Web-based Supplementary Materials for "Improving efficiency of inferences on treatment effect in two-arm randomized trials using auxiliary covariates"

Min Zhang*, **Anastasios A. Tsiatis, and Marie Davidian**

Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695-8203, U.S.A.

**email:* mzhang4@stat.ncsu.edu

Web Appendix A: Demonstration of the Existence of Joint Densities Satisfying the Semiparametric Model

At the end of Section 2 of the main paper, the semiparametric model framework within which we derive the proposed methods is stated, and is repeated here for convenience.

The data from a clinical trial are denoted by (Y_i, X_i, Z_i), $i = 1, \ldots, n$, assumed iid across i, where Y denotes the response of interest; X is a vector of baseline auxiliary covariates; and $Z = 1, \ldots, k$ depending on to which of k possible treatment groups subject i was randomized, with randomization probabilities $P(Z = g) = \pi_g$, $g = 1, \ldots, k$, $\sum_{g=1}^{k} \pi_g = 1$. Randomization guarantees that $Z \perp\!\!\!\perp X$. We assume that interest focuses on a parameter β involved in characterizing treatment comparisons, defined in the context of a model for the conditional density of Y given Z, $p_{Y|Z}(y|z; \theta, \eta)$, where $\theta = (\beta^T, \gamma^T)^T$. Here, γ represents a finite-dimensional vector of possible additional parameters, and η is a finite- or infinite-dimensional nuisance parameter required to describe fully the class of models under consideration. Examples of such models are discussed in Section 2 of the main paper.

The semiparametric model introduced in Section 2 of the main paper consists of all joint densities $p_{Y,X,Z}(y, x, z\,; \theta, \eta, \psi, \pi) = p_{Y,X|Z}(y, x \,|\, z; \theta, \eta, \psi) p_Z(z; \pi)$ such that the conditional densities for (Y, X) given Z, denoted $p_{Y,X|Z}(y, x \,|\, z; \theta, \eta, \psi)$, satisfy

$$\text{(i)} \qquad \int p_{Y,X|Z}(y, x \,|\, z; \theta, \eta, \psi)\, dx = p_{Y|Z}(y|z; \theta, \eta), \qquad \text{(A.1)}$$

$$\text{(ii)} \qquad \int p_{Y,X|Z}(y, x \,|\, z; \theta, \eta, \psi)\, dy = p_X(x), \qquad \text{(A.2)}$$

where $p_X(x)$ refers to any arbitrary marginal density for the auxiliary covariates, and (ii) follows because $Z \perp\!\!\!\perp X$. The additional nuisance parameters ψ and η together are used to specify conditional densities satisfying (i) and (ii).

We now demonstrate that joint distributions for (Y, X, Z) satisfying conditions (i) and (ii) in (A.1) and (A.2) may be constructed. For simplicity, we consider scalar Y and a two-armed trial, $k = 2$; extensions to vector Y and arbitrary $k \geq 2$ are straightforward. Begin with a given marginal density $p_X(x)$ for X and the conditional density $p_{Y|Z}(y|z;\theta,\eta)$ of interest. A joint distribution for (Y, X, Z) may then be developed through the following steps:

1. Generate X from $p_X(x)$.

2. Generate W_0 and W_1 from any arbitrary conditional densities $p_{W_0|X}(w_0 \,|\, x)$ and $p_{W_1|X}(w_1 \,|\, x)$, where, clearly, W_0 and W_1 have marginal densities

$$p_{W_k}(w_k) = \int p_{W_k|X}(w_k \,|\, x) p_X(x)\, dx, \quad k = 1, 2.$$

   A transformation of these variables is used in step 4 below to derive the response variable Y.

3. Generate a random Bernoulli Z random variable (taking values 1 and 2) independently of X, W_0, and W_1, with "success" probability $P(Z = 2) = \pi_2$. (Here, $P(Z = 1) = \pi_1 = 1 - \pi_2$.)

4. Let $F_{W_k}(u) = P(W_k \leq u)$, $k = 1, 2$, be the cumulative distribution functions (cdfs) for W_k, $k = 1, 2$, and write $F_{Y|Z=k}(u;\theta,\eta) = P(Y \leq u \,|\, Z = k;\theta,\eta)$, $k = 1, 2$, the cdfs corresponding to $p_{Y|Z}(y|z;\theta,\eta)$. Generate Y as

$$Y = I(Z = 1) F^{-1}_{Y|Z=1}\{F_{W_1}(W_1);\theta,\eta\} + I(Z = 2) F^{-1}_{Y|Z=2}\{F_{W_2}(W_2);\theta,\eta\}.$$

This construction guarantees that $Z \perp\!\!\!\perp X$, that the conditional distribution of Y given Z has the required density $p_{Y|Z}(y|z;\theta,\eta)$, and allows for flexible relationships for (Y, X) given Z. The derivation may be generalized straightforwardly to vector Y, as in the case of longitudinal response.

