Han F, et al., HSOA J Altern Complement Integr Med 2025, 11: 626
DOI: 10.24966/ACIM-7562/100626

HSOA Journal of
Alternative, Complementary & Integrative Medicine

Research Article

Selection of Optimal Treatment Regimen for Individuals Using Multivariate Assessments During Clinical Follow-up

Feng Han[1], Linlin Zhang[2], Danhui Yi[3]*, Jingsheng Zhang[1] and Kaishan Han[4]

[1]*Institute of Mathematics and Physics, Beijing Union University, Beijing, China*

[2]*Foreign Languages Department, Beijing Union University, Beijing, China*

[3]*School of Statistics, Renmin University of China, Beijing, China*

[4]*School of Mathematics, North University of China, Taiyuan, China*

Abstract

Background: In many cases, the treatment of the patients is multi-course, during which varying characteristics and indicators are present. Therefore, the clinical data motivated by ongoing trials tends to be high-dimensional. This is where longitudinal single biomarker Covariate-Specific Treatment Effect (CSTE) curves are universally employed to graphically explore the selection of optimal treatment regimen and provide information about the effect of a therapeutic intervention.

Methods: We extended the single biomarker clinical follow-up studies to multi-biomarker ones and devised the CSTE function using separable univariate additive model, spline function, and the generalized estimating equation (GEE). When the data is increasingly high-dimensional, penalised GEE is taken into account. At the same time, the discretization of continuous variables is considered, the sub-indicators of the CSTE curve of individual treatment are examined, and the confidence band is constructed by means of Hotelling Tube methods.

Results: Simulation studies show that as the sample size increases, the estimated values approximate the true values, the length of the confidence intervals decreases while the coverage rate increases, indicating good large-sample properties.

Conclusion: The results of the simulation studies show that the proposed method is feasible in handling high-dimensional covariates. It can quantify variability associated with treatment selection and serve as a reliable and uniform inferential tool in clinical follow-up studies.

Keywords: Clinical Follow Up; High-Dimensional Data; Multivariate; Optimal Treatment Regimen

***Corresponding author:** Danhui Yi, School of Statistics, Renmin University of China, Beijing, China, Tel: +86 13701214116; E-mail: xueyi905@aliyun.com

Citation: Han F, Zhang L, Yi D, Zhang J, Han K (2025) Selection of Optimal Treatment Regimen for Individuals Using Multivariate Assessments During Clinical Follow-up. HSOA J Altern Complement Integr Med 11: 626.

Received: August 26, 2025; **Accepted:** September 09, 2025; **Published:** September 16, 2025

Background

In medical practice, doctors and patients will have to decide on the best treatment regimen to maximize treatment benefit. In some cases, it is a matter of "do or die". Crucial as it is, the treatment selection criteria suffer from "the curse of dimensionality", which means due to factors known and unknown, some patients have higher likelihood of cure after immediate surgery while others respond positively to less invasive treatments. To tackle the problem of uncertainty associated with treatment selection, statistic techniques such as causal inference studies are conventionally employed, despite the fact that causal effect is usually a constant rather than a function, hence a lack of predictive ability of covariates in evaluating individualized treatment selection of heterogeneous patients. In the literature, two approaches and their hybrids made it possible to derive an optimal treatment rule. The first approach targets at predicting the average result difference of biomarkers in two comparison groups or multiple comparison groups. Notably, the average result difference is not a constant but a biomarker-dependent function, and these biomarkers may be single, multiple or even high-dimensional. For the selection of an optimal treatment regimen based on a single biomarker, Ma and Zhou [1] proposed a method using Covariate-Specific Treatment Effect (CSTE) curves, which was followed up by Han et al., [2] based on the cases where the outcome variables are binary.

Guo et al., [3] extended their research to high-dimensional data which is cross-sectional rather than longitudinal, and failed to build statistical inference upon the situation when the data was obtained from actual treatments in multiple follow-ups and thus longitudinal. To advance their study, Gao [4] investigated the longitudinal data from repeated measurements of outcomes of a single biomarker by using the spline-based Generalised Estimating Equation (GEE) method to estimate the CSTE curve, and identified its asymptotic nature. Based on the asymptotic normality of CSTE curves, the confidence bands in the CSTE curves were constructed. The results of numerical simulation and case analysis show that the estimation of the CSTE curve and its confidence band proved the efficiency of the estimators.

In practical scenarios, longitudinal follow-up studies often involve multidimensional or high-dimensional covariates, attributable to the inclusion of multiple biomarkers or high-dimensional biomarkers. Based on this, an attempt was made to expend previous longitudinal single-biomarker studies to multiple biomarker ones, and corresponding CSTE curve estimation methods were presented. Regarding the parameter estimation of CSTE curves for high-dimensional

longitudinal data, the Smoothly Clipped Absolute Deviation (SCAD) penalty was employed, specifically, the penalised GEE (PGEE) method utilized by Wang et al., [5] to estimate the parameters.

