Original Article

Synergistic markers based on inter-covariate association estimate treatment option with lower propensity to covariates

Lawrence Wing Chi Chan[1]^, Alan Sihoe[2]

[1]Department of Health Technology and Informatics, Hong Kong Polytechnic University, Hong Kong, China; [2]Cardiothoracic Surgery, Gleneagles Hong Kong Hospital, Hong Kong, China

Contributions: (I) Conception and design: Chan LWC; (II) Administrative support: Both authors; (III) Provision of study materials or patients: Chan LWC; (IV) Collection and assembly of data: Chan LWC; (V) Data analysis and interpretation: Both authors; (VI) Manuscript writing: Both authors; (VII) Final approval of manuscript: Both authors.

Correspondence to: Lawrence Wing Chi Chan. Department of Health Technology and Informatics, Hong Kong Polytechnic University, Hong Kong, China. Email: wing.chi.chan@polyu.edu.hk.

Background: Propensity constitutes a common problem in identifying clinical outcome prediction model whose covariates include the treatment option, which is assumed to be randomly assigned but indeed dependent of other covariates in the training data. The genuine effect of treatment option cannot be elucidated under the influence of propensity. Existing approaches, such as matched-pairs study design, still cannot solve the problem for imbalanced or small datasets.

Methods: This work proposed an anti-propensity estimate of treatment option, which is generated by support vector classifier based on two synergistic markers that represent the lower and upper limits of inter-covariate association level. The algorithm for generating the synergistic markers was illustrated and the performance was evaluated on a public dataset of gene expression levels, which were obtained from surgically excised tumor samples in non-small cell lung cancer (NSCLC) patients where treatment option, i.e., adjuvant therapy or not, was known.

Results: Six covariates represented by the expression levels of *ZNF217*, *ERCC3*, *PMS1*, *PIK3CB*, *BARD1*, and *MAPK1*, were selected to generate two synergistic markers and classifier for estimating the adjuvant therapy option with substantially attenuated propensity. The estimation accuracy attained an area under the receiver-operating characteristics curve, 0.78, in the test set.

Conclusions: The proposed synergistic markers demonstrated a parsimonious and anti-propensity estimation of treatment option, which is ready for the further evaluation and application in the clinical outcome prediction model.

Keywords: Synergistic markers; treatment option; propensity; inter-covariate association

Submitted Oct 11, 2022. Accepted for publication Mar 13, 2023. Published online Apr 12, 2023.
doi: 10.21037/atm-22-5006
View this article at: https://dx.doi.org/10.21037/atm-22-5006

Introduction

Besides the clinical experience and guidelines, decision of a treatment option, particularly surgery and adjuvant therapy, is usually supported by the statistical analyses, including Kaplan-Meier (KM) estimators, Cox regression model and logistic regression model, and, in recent years, by the emerging artificial intelligent tools (1,2). Those supporting approaches examine the causal effect of the treatment on the clinical outcome or benefit. To predict the outcome

^ ORCID: 0000-0001-6451-2273.

in terms of risk score or dichotomy, such as hazard rate or recurrence, the model is built with a fixed panel of selected covariates (3). The candidate covariates include but not limited to treatment option, patient demographics, clinical information, and tumor characteristics. In the training stage of model, the candidate covariates are prioritized according to their effects on the outcome and the top covariates are selected for building the model.

The above-mentioned models, whose covariates include treatment option, could be easily trained and the implementation is straightforward, based on the assumption that the treatment assignment is randomized and independent of the other covariates. In practice, particularly for observational studies, the treatment is not randomized but assigned by the clinical deliberation with reference to the other covariates. The treatment selection is usually affected by many factors including age, sex, comorbidities, and genomic profile (4,5). For example, younger patients (18–54 years) were more likely to receive adjuvant chemotherapy (31%) than post-operative observation (21%) (6). Such dependence is illustrated by the fact that the covariate distributions depart substantially between the treatment and control groups. Therefore, the trained model is biased to the other covariates rather than learning the genuine effect of treatment on the outcome.

Highlight box

Key findings

- The proposed synergistic markers can estimate the adjuvant therapy option with substantially attenuated propensity to covariates.

What is known and what is new?

- The existence of propensity is known in observational studies and clinical studies where the treatment is not randomized but assigned by the clinical deliberation with reference to the other covariates. Matched-pairs study design is susceptible to generalization problems on small or imbalanced samples.
- Anti-propensity estimate of treatment option can be generated by the supervised learning model of the proposed synergistic markers.

What is the implication, and what should change now?

