives on health policy, managerial decision-making and individual medication. She completed her PhD from National University of Singapore, with the pleasure being advised by Joel Goh, Melvyn Sim, Donald Rubin. She has expertise in develops and applies causal inference and optimization methods with the large real-world data of health system. Her research has appeared in Annals of Internal Medicine.

Donald B. Rubin is currently Professor in the Yau Center for Mathematical Sciences, Tsinghua University; Murray Schusterman Senior Research Fellow, Fox Business School, Temple University; and Professor of Statistics Emeritus, Harvard University. He has been elected to be a Fellow/Member/Honorary Member of: the Woodrow Wilson Society, Guggenheim Memorial Foundation, Alexander von Humboldt Foundation, American Statistical Association, Institute of Mathematical Statistics, International Statistical Institute, American Association for the Advancement of Science, American Academy of Arts and Sciences, European Association of Methodology, the British
Academy, and the U.S. National Academy of Sciences. As of 2021, he has authored/coauthored nearly 500 publications (including ten books), has four joint patents, and for many years has been one of the most highly cited authors in the world, with currently over 340,000 citations, and over 2,500 per year in recent years (Google Scholar). Of his many publications with over 1,000 citations each, over ten of them are solely authored by Rubin. He has received honorary doctorate degrees from Otto Friedrich University, Bamberg, Germany; the University of Ljubljana, Slovenia; Universidad Santo Tomás, Bogotá, Colombia; Uppsala University, Sweden; and Northwestern University, Evanston, Illinois. He has also received honorary professorships from the University of Utrecht, The Netherlands; Shanghai Finance University, China; Nanjing University of Science & Technology, China; Xi’an University of Technology, China; and University of the Free State, Republic of South Africa. He is a widely sought international lecturer and consultant on a variety of statistical topics.

ORCID

Shasha Han  http://orcid.org/0000-0001-7388-8125

References

[1] Haukoos JS, Lewis RJ. The propensity score. JAMA. 2015;314:1637–1638.
[2] Rosenthal R, Rosnow RL, Rubin DB. Contrasts and effect sizes in behavioral research: a correlational approach. Cambridge University Press; 2000.
[3] Snedecor GW, Cochran WG. Statistical methods. 1st ed. Iowa State University Press; 1967.
[4] Wu CJ, Hamada MS. Experiments: planning, analysis, and optimization. Vol. 552. John Wiley & Sons; 2011.
[5] Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983;70:41–55.
[6] Austin PC. A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. Stat Med. 2008;27:2037–2049.
[7] Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate Behav Res. 2011;46:399–424.
[8] Brookhart MA, Schneeweiss S, Rothman KJ, et al. Variable selection for propensity score models. Am J Epidemiol. 2006;163:1149–1156.
[9] Caliendo M, Kopeinig S. Some practical guidance for the implementation of propensity score matching. J Econ Surv. 2008;22:31–72.
[10] Cintina I, Love I. Re-evaluating microfinance: evidence from propensity score matching. World Bank Econ Rev. 2019;33:95–115.
[11] d’Agostino RB. Tutorial in biostatistics: propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. Stat Med. 1998;17:2265–2281.
[12] Dehejia RH, Wahba S. Propensity score-matching methods for nonexperimental causal studies. Rev Econ Stat. 2002;84:151–161.
[13] Guo S, Fraser MW. Propensity score analysis: statistical methods and applications. Vol. 11. SAGE publications; 2014.
[14] Klompmaker S, van Hilst J, Wellner UF, et al. Outcomes after minimally-invasive versus open pancreatoduodenectomy: a pan-european propensity score matched study. Ann Surg. 2020;271:356–363.
[15] Thourani VH, Kodali S, Makkar RR, et al. Transcatheter aortic valve replacement versus surgical valve replacement in intermediate-risk patients: a propensity score analysis. The Lancet. 2016;387:2218–2225.
[16] Imbens GW. The role of the propensity score in estimating dose-response functions. Biometrika. 2000;87:706–710.
[17] Horvitz DG, Thompson DJ. A generalization of sampling without replacement from a finite universe. J Am Stat Assoc. 1952;47:663–685.
[18] Lechner M. Identification and estimation of causal effects of multiple treatments under the conditional independence assumption. In: Lechner M, Pfeiffer F, editors. Econometric evaluation of labour market policies. Heidelberg: Physica; 2001. p. 43–58.

[19] Imai K, van Dyk DA. Causal inference with general treatment regimes. J Am Stat Assoc. 2004;99:854–866.

[20] LaLonde RJ. Evaluating the econometric evaluations of training programs with experimental data. Am Econ Rev. 1986;76:604–620.

[21] Kaplan K, Mashash M, Williams R, et al. Effect of light flashes vs sham therapy during sleep with adjunct cognitive behavioral therapy on sleep quality among adolescents: a randomized clinical trial. JAMA Netw Open. 2019;2:Article ID e1911944.

[22] Rubin DB. Discussion of randomization analysis of experimental data in the fisher randomization test. J Am Stat Assoc. 1980;75:591–593.

[23] Rubin D. Estimating causal effects of treatments in randomized and nonrandomized studies. J Educ Psychol. 1974;66:688–701.

[24] Imbens GW, Rubin DB. Causal inference in statistics, social, and biomedical sciences. Cambridge University Press; 2015.

[25] Rubin DB. Formal modes of statistical inference for causal inference. J Stat Plan Inference. 1990;25:279–292.

[26] Gutman R, Rubin D. Estimation of causal effects of binary treatments in unconfounded studies with one continuous covariate. Stat Methods Med Res. 2017;26:1199–1215.

[27] Gutman R, Rubin DB. Robust estimation of causal effects of binary treatments in unconfounded studies with dichotomous outcomes. Stat Med. 2013;32:1795–1814.

[28] Rubin DB. Matching to remove bias in observational studies. Biometrics. 1973;29(1):159–183.

[29] Cochran WG, Rubin DB. Controlling bias in observational studies: a review. Sankhýa. 1973;35:417–446.