Despite being a robust strategy, the CSTE curves cannot generate optimal treatment policy without the estimation of confidence bands. Song [6] determined optimal treatment regimens via the Selection Influence Curve (SIC) and employed the bootstrap method to obtain the confidence band. Ma and Zhou [1] used the Resampling (RS) technique to construct the confidence bands for the CSTE curve with continuous response variables. Moreover, Zhou et al., [7] proved the asymptotic normality of the regression spline and applied the Hotelling Tube method to construct the confidence bands. Subsequently, Loader et al., [8] and Krivobokovaa et al., [9] followed a similar approach in constructing the confidence bands of the estimators.

Han et al., [2] applied this method to construct the confidence band of the CSTE curve for cross-sectional data of single biomarker with binary outcome. Gao also applied it to construct confidence band of the CSTE curves with longitudinal single biomarkers, and Guo et al., constructed confidence bands of the CSTE curves using the Spline Back-Kernel (SKB) estimators. In this paper, we applied Hotelling Tube method to construct confidence band of the CSTE curves for longitudinal data of multiple biomarkers and high-dimensional biomarkers.

This paper is organized as follows:

In Section 2, we introduce the optimal treatment regimen selection method based on CSTE curves using multivariate and high-dimensional clinical follow-up data. Section 3 presents the simulation study, and a conclusion is given in Section 4.

Optimal Treatment Regimen Selection Based on CSTE Curves with Multivariate and High-Dimensional Clinical Follow-Up Data

Defining CSTE curve with multivariate clinical follow-up data

Notations are given here for the sake of convenience in the following discussions.

We consider a sample of N subjects ($i = 1,2,\cdots\cdots,N$), each of whom was observed for M times. Let $Y_{ij}(k)$ be the outcome variable of the j-th observation of the i-th individual receiving the k-th treatment regeimen, and the indicator variable W_{ijl} represents whether the i-th subject select the l-th treatment regimen in the j-th observation. A value of 1 indicates acceptance of the l-th treatment regimen, and 0, otherwise. $(W_{ij1}, W_{ij2}, \cdots\cdots, W_{ijd}) = W_{ij}^T$ indicates the d-th treatment regimen available for the j-th observation of the i-th subject. If there is only one treatment option available, each subject has a pair of potential outcomes in the j-th observation, denoted as $(Y_{ij}(0), Y_{ij}(1))$. If the i-th subject is in the control group, then $Y_{ij}(1)$ is an unobservable latent variable and $Y_{ij}(0)$ is an observable variable. If the i-th subject is in the treatment group, then the opposite is true. Namely, $Y_{ij} = W_{ij} Y_{ij}(1) + (1-W_{ij})$ is the measurable outcome variable of the i-th individual's j-th observation. X_{ij} represents that the covariate of the j-th measurement of the i-th subject is a vector, denoted as $X_{ij} = (X_{ij1}, X_{ij2}, \cdots\cdots, X_{ijP})$.

Underlying Hypothesis 1: Stable unit treatment value, which implies that subjects who receive treatments do not interfere with each other's potential outcome, and for any concerned subject, each treatment corresponds to only one potential outcome.

Underlying Hypothesis 2: The ignorability assumption, denoted as, $W \perp (Y(0), Y(1)) \mid X$, also known as the unconfoundedness assumption, was proposed by Rosenbaum and Rubin (1983), who observed that the treatment assignment is ignorable and does not depend on the potential outcome.

Single treatment programs: Based on these hypotheses, the outcome variables in a single binary treatment program, W, are:

$$Y_{ij} = W_{ij} Y_{ij}(1) + (1-W_{ij}) Y_{ij}(0)$$

Let $E(Y_{ij}) = \mu_{ij}$, hypothetically, g(.) is assumed to be a joint function with strict monotonicity, and there exists an inverse function denoted by $g^{-1}(.)$, $\mu_{ij} = g(W_{ij}^T \beta(x_{ij}) + h(x_{ij}))$.

The CSTE curve for a longitudinal univariate single-treatment regimen is assumed to be (Gao, 2020):

$$CSTE(x) = g^{-1}\big(E(Y|X = x, W = 1)\big) - g^{-1}\big(E(Y|X = x, W = 0)\big)$$

However, the above definition was based on single covariate, when there are multiple covariates, which is often the case in practice, we redefine longitudinal multivariate CSTE function as:

$$CSTE(X) = g^{-1}\big(E(Y|X = X, W = 1)\big) - g^{-1}\big(E(Y|X = X, W = 0)\big)$$

Here, X is a vector.