- The proposed synergistic markers can be presented as continuous covariates in clinical outcome prediction model where the model could quantify individually the patient benefit subject to the treatment option. In addition to traditional statistical patient stratification, prediction model supported by synergistic markers can advance the personalized treatment with optimal patient benefit.

To suppress the bias of covariate, researchers developed methods for estimating the propensity score for each subject using discriminant analysis or logistic regression of treatment option on covariates. The propensity score could help make a valid causal inference by implementing matched-pairs study design, weighting the cases in training the model or acting as an additional covariate in the model (7). However, the estimation of propensity score is susceptible to generalization problems of parametric model based on small or imbalanced samples. The interactions between covariates, which could have been considered in treatment decision, are also ignored by the above-mentioned remedial methods. Therefore, it is of critical importance to develop an algorithmic method for integrating a set of covariates to generate so-called synergistic markers in this paper that can truly identify the difference between treatment and control groups, to get rid of the propensity to individual covariates. In this work, the anti-propensity estimate of treatment option is defined as the output generated by a supervised learning model, such as support vector machine (SVM), where the synergistic markers constitute the inputs. The output possesses a score and its dichotomized estimate so that the dependence on covariates is highly reduced when compared with the original treatment option.

Methods

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). For developing outcome prediction model, data including the clinical information, markers, features, facts, treatment option and outcome are collected and used to train and test the model. The clinical information, markers, features, and facts are model covariates where the i^{th} covariate of the k^{th} subject or case is denoted by $x_i(k)$. The distributions of covariates may largely deviate from normal distribution so that the model may be predisposed to biased prediction results if left uncorrected. Methods, such as rank-based inverse normal transformation (INT), can be applied to symmetrize and concentrate the distribution to standard normal, N(0,1). The values of the i^{th} covariate across N subjects form the following vector.

$$u_i = \left[z_i(1), z_i(2), \ldots, z_i(n) \right]^T \quad [1]$$

where the covariate, $z_i(k)$, follows a near-normal distribution, ~N(0,1), across subjects. The dot product, $u_i \cdot u_i$

I

	A	B	C	D	E
A	0	0.875513	0.761413	0.704578	0.635384
B	0.875513	0	0.623233	0.620385	0.50633
C	0.761413	0.623233	0	0.637049	0.787873
D	0.704578	0.620385	0.637049	0	0.477486
E	0.635384	0.50633	0.787873	0.477486	0

II

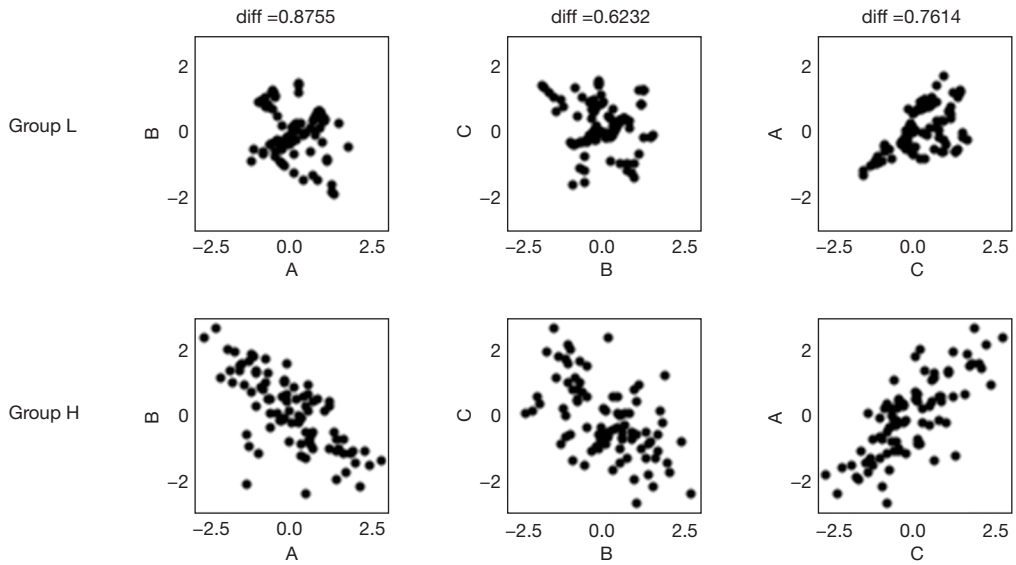

Figure 1 Hypothetical example of five covariates. (I) Matrix D; (II) scatter plots of covariate pairs (A, B), (B, C), and (C, A) of groups L and H.

and $u_i \cdot u_j$, tend to, respectively, 1 and the Pearson correlation coefficient between the ith and jth covariates when n is large enough, approaching the population size.