Multiple treatment programs

Gao (2020) assumed that the CSTE curves for the k-th treatment option of longitudinal univariate are:

$$CSTE(x) = g^{-1}\big(E(Y|X = x, W = I_K)\big) - g^{-1}\big(E(Y|X = x, W = 0)\big)$$

In the above equation, the I_K is a vector with the k-th element being 1 and the remaining elements being 0. The monotonicity of $g^{-1}(.)$ can be obtained from the monotonicity of $g(.)$. However, the above definition is only based on one covariant and become unapplicable in case of multiple covariates; Accordingly, we provide a definition of multivariate variables. Assuming the CSTE curves for the k-th treatment regimen of the longitudinal multivariate are:

$$CSTE_K(X) = g^{-1}\big(E(Y|X = X, W = I_K)\big) - g^{-1}\big(E(Y|X = X, W = 0)\big)$$

Here, X is a vector.

CSTE curve estimation based on multivariate clinical follow-up data

Estimation of CSTE function: B-spline function approximation

Applying the B-spline function to approximate our estimation of the CSTE function leads to the following assumptions:

Assumption 1: $X_{ij} = (X_{ij1}, X_{ij2}, \cdots\cdots, X_{ijP})$, in which the domain of each component is a compact set.

Assumption 2: The distance between r nodes of B-spline basis function with order m+1 on the interval $[a, b]$ is equal.

Assumption 3: The CSTE curve $CSTE_K(X)$ is $m+1$ order continuous in the defined domain.

Assumption 4: The presence of a $C > 0$ maker = $CN^{1/(2m+1)}$.

Suppose the biological indicators of the subjects are denoted as X, then the value for each component is respectively $[a_1,b_1]\cdots\cdots[a_p b_p]$:

$$B_{(1)}(X_{ij1})_r^{m+1} = \left(B_{(1)1}(X_{ij1}), B_{(1)2}(X_{ij1}), \cdots\cdots, B_{(1)r+m+1}(X_{ij1})\right)^T \cdots\cdots$$

where $B_{(P)}(X_{ijP})_r^{m+1} = \left(B_{(P)1}(X_{ijP}), B_{(P)2}(X_{ijP}), \cdots\cdots, B_{(P)r+m+1}(X_{ijP})\right)^T$ is a vector of B-spline basis functions of order m+1 with r nodes on each interval. Next, we use this vector to approximate:

$$\beta_k(x) = \beta_{(1)k}(x_1) + \cdots\cdots + \beta_{(p)k}(x_p) \approx \sum_{i=1}^{r+m+1} v_{(1)ik} B_{(1)i}(x_1) + \cdots\cdots + \sum_{i=1}^{r+m+1} v_{(p)ik} B_{(p)i}(x_p) = V_{(1)ik} B_{(1)i}(x_1)^T + \cdots V_{(p)ik} B_{(p)i}(x_p)^T$$

$$h(X) = h_{(1)}(X_1) + \cdots\cdots + h_{(p)}(X_p)$$

$$\approx \sum_{i-1}^{r+m+1} v_{(1)ik+1} B_{(1)i}(X_1) + \cdots\cdots + \sum_{i-1}^{r+m+1} v_{(p)ik+1} B_{(p)i}(X_p) = V_{(1)k+1} B_{(1)}(X_1)^T + \cdots\cdots + V_{(p)k+1} B_{(p)}(X_p)^T$$

Here, $V_{(1)K} = \left(v_{(1)1k}, v_{(1)2k}, \cdots\cdots, v_{(1)r+m+1k}\right)^T$. And others abide by the same token.

The original model can be approximated as follows:

$$\mu_{ij} = g(W_K^T \beta_k(X) + h(X))$$

$$= g(W_K^T (V_{(1)k} B_{(1)}(X_1)^T + \cdots\cdots V_{(p)k} B_{(p)}(X_p)^T) + V_{(1)k+1} B_{(1)}(X_1)^T) + \cdots\cdots V_{(p)k+1} B_{(p)}(X_p)^T)$$

$$CSTE_K(X) = g^{-1}\left(E(Y|X = X, W = I_K)\right) - g^{-1}\left(E(Y|X = X, W = 0)\right)$$

$$= \beta_K(X) = V_{(1)K} B_{(1)}(X_1)^T + \cdots\cdots + V_{(P)K} B_{(P)}(X_P)^T$$

Also denoted as:

$$\pi(X)^T = \left(W_{ij1}(B_{(1)}(X_1)^T + \cdots\cdots + B_{(P)}(X_P)^T), W_{ij2}(B_{(1)}(X_1)^T + \cdots\cdots + B_{(P)}(X_P)^T)^T, \cdots\cdots, W_{ijK}(B_{(1)}(X_1)^T + \cdots\cdots + B_{(P)}(X_P)^T), (B_{(1)}(X_1)^T + \cdots\cdots + B_{(P)}(X_P)^T)\right)$$

and $V = (V_K, V_{K+1})^T$,

Thus, $E(Y_{ij}) = \mu_{ij} = g(W_K^T \beta_K(X) + h(X)) = g(\pi(X)^T V)$.

What is worth noticing is that the coupling function $g(.)$ has strict monotonicity and there exists an inverse function for that $g^{-1}(.)$, both of which have the same monotonicity.

Model parameter estimation

Denoting $var(Y_{ij}) = g'\left(W_{ij}^T \beta(X_{ij}) + h(X_{ij})\right)\varphi = g'(\pi(X)^T V)\varphi$, where Y_{ij} covariance matrix is assumed to be a function of marginal mean and parameter α:

$$Cov(Y_{iu}, Y_{iv}) = c(\mu_{iu}, \mu_{iv}; \alpha) = c(g(\pi(X_{iu})^T V), g(\pi(X_{iv})^T V); \alpha) = R(\alpha)$$

where c(.) is a known function and u and v are measurement No. u and No. v respectively.

Assumption 5 Given V and φ, the estimation of the parameter α in the work correlation matrix $\hat{\alpha}$ has $\sqrt{N}$ consistency.

Assumption 6 Given V and φ, the estimation of the $\hat{\varphi}$ has $\sqrt{N}$ consistency.

Assumption 7 $$\frac{\partial \hat{\alpha}(V, \varphi)}{\partial \varphi} \leq H(Y, V) + O(1)$$

The GEE method was applied for parameter estimation, while we introduced several variables.

Let: $A_i = diag\{g'(\pi(X_{i1})^T V), g'(\pi(X_{i2})^T V), \cdots\cdots, g'(\pi(X_{iM})^T V)\}$

The estimated equation is as follows:

$$S(V; \alpha, \varphi) = \sum_{i=1}^{N} \left(\frac{\partial g(\pi(X_i)^T V)}{\partial V}\right)^T \left(A_i^{\frac{1}{2}} R(\alpha) A_i^{\frac{1}{2}}\right)^{-1} \left(Y_i - g(\pi(X_i)^T V)\right) = O_p$$

where the estimation of the given work-related matrix in practice is $\widehat{R(\alpha)}$ and the estimation of variance is $\hat{A}_i$. In this case, the solution for the estimated equation can be obtained by iterating using Fisher's scoring method for parameter estimation.

CSTE estimation of asymptotic normality

The asymptotic normality of the estimation of longitudinal univariate CSTE curves was proven by Gao (2020). On this basis, under the above assumptions in this paper, it was not difficult to obtain the following properties of $CSTE_K$:

$\widehat{CSTE}_K(X)$ uniform convergence in $CSTE_K(X)$ and

$\widehat{CSTE}_K(X) \to N(CSTE_K(X), B(X)^T e_K \sum e_K^T B(X))$,

$\sum = \{\sum_{i=1}^N D_i^T V_i^{-1} D_i\}^{-1} \sum_{i=1}^N D_i^T V_i^{-1}(Y_i - g(\pi(X_i)^T V))(Y_i - g(\pi(X_i)^T V))^T V_i^{-1} D_i^T \{\sum_{i=1}^N D_i^T V_i^{-1} D_i\}^{-1}$,

Among them,

$D_i = \frac{\partial g(\pi(X_i)^T V)}{\partial V}$.

Next, we apply Hotelling Tube method to construct the confidence band for the simultaneous promotion of CSTE curves over the entire domain of definition.

Construction of confidence bands for the generalization of CSTE estimation

In fact, if there are two covariates, the geometric meaning of the CSTE function is no longer a curve but a surface. For the convenience of research and to visually demonstrate how to select the optimal treatment regimen as different covariates change, when we examine a certain covariate, we might as well take other covariates as a certain constant in the definition domain. For example, if there are two covariates, one is discrete as X_1 with values of 1 and 2, and the other is continuous as X_2. We set the value of X_1 as 1, *CSTE (X) = CSTE (X_1= 1,X_2)* remains a curve, and the treatment regimen is selected based on the change in the value of the biological indicators X_2. Similarly, when X_1 has a value of 2, *CSTE (X) = CSTE (X_1 = 2,X_2)* remains a curve, and the treatment regimen is selected based on the change in the value of the biological indicator X_2.