The association level between any two covariates for the treatment group is denoted by $C_T(i, j)$, and that for the non-treatment group, $C_N(i, j)$, given by the following.

$$C_T(i,j) = \left|u_{Ti} \cdot u_{Tj}\right| = \left|\frac{1}{n_T}\sum_{k_T=1}^{n_T} z_i(k_T) z_j(k_T)\right| \quad [2]$$

$$C_N(i,j) = \left|u_{Ni} \cdot u_{Nj}\right| = \left|\frac{1}{n_N}\sum_{k_N=1}^{n_N} z_i(k_N) z_j(k_N)\right| \quad [3]$$

where u_{Ti} and u_{Tj} represent the vectors containing the i^{th} and j^{th} covariates across the treatment group; u_{Ni} and u_{Nj} represent the vector containing the ith and jth covariates across the non-treatment group. The number of candidate covariates is denoted by m. Two groups are further defined as group H with higher overall association level and group L with lower overall association level, subject to the direction of the difference in overall association level given by the following formula.

$$\Delta = \sum_{i\neq j, i=1, j=1}^{i=m, j=m} C_T(i,j) - \sum_{i\neq j, i=1, j=1}^{i=m, j=m} C_N(i,j) \quad [4]$$

If $\Delta \geq 0$, the treatment group is defined as group H and the non-treatment group, L. Otherwise, the treatment group is defined as group L and the non-treatment group, H. The difference in association level between the ith and jth covariates is formulated by the (i,j)th element of a matrix, D, defined by the following formula.

$$D(i,j) = \begin{cases} C_H(i,j) - C_L(i,j) & for\ i \neq j \\ 0 & for\ i = j \end{cases} \quad [5]$$

An example of D, a 5×5 matrix generated from the hypothetical data of five covariates, is given as *Figure 1A*. The scatter plots of (A, B), (B, C) and (C, A) of groups L and H are shown in *Figure 1B*. When the association between covariates in group L is substantially weaker than that in group H, the corresponding value in matrix D is relatively high.

In the covariate sorting process, half of the off-diagonal elements, either upper or lower triangular, are extracted to

I

	{A,B}	C	D	E
	0.875513	0.761413	0.704578	0.635384
	0.875513	0.623233	0.620385	0.50633
C	1.384646	0	0.637049	0.787873
D	1.324963	0.637049	0	0.477486
E	1.141714	0.787873	0.477486	0

II

	{A,B,C}	D	E
	1.636926	0.704578	0.635384
	1.498746	0.620385	0.50633
C	1.384646	0.637049	0.787873
D	1.962012	0	0.477486
E	1.929587	0.477486	0

Figure 2 Covariate sorting process. (I) Covariates A and B are selected with a maximum difference, 0.8755, and merged to form list L and the first column is re-calculated; (II) covariate C joins list L with a maximum difference, 0.7614, and the first column is re-calculated.

form a list. The maximum of the list and the corresponding covariate pair are identified. The selected covariate list with m' covariates is denoted by $L_{m'}$. For the above example of D, the maximum is 0.8755, the covariates A and B are selected and L_2 is {A,B}.

The third covariate is added to L_2 in condition that the sum of its D(i,j) values with A and B is the highest amongst the other covariates. To find the highest sum, columns A and B of matrix D are added element by element. The result of column addition is shown in *Figure 2A*.

From the first column of the result, the covariate C yields the highest sum of D(i,j) values with A and B so that C is added to the list, giving L_3, {A,B,C}. To determine the fourth covariate, columns {A,B} and C are added element-by-element to give the result as shown in *Figure 2B*.

From the first column again, the covariate D yields the highest sum of D(i,j) values with A, B and C so that D is added to the list, giving L_4, {A,B,C,D}.

For adding the subsequent covariates to the list, the above steps of column addition and optimal value search are repeated. For m' ranging from 2 to m, an ordered list of covariates can be formed in the descending order of the corresponding difference in cumulative association level, $\Delta_{m'}$, given as below.

$$\Delta_{m'} = \sum_{i\neq j, i=1, j=1}^{i=m', j=m'} C_H(i,j) - \sum_{i\neq j, i=1, j=1}^{i=m', j=m'} C_L(i,j) \quad [6]$$

The synergistic markers are derived as follows. For m' ranging from 2 to m, the cumulative association level of group H or group L must fall within an interval whose lower and upper bounds are given by the sample means of two synergistic markers, s_1 and s_2. For m' covariates, twice of the cumulative association level is elaborated to give the lower bound by the following inequality.

$$\begin{aligned} 2\sum_{i\neq j, i=1, j=1}^{i=m', j=m'} C(i,j) &= 2\sum_{i\neq j, i=1, j=1}^{i=m', j=m'} \left|\frac{1}{n}\sum_{k=1}^{n} z_i(k) z_j(k)\right| \\ &\geq \sum_{i\neq j, i=1, j=1}^{i=m', j=m'} \frac{1}{n}\sum_{k=1}^{n} 2 z_i(k) z_j(k) \\ &\geq \frac{1}{n}\sum_{k=1}^{n}\left[\left(\sum_{i=1}^{m'} z_i(k)\right)^2 - \sum_{i=1}^{m'}\left(z_i(k)\right)^2\right] \\ &\geq \frac{1}{n}\sum_{k=1}^{n} s_1(k) \end{aligned} \quad [7]$$

where $s_1(k)$ is the first synergistic marker given by the following formula.

$$s_1(k) = \left(\sum_{i=1}^{m'} z_i(k)\right)^2 - \sum_{i=1}^{m'}\left(z_i(k)\right)^2 \quad [8]$$

The upper bound is elaborated by the following inequality.

$$\begin{aligned} 2\sum_{i\neq j, i=1, j=1}^{i=m', j=m'} C(i,j) &= 2\sum_{i\neq j, i=1, j=1}^{i=m', j=m'} \left|\frac{1}{n}\sum_{k=1}^{n} z_i(k) z_j(k)\right| \\ &\leq \sum_{i\neq j, i=1, j=1}^{i=m', j=m'} \frac{1}{n}\sum_{k=1}^{n} 2\left|z_i(k)\right|\left|z_j(k)\right| \\ &\leq \frac{1}{n}\sum_{k=1}^{n}\left[\left(\sum_{i=1}^{m'}\left|z_i(k)\right|\right)^2 - \sum_{i=1}^{m'}\left(z_i(k)\right)^2\right] \\ &\leq \frac{1}{n}\sum_{k=1}^{n} s_2(k) \end{aligned} \quad [9]$$

where $s_2(k)$ is the second synergistic marker given by the following formula.

$$s_2(k) = \left(\sum_{i=1}^{m'}\left|z_i(k)\right|\right)^2 - \sum_{i=1}^{m'}\left(z_i(k)\right)^2 \quad [10]$$

The number of covariates in the ordered list to be included for generating the markers can be estimated by machine learning. If SVM is used, the support vector classifier (SVC) is trained with the inputs $S_{m'}(k) = [s_1(k),$

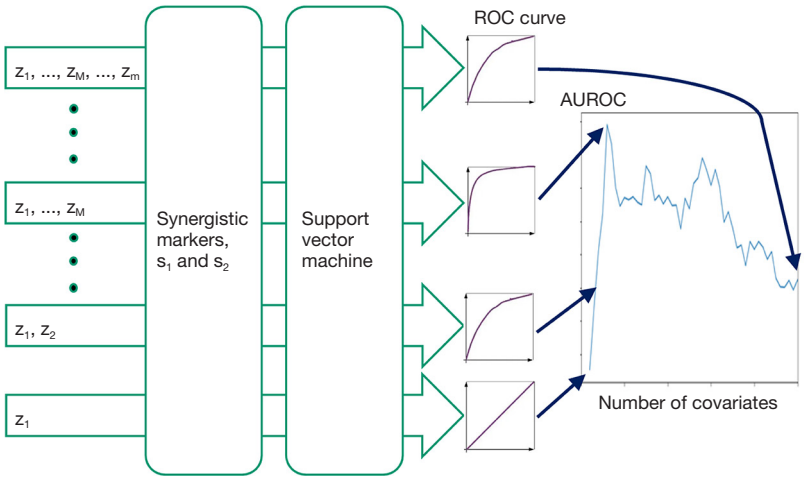

Figure 3 Implementation of iterative procedure for obtaining the optimal number of covariates, M, constituting the synergistic markers. ROC, receiver operating characteristics; AUROC, area under ROC curve.

$s_2(k)]^T$ generated by m' covariates and the output given by a score, $y_{m'}(k)$, and its estimate of the binary treatment option. The trained classifier is represented by,

$$y_{m'}(k) = \sum_{i \in SV} \alpha_i \Phi\left(S_{m'}(i), S_{m'}(k)\right) + b \quad [11]$$

where SV is a set of indices of support vectors, α_i is the coefficient corresponding to the i^{th} support vector, $S_{m'}(i)$, and b is a constant. The kernel, Φ, can be represented by a nonlinear function, such as radial basis function (RBF), polynomial, and sigmoid function. The output score, $y_{m'}(k)$, is evaluated against the original treatment option through the receiver operating characteristics (ROC) analysis.