At the same time, as is well-known, continuous variables can be discretised, especially in actual treatment. For the normal range of some biological indicators, we may as well mark the value as 1, and mark the non-normal range as 2. Therefore, the continuous variable X_2 can be discretized according to the practical significance of its value. Assuming that the values are 1 and 2, respectively, we get *CSTE (X) = CSTE (X_1,X_2=1)* and *CSTE (X) = CSTE (X_1,X_2=2)* separately, and make the choice of treatment according to the change of the value of the biological indicator X_1. For multiple covariates, the method is analogous. In the following discussion, we assume that the variable which is not a constant in the multi-dimensional CSTE curve estimation is X_1, with the value of and the domain of definition is [a,b].

Based on the assumptions of this paper and a sufficiently large sample size, it was not difficult to obtain the following conclusions by applying Hotelling Tube method:

$$P\left(\sup_{a \leq x \leq b} \frac{|\hat{\beta}_K(X) - \beta_K(X)|}{\sqrt{Var(\hat{\beta}_K(X))}} \leq c_\alpha\right) \geq 1 - \alpha$$

where $0<\alpha<1, c_\alpha$, satisfies:

$$\alpha=\frac{|\gamma|}{\pi}e^{-\frac{c_\alpha^2}{2}}+2(1-\varphi(c_\alpha))$$

$$\gamma=\int_{[a,b]}\left\|\frac{d}{dX}\left(\frac{\Sigma^{1/2}e_K^TB(X)}{|\Sigma^{1/2}e_K^TB(X)|}\right)\right\|dX$$

$$B(X)=B_{(1)}(X_1=x)^T+B_{(2)}(X_2=a_1)^T+.....+B_{(p)}(X_p=a_{p-1})^T$$

Where by the confidence bands for the CSTE curve are:

$$\left[\beta_k(X)-c_\alpha\sqrt{Var(\beta_k(X))},\beta_k(X)+c_\alpha\sqrt{Var(\beta_k(X))}\right].$$

CSTE curve estimation for high-dimensional clinical follow-up data

In this section, we examined the CSTE curve estimation in the case of high-dimensional data. The CSTE estimation is still in accordance with the above discussed B-spline function approximation. In addition, we resolved the high-dimensional problem using the SCAD penalty in the parameter estimation, that is, we applied the PGEE method in the parameter estimation:

$$\mathrm{U}(V)=\mathrm{S}(V)-\mathrm{p}_\lambda^{'}(|V|)\mathrm{sign}(V)$$

$$=\sum_{i=1}^{N}\left(\frac{\partial g(\pi(X_i)^TV}{\partial V}\right)^T\left(A_i^{\frac{1}{2}}R(\alpha)A_i^{\frac{1}{2}}\right)^{-1}(Y_i-g(\pi(X_i)^TV))-P_\lambda(|V|)sign(V)=0$$

where

$$p_\lambda^{'}(\theta)=\lambda\left\{I(\theta\leq\lambda)+\frac{(a\lambda-\theta)_+}{(a-1)\lambda}I(\theta>\lambda)\right\}.$$

Simulation Study

In this section, we investigate the feasibility and large sample property of our proposed method via simulated datasets.

Simulation study of the multivariate clinical follow-up data

300 datasets were randomly generated in order to examine the three covariates, then, the specific covariates and the treatment program indicator variables were generated as follows:

$$X_1\sim Uniform(0,0.8),X_2\sim Uniform(0,0.9),X_3\sim N(0.5,0.16),T\sim B(0.6)$$

where the outcome variables are: $Y_{ij}=g(X)T+h(X)+\varepsilon_{ij}$

$g(X)=0.1(X_1)^2+0.2(X_2)^4+0.3(X_3)^4$and $\mathrm{h}(X)=0.3(X_1)^4+0.1(X_2)^6+0.2(X_3)^2$

$\varepsilon_{ij}\sim N(0,\Sigma)$,

where $\sum$ for the structural correlation matrix is as follows:

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	1.0000000	0.667000	0.444889	0.296741	0.1979262
[2,]	0.6670000	1.000000	0.667000	0.444889	0.2967410
[3,]	0.4448890	0.667000	1.000000	0.667000	0.4448890
[4,]	0.2967410	0.444889	0.667000	1.000000	0.6670000
[5,]	0.1979262	0.296741	0.444889	0.667000	1.0000000

The evaluation of the goodness-of-fit of the GEE adopts the QIC and QICC criteria based on the quasi-likelihood function, as the likelihood function is no longer applicable to the GEE. Here, QIC was used to select the correlation matrix under a given model, while QICC was used to select the model under a given correlation matrix. For both indices, the smaller the value, the better the corresponding structure or model.

Factoring in the estimation method of the correlation matrix to three settings, namely, independent structure, unspecified structure and autocorrelation structure when generating simulation data. 300 experiments were conducted for each setting to observe the estimation of the parameters and the CSTE curves. In the experiments, the smoothing parameters, *m* and *r*, were determined by cross-validation, and the QICs estimated by different combinations of *m* and *r* were compared to select the smoothing parameter combination that minimises the QIC. table 1 shows the mean square error (MSE) and mean absolute error (MAE) in the CSTE curve estimation.

Sample size	Associated array types	MSE	MAE	CP	Length
500	Exchangeable	0.03231677	0.12751007	0.99031200	0.91701638
	Autocorrelation structure	0.01602450	0.07993835	0.98682800	0.53886455
	Independence	0.1071688	0.2419963	0.9958240	1.7860859
	Without specifying the structure	0.1052969	0.2000057	0.9866958	6.9972548
1000	Exchangeable	0.02048433	0.09591628	0.99368200	0.67388231
	Autocorrelation structure	0.01221517	0.06345883	0.98803200	0.39740128
	Independence	0.05670803	0.17538938	0.99692600	1.31077785
	Without specifying the structure	0.06650566	0.14969630	0.98200419	3.31338835
2000	Exchangeable	0.01458137	0.07549826	0.98989300	0.49166377
	Autocorrelation structure	0.01042161	0.05396849	0.97826800	0.29541488
	Independence	0.03272665	0.12897410	0.99672700	0.95677027
	Without specifying the structure	0.03545137	0.11072930	0.98594726	1.67024123

Table 1: Estimates of the CSTE function for Simulation 1: Multivariate clinical follow-up and confidence bands.

Abbreviations: CP, coverage probabilities; MAE, mean absolute error; MSE, mean square error.

It can be seen from table 1 that as the sample size increases, the MSE and MAE decrease, the length of the confidence interval decreases, and the coverage rate increases. Whether the correlation matrix type is correctly specified or not, both MSE and MAE are relatively small, and they are the smallest when the correlation matrix type is correctly specified.

In figure 1, we discretised variables X_1 and X_3, If both ≤0.5, they take the value of 1, otherwise, 0. We consider the case where X_1 takes the value of 1 and X_3 takes the value of 0. By examining the CSTE

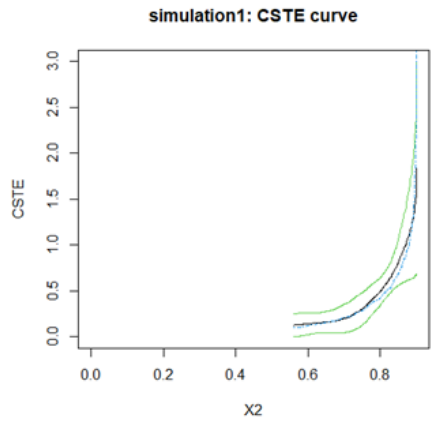

Figure 1: Simulation 1: CSTE curve.

curves with the change of X_2, it can be seen that the curves are all >0 within the range of the X_2 values, so the treatment regimen should be selected.

Simulation study of high-dimensional clinical follow-up data

300 datasets were randomly generated in order to examine the three covariates, then, the specific covariates and the treatment program indicator variables were generated as follows:

$$X_1 \sim U(0,0.1), X_2 \sim U(0,0.5), X_3 \sim U(0,0.1), X_4 \sim U(0,0.2)$$
$$X_5 \sim U(0,0.3), X_6 \sim U(0,0.4), X_7 \sim U(0,0.5), X_8 \sim U(0,0.6)$$
$$X_9 \sim U(0,0.7), X_{10} \sim U(0,0.8), X_{11} \sim U(0,0.15), X_{12} \sim U(0,0.8), T \sim B(0.6);$$

The outcome variables are: $Y_{ij} = g(X)T + h(X) + \varepsilon_{ij}$

$$g(X) = 0.1X_1 + 0.2X_2 + 0.3X_3 + 0.2X_4 + 0.1X_5 + 0.3X_6 + 0.2X_7 + 0.1X_8 + 0.1X_9 + 0.3X_{10} + 0.2X_{11} + 0.1X_{12}$$

$$h(X) = 0.3X_1 + 0.1X_2 + 0.2X_3 + 0.1X_4 + 0.3X_5 + 0.2X_6 + 0.1X_7 + 0.3X_8 + 0.3X_9 + 0.2X_{10} + 0.1X_{11} + 0.3X_{12}$$

$$\varepsilon_{ij} \sim N(0, \Sigma)$$

where ∑ for the structural correlation matrix is as follows:

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	1.0000000	0.667000	0.444889	0.296741	0.1979262
[2,]	0.6670000	1.000000	0.667000	0.444889	0.2967410
[3,]	0.4448890	0.667000	1.000000	0.667000	0.4448890
[4,]	0.2967410	0.444889	0.667000	1.000000	0.6670000
[5,]	0.1979262	0.296741	0.444889	0.667000	1.0000000

As can be seen from table 2, MSE and MAE are decreasing as the sample size increases, and both MSE and MAE are smaller with or without the correctly specified correlation matrix type, with the smallest in the correctly specified case.