For each m' increasing from 2 to m, the area under the ROC curve (AUROC) is recorded as the performance of the SVC, denoted by A(m'). The optimal number of covariates, M, and thus the corresponding synergistic markers, $s_1(k)$ and $s_2(k)$, are identified by the highest A(m'), i.e., A(M). The implementation of the iterative procedure is illustrated in *Figure 3*.

Statistical analysis

ROC analysis was performed to estimate the AUROC as the probability that the model ranks a random treatment case more highly than a random control case. The propensity score is defined as the probability of an individual being treated (or not being treated) in condition to covariate values. Logistic regression was used to estimate the conditional probability.

Results

The public dataset 'non-small cell lung cancer (NSCLC) Radiogenomics' was derived from Stanford University School of Medicine and Palo Alto Veterans Affairs Healthcare System (8), with 186 NSCLC subjects' RNA sequencing data and thus gene expression data, obtained from surgically excised tumor samples. Subjects were recruited between April 7th, 2008, and September 15th, 2012, and all were in the early stage of NSCLC. Treatment option, i.e., adjuvant therapy or not, was included for analysis. Platinum-doublet chemotherapy represents standard adjuvant treatment for early-stage resected NSCLC. In clinical trials, adjuvant epidermal growth factor receptor (EGFR) tyrosine kinase inhibitors demonstrated prolonged disease-free survival (DFS) and comparable clinical effectiveness in EGFR-mutated NSCLC patients (9). After data pre-processing, 195 genes' expression levels representing the covariates for each case were extracted from the dataset. The synergistic markers were generated from the training set of 166 cases and evaluated by the test set of 20 cases.

The association levels of 18,915 unique covariate pairs were computed for each of the treatment and non-treatment groups. The distributions of association levels are shown and compared in *Figure 4A*. The sum of association levels of the treatment group is higher than that of the non-treatment group. The treatment group is thus defined as group H and non-treatment group, group L. The covariate pair, ('ZNF217', 'ERCC3'), gave the highest difference

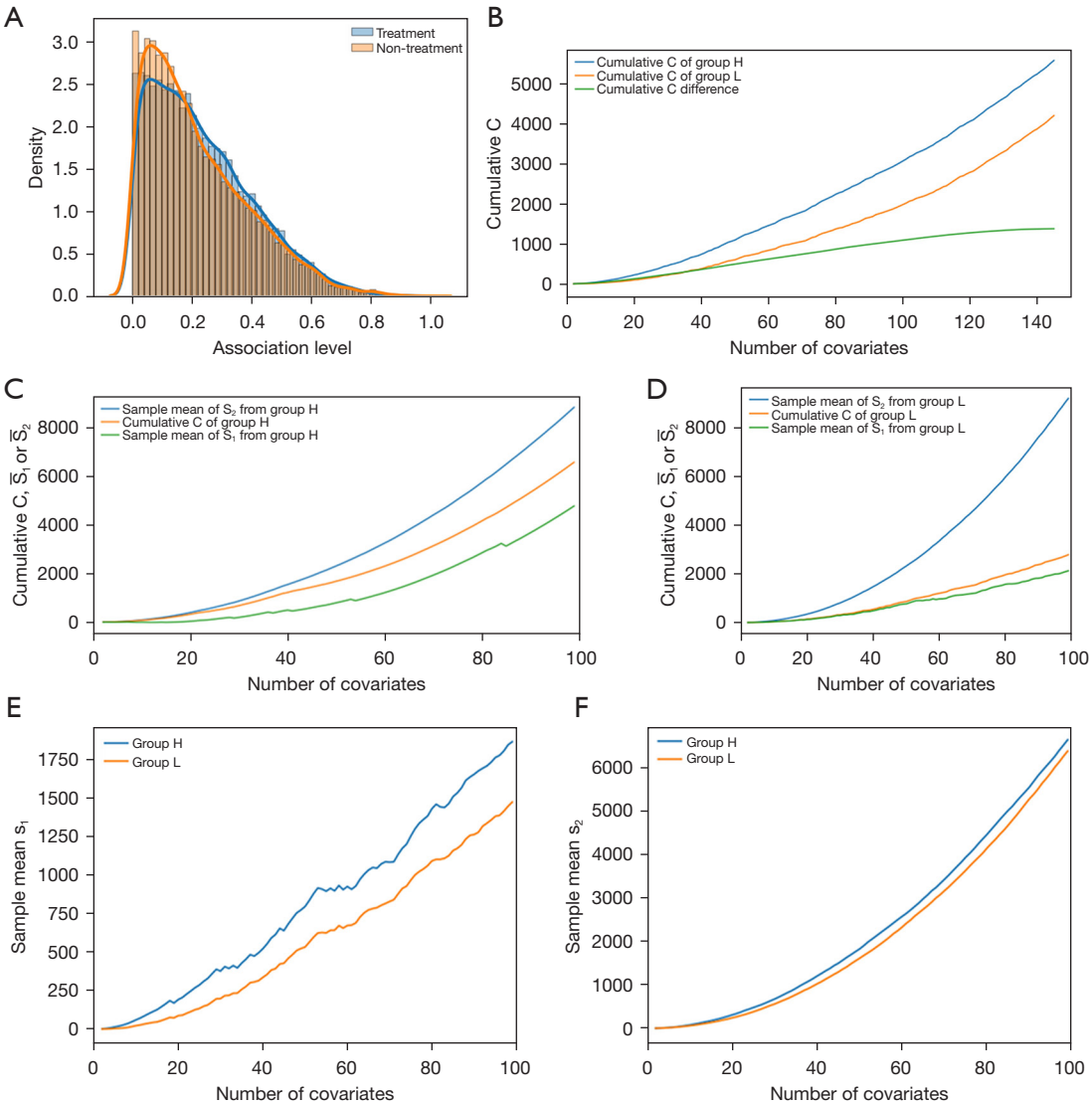

Figure 4 Association levels of covariates. (A) Distributions across all possible covariate pairs; (B) trends of cumulative association level of groups H and L and their difference; (C) trends of sample means of synergistic markers s_1 and s_2, and cumulative association level of group H; (D) trends of sample means of synergistic markers s_1 and s_2, and cumulative association level of group L; (E) sample means of s_1; (F) sample means of s_2.

in association level between groups H and L, C_H-C_L. The ordered list was initialized by this pair. The subsequent covariates were added to the list one by one according to their cumulative association levels. *Figure 4B* shows the increasing trends of cumulative association level of groups H and L and their difference when the number of covariates in the ordered list increases.

As shown in *Figure 4C* and *Figure 4D*, the sample means of the synergistic markers z_1 and z_2 serve as the lower and upper bounds of the cumulative association level of their corresponding groups for any number of covariates in the ordered list. It is shown in *Figure 4E* that the sample mean of z_1 in group H is higher than in group L and the difference increases with the number of covariates in the ordered list. The same observation is shown in *Figure 4F*.

SVC was trained with the synergistic markers, s_1 and s_2, as input and the treatment option as target output. RBF was used as kernel. For each covariate number, AUROC was computed to evaluate the performance of SVC on training data. In *Figure 5A*, the AUROC is plotted against the