Sample size	Associated array types	MSE	MAE
500	Exchangeable	0.04512008	0.14262900
	Autocorrelation structure	0.1081171	0.2149774
	Independence	0.02697552	0.12490880
	Without specifying the structure	0.1081171	0.2149774
1000	Exchangeable	0.02715838	0.10864697
	Autocorrelation structure	0.07464237	0.1698040
	Independence	0.02402516	0.11848331
	Without specifying the structure	0.07464237	0.1698040
2000	Exchangeable	0.01604547	0.08131358
	Autocorrelation structure	0.05775575	0.1392986
	Independence	0.01240049	0.08403477
	Without specifying the structure	0.05775575	0.1392986

Table 2: Estimates of the CSTE function for Simulation 2: high-dimensional clinical follow-up.

Abbreviations: MAE, mean absolute error; MSE, mean square error.

Conclusion

In this paper, we proposed the estimation method for multivariate CSTE curves in clinical follow-up by taking account of the definition of longitudinal univariate CSTE curves and using GEE and spline function estimation with an additive model. The confidential band we constructed may grant a higher percentage of success in therapeutical decisions. As for high-dimensional CSTE curves, PGEE is exploited to model the selection impact curve. Simulation studies show that as the sample size increases, the estimated values approximate the true values, the length of the confidence intervals decreases while the coverage rate increases, indicating good large-sample properties.

Declarations

Ethics approval and consent to participate

This research complied with national policies and regulations.

Consent for publication

We declare that there are no conflicts of interest regarding the publication of this manuscript.

Availability of data and materials

The data that support the fundings of this research are available from the corresponding author upon reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was supported by the National Social Science Foundation of China (Grants No. 18FTJ003), the Academic Research Projects of Beijing Union University (No. ZKZD202306), and the Research Project on Water Price Pricing Model of Shanxi Large Water Network Project.

Author's Contribution

Except for the corresponding author, the other authors are ranked according to their contribution.(Corresponding Author:Yi Danhui,-First Author:Han Feng,Second Author:Zhang Lin lin,Co - Fourth Author:Zhang Jingsheng and Han Kaishan).

Acknowledgement

We would like to express our deepest gratitude to our families. They have always been our strongest supporters, and their encouragement and companionship are the driving force of our academic pursuit. In addition, we are lucky to have each other in the research team to work closely and learn from each other.

References

1. Ma Y, Zhou XH (2017) Treatment selection in a randomized clinical trial via covariate-specific treatment effect curves. Statistical Methods in Medical Research 26: 124-141.
2. Han K, Zhou X-H, Liu B (2017) CSTE Curve for Selection the Optimal Treatment When Outcome Is Binary. Scientia Sinica Mathematica 47: 497-514.
3. Wenchuan G, Hua ZX, Shujie M (2021) Estimation of Optimal Individualized Treatment Rules Using a Covariate-Specific Treatment Effect Curve with High-Dimensional Covariates. Journal of the American Statistical Association 116: 309-321.
4. Yanxia G (2020) Using CSTE curve to select the optimal scheme under longitudinal data. East China Normal University, China.
5. Lan W, Jianhui Z, Annie Q (2012) Penalized generalized estimating equations for high-dimensional longitudinal data analysis. Biometrics 68: 353-360.
6. Song X, Pepe MS (2004) Evaluating markers for selecting a patient's treatment. Biometrics 60: 874-83.
7. Zhou S, Shen X, Wolfe DA (1998) Local asymptotics for regression splines and confidence regions. The annals of statistics 26: 1760-1782.
8. Loader C, Sun J (2012) Robustness of Tube Formula Based Confidence Bands. Journal of Computational and Graphical Statistics 6: 242-250.
9. Krivobokova T, Kneib T, Claeskens G (2010) Simultaneous Confidence Bands for Penalized Spline Estimators. Journal of the American Statistical Association 105: 852-863.