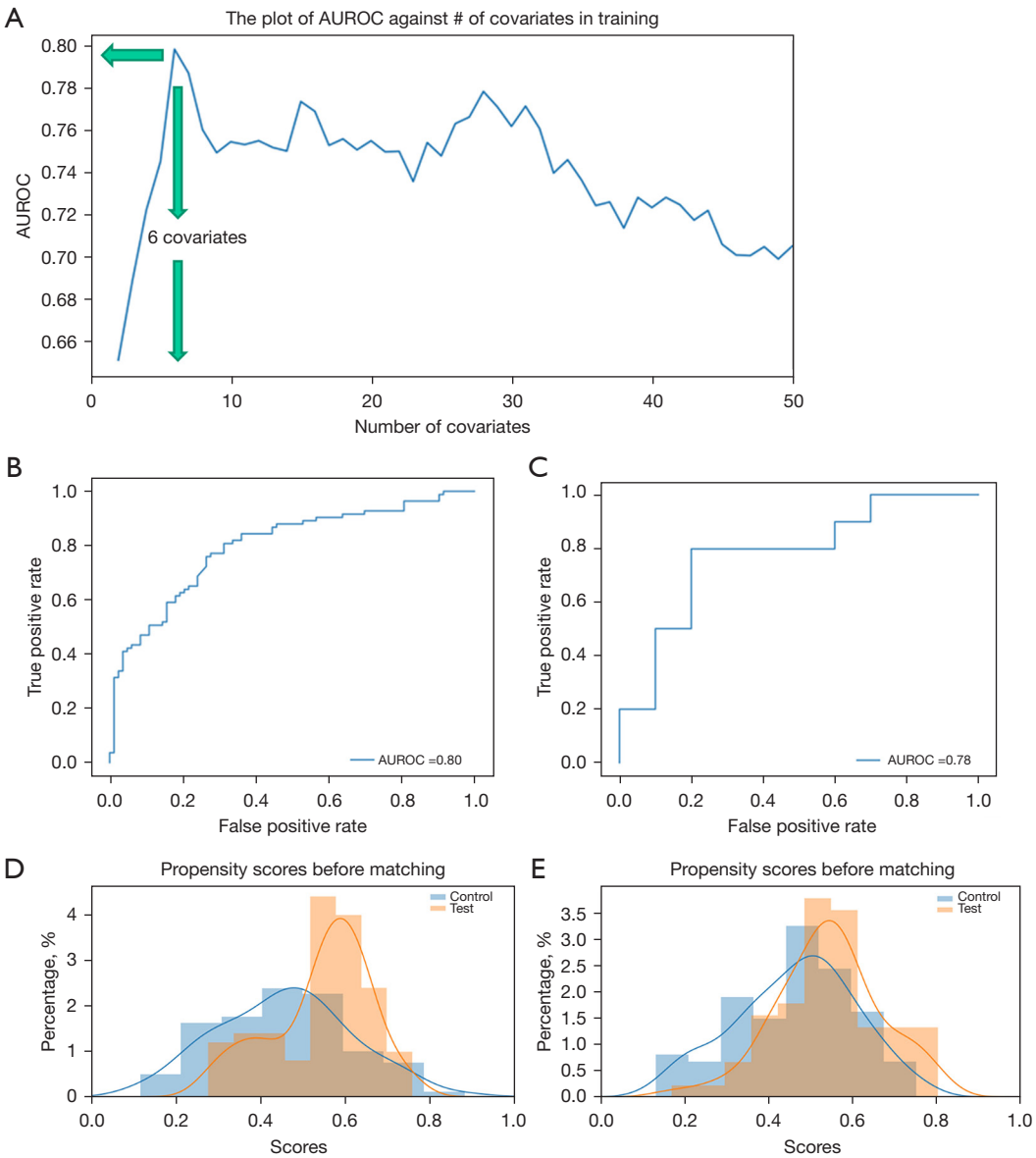

Figure 5 (A) The plot of AUROC against the number of covariates; (B,C) ROC curves of SVC using synergistic markers on training and test data; (D,E) propensity score distributions of the original treatment option and its SVC estimate. ROC, receiver operating characteristics; AUROC, area under ROC curve; SVC, support vector classifier.

number of covariates, which were used for generating the synergistic markers. It was shown that the AUROC attained the maximum, 0.80, when six covariates in the ordered list was used to generate the synergistic markers. The genes, *ZNF217*, *ERCC3*, *PMS1*, *PIK3CB*, *BARD1*, and *MAPK1*, were selected where their expression levels represent the six covariates.

Using training and test sets, the ROC curves of the trained SVC with synergistic markers based on 6 covariates were plotted in *Figure 5B* and *Figure 5C* respectively. The test performance attained 0.78, which was close to the training performance.

The python module, "pymatch" (https://github.com/benmiroglio/pymatch), was used to assess the propensity of covariates on the actual treatment option and compare with that on the SVC prediction. The propensity scores were

computed based on these six covariates in the ordered list to avoid overfitting of regression model. The distributions of propensity scores were compared between treatment and non-treatment groups based on the actual treatment option and the predicted treatment in *Figure 5D* and *Figure 5E* respectively. A significant difference in median propensity score between treatment and non-treatment groups was found on the actual treatment option ($P<0.05$), but not on that predicted by the synergistic markers ($P>0.05$).

Discussion

In developing outcome prediction model, the clinical information, markers, features, patient demographics and treatment option constitute the model covariates and the clinical outcome represents the target for training, test and implementing the model. The distributions of covariates would largely deviate between the treatment and control groups and therefore the model will generate biased prediction results if such propensity is left uncorrected.

This work proposes the synergistic markers that predict the treatment option based on the inter-covariate association level, instead of the magnitudes of individual covariates. Such prediction can get rid of the propensity to certain covariates influencing the clinical decision. Non-parametric method is used to generate the synergistic markers based on a relatively high-dimensional but small dataset (e.g., covariates >50, $N<200$). It avoids the curse of dimensionality and overfitting problem caused by parametric model. In the training process, the original treatment option is replaced by the synergistic markers' estimate that can impose the genuine treatment effect in the outcome prediction model (10).