Advances In Industrial Biotechnology | ISSN: 2639-5665

Advances In Microbiology Research | ISSN: 2689-694X

Archives Of Surgery And Surgical Education | ISSN: 2689-3126

Archives Of Urology

Archives Of Zoological Studies | ISSN: 2640-7779

Current Trends Medical And Biological Engineering

International Journal Of Case Reports And Therapeutic Studies | ISSN: 2689-310X

Journal Of Addiction & Addictive Disorders | ISSN: 2578-7276

Journal Of Agronomy & Agricultural Science | ISSN: 2689-8292

Journal Of AIDS Clinical Research & STDs | ISSN: 2572-7370

Journal Of Alcoholism Drug Abuse & Substance Dependence | ISSN: 2572-9594

Journal Of Allergy Disorders & Therapy | ISSN: 2470-749X

Journal Of Alternative Complementary & Integrative Medicine | ISSN: 2470-7562

Journal Of Alzheimers & Neurodegenerative Diseases | ISSN: 2572-9608

Journal Of Anesthesia & Clinical Care | ISSN: 2378-8879

Journal Of Angiology & Vascular Surgery | ISSN: 2572-7397

Journal Of Animal Research & Veterinary Science | ISSN: 2639-3751

Journal Of Aquaculture & Fisheries | ISSN: 2576-5523

Journal Of Atmospheric & Earth Sciences | ISSN: 2689-8780

Journal Of Biotech Research & Biochemistry

Journal Of Brain & Neuroscience Research

Journal Of Cancer Biology & Treatment | ISSN: 2470-7546

Journal Of Cardiology Study & Research | ISSN: 2640-768X

Journal Of Cell Biology & Cell Metabolism | ISSN: 2381-1943

Journal Of Clinical Dermatology & Therapy | ISSN: 2378-8771

Journal Of Clinical Immunology & Immunotherapy | ISSN: 2378-8844

Journal Of Clinical Studies & Medical Case Reports | ISSN: 2378-8801

Journal Of Community Medicine & Public Health Care | ISSN: 2381-1978

Journal Of Cytology & Tissue Biology | ISSN: 2378-9107

Journal Of Dairy Research & Technology | ISSN: 2688-9315

Journal Of Dentistry Oral Health & Cosmesis | ISSN: 2473-6783

Journal Of Diabetes & Metabolic Disorders | ISSN: 2381-201X

Journal Of Emergency Medicine Trauma & Surgical Care | ISSN: 2378-8798

Journal Of Environmental Science Current Research | ISSN: 2643-5020

Journal Of Food Science & Nutrition | ISSN: 2470-1076

Journal Of Forensic Legal & Investigative Sciences | ISSN: 2473-733X

Journal Of Gastroenterology & Hepatology Research | ISSN: 2574-2566

Journal Of Genetics & Genomic Sciences | ISSN: 2574-2485

Journal Of Gerontology & Geriatric Medicine | ISSN: 2381-8662

Journal Of Hematology Blood Transfusion & Disorders | ISSN: 2572-2999

Journal Of Hospice & Palliative Medical Care

Journal Of Human Endocrinology | ISSN: 2572-9640

Journal Of Infectious & Non Infectious Diseases | ISSN: 2381-8654

Journal Of Internal Medicine & Primary Healthcare | ISSN: 2574-2493

Journal Of Light & Laser Current Trends

Journal Of Medicine Study & Research | ISSN: 2639-5657

Journal Of Modern Chemical Sciences

Journal Of Nanotechnology Nanomedicine & Nanobiotechnology | ISSN: 2381-2044

Journal Of Neonatology & Clinical Pediatrics | ISSN: 2378-878X

Journal Of Nephrology & Renal Therapy | ISSN: 2473-7313

Journal Of Non Invasive Vascular Investigation | ISSN: 2572-7400

Journal Of Nuclear Medicine Radiology & Radiation Therapy | ISSN: 2572-7419

Journal Of Obesity & Weight Loss | ISSN: 2473-7372

Journal Of Ophthalmology & Clinical Research | ISSN: 2378-8887

Journal Of Orthopedic Research & Physiotherapy | ISSN: 2381-2052

Journal Of Otolaryngology Head & Neck Surgery | ISSN: 2573-010X

Journal Of Pathology Clinical & Medical Research

Journal Of Pharmacology Pharmaceutics & Pharmacovigilance | ISSN: 2639-5649

Journal Of Physical Medicine Rehabilitation & Disabilities | ISSN: 2381-8670

Journal Of Plant Science Current Research | ISSN: 2639-3743

Journal Of Practical & Professional Nursing | ISSN: 2639-5681

Journal Of Protein Research & Bioinformatics

Journal Of Psychiatry Depression & Anxiety | ISSN: 2573-0150

Journal Of Pulmonary Medicine & Respiratory Research | ISSN: 2573-0177

Journal Of Reproductive Medicine Gynaecology & Obstetrics | ISSN: 2574-2574

Journal Of Stem Cells Research Development & Therapy | ISSN: 2381-2060

Journal Of Surgery Current Trends & Innovations | ISSN: 2578-7284

Journal Of Toxicology Current Research | ISSN: 2639-3735

Journal Of Translational Science And Research

Journal Of Vaccines Research & Vaccination | ISSN: 2573-0193

Journal Of Virology & Antivirals

Sports Medicine And Injury Care Journal | ISSN: 2689-8829

Trends In Anatomy & Physiology | ISSN: 2640-7752