One of the key features of the algorithm is the generation of an order list of covariates, through which the difference in inter-covariate association levels between two groups could be prioritized and becomes monotonically increasing. As demonstrated by fitting the public data, the two synergistic markers form the lower and upper bounds of the inter-covariate association level in both treatment and control groups. For each synergistic marker, its value of treatment group is always higher than that of control group. With such discriminating ability, machine learning was applied to train a classifier based on the synergistic markers as input and the treatment option as the target. The distributions of propensity score of the predicted treatment option to the selected covariates demonstrated no significant difference between two groups. In contrast, the distributions for the original treatment option showed significant deviation. The results demonstrated that the treatment option predicted by the synergistic markers can reduce or eliminate the propensity of the actual treatment option to covariates.

In survival analysis, KM curves for two or more treatment levels are plotted and compared by the log-rank test. Two treatment levels may be represented by adjuvant therapy and POB (i.e., no adjuvant therapy). The outcome may be survival or disease relapse time. The significant difference in the clinical outcome between the treatment levels can be examined by the survival analysis. For example, it was found by Salazar *et al.* that NSCLC patients who received adjuvant chemotherapy later had a significantly better survival when compared with patients treated with surgery alone (6). Such analysis cannot quantify the change in survival or relapse time subject to the treatment option, and therefore cannot indicate the individual's benefit. In contrast, the proposed synergistic markers can be presented as continuous covariates in an outcome prediction model. As such, the model could quantify the patient benefit subject to the treatment option, instead of indicating the significant difference only.

Test of the algorithm on different clinical datasets is continuing. Future work is to look for large dataset with relatively balanced ratio between treatment and non-treatment group sizes. Oversampling could be further enhanced so that the prediction of treatment option is robust on the data with highly imbalanced group sizes.

Conclusions

Synergistic markers are proposed to generate an anti-propensity estimate of treatment option using the supervised learning. The outcome prediction model incorporating such anti-propensity estimate can personalize the patient benefit subject to the treatment option.

Acknowledgments

The authors would like to thank Mr. Jiqiao LU at Hong Kong Polytechnic University for technical assistance and advice.

Funding: None.

Footnote

Conflicts of Interest: Both authors have completed the

ICMJE uniform disclosure form (available at https://atm.amegroups.com/article/view/10.21037/atm-22-5006/coif). LWCC reports that US Non-Provisional Patent (17/936,892) and China Invention Patent (202211221439.8) were filed on 8 Oct, 2021. AS has no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

References

1. Katzman JL, Shaham U, Cloninger A, et al. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. BMC Med Res Methodol 2018;18:24.
2. Zhou C, Hu J, Wang Y, et al. A machine learning-based predictor for the identification of the recurrence of patients with gastric cancer after operation. Sci Rep 2021;11:1571.
3. Little RJ, Rubin DB. Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. Annu Rev Public Health 2000;21:121-45.
4. Kim YJ, Oremus M, Chen HH, et al. Factors affecting treatment selection and overall survival for first-line EGFR-tyrosine kinase inhibitor therapy in non-small-cell lung cancer. J Comp Eff Res 2021;10:193-206.
5. Li T, Kung HJ, Mack PC, et al. Genotyping and genomic profiling of non-small-cell lung cancer: implications for current and future therapies. J Clin Oncol 2013;31:1039-49.
6. Salazar MC, Rosen JE, Wang Z, et al. Association of Delayed Adjuvant Chemotherapy With Survival After Lung Cancer Surgery. JAMA Oncol 2017;3:610-9.
7. West SG, Thoemmes F. Campbell's and Rubin's perspectives on causal inference. Psychol Methods 2010;15:18-37.
8. Bakr S, Gevaert O, Echegaray S, et al. A radiogenomic dataset of non-small cell lung cancer. Sci Data 2018;5:180202.
9. He Q, Liu J, Cai X, et al. Comparison of first-generation EGFR-TKIs (gefitinib, erlotinib, and icotinib) as adjuvant therapy in resected NSCLC patients with sensitive EGFR mutations. Transl Lung Cancer Res 2021;10:4120-9.
10. Louhimo R, Laakso M, Heikkinen T, et al. Identification of genetic markers with synergistic survival effect in cancer. BMC Syst Biol 2013;7 Suppl 1:S2.

Cite this article as: Chan LWC, Sihoe A. Synergistic markers based on inter-covariate association estimate treatment option with lower propensity to covariates. Ann Transl Med 2023;11(10):348. doi: 10.21037/atm-22-5006