Web Appendix B: Derivation of Estimating Functions for Treatment Effect

We consider the semiparametric framework given at end of Section of 2 in the main paper, restated in Web Appendix A, and apply the principles of semiparametric theory to derive equation (12) of the main paper.

Before we present the detailed argument, we summarize the general approach. Under the semiparametric theory perspective, one views estimating functions as elements of the Hilbert space $\mathcal{H}$ consisting of all functions $h(Y, X, Z)$ such that $E\{h(Y, X, Z)\} = 0$ and $E\{h(Y, X, Z)^T h(Y, X, Z)\} < \infty$ (e.g., Tsiatis, 2006, Chapter 2). The advantage of considering estimating functions as elements of $\mathcal{H}$ is that geometric principles may be used to derive the form of all estimating functions and to assess the relative efficiencies of the estimators corresponding to them. The derivation makes use of the critical result that, under suitable regularity conditions, all estimating functions for estimators of finite-dimensional parameters of interest in semiparametric models are orthogonal to the so-called nuisance tangent space, a certain linear subspace of $\mathcal{H}$ (see Tsiatis, 2006, Chapter 4, for general discussion). Thus, the argument involves characterizing the nuisance tangent space and its orthogonal complement, in which estimating functions for a particular problem lie. In our case, then, the key to deriving semiparametric estimators for the parameter θ in our framework is to describe the nuisance tangent space and find the form of elements in its orthogonal complement, which will be of the form of estimating functions used to construct estimating equations for θ based on (Y, X, Z).

We remark in response to a query by a reviewer that the argument given in the rest of this appendix is not simply a special case of a general theory. Rather, it results from applying the semiparametric theory perspective above to this problem. The argument subsumes and represents a significant advance beyond that given by Leon et al. (2003) and in Appendix A.2 of Davidian, Tsiatis, and Leon (2005) for the particular case of estimating the difference of $k = 2$ treatment means, β_2, defined in (1) of the main paper.

We present the argument for the particular case scalar Y and infinite dimensional η; similar developments are possible in the cases of multivariate Y and null η. Formally, the nuisance tangent space we seek is defined as the mean-square closure of parametric submodel nuisance tangent spaces. A parametric submodel is a finite-dimensional parametric model that

(a) Is contained in the semiparametric model and

(b) Contains the truth; i.e., the distribution that generates the data.

The parametric submodel nuisance tangent space is the space spanned by the nuisance score vector of the parametric submodel. Here, we denote such a parametric submodel as

$$p_{Y,X|Z}(y, x|z; \theta, \xi_\eta, \xi_\psi),$$

satisfying conditions analogous to (A.1) and (A.2), i.e.,

$$\text{(i)} \qquad \int p_{Y,X|Z}(y, x|z; \theta, \xi_\eta, \xi_\psi) dx = p_{Y|Z}(y|z; \theta, \xi_\eta) \tag{B.1}$$

$$\text{(ii)} \qquad \int p_{Y,X|Z}(y, x|z; \theta, \xi_\eta, \xi_\psi) dy = p_X(x; \theta, \xi_\eta, \xi_\psi), \tag{B.2}$$

where $p_{Y,X|Z}(y, x|z; \theta_0, \xi_{\eta 0}, \xi_{\psi 0})) = p_{0Y,X|Z}(y, x|z)$, the true density of Y, X given Z (i.e., that generating the data); ξ_η is a finite-dimensional parameter defined so that $p_{Y|Z}(y|z; \theta, \xi_\eta)$ is a parametric submodel for the semiparametric model $p_{Y|Z}(y|z; \theta, \eta)$; and ξ_ψ is an additional finite dimensional parameter describing the joint distribution of Y and X given Z such that (i) and (ii) in (B.1) and (B.2) are satisfied.

The parametric submodel nuisance tangent space is the space spanned by the nuisance score vector $\{S_{\xi_\eta}^T(Y,X,Z), S_{\xi_\psi}^T(Y,X,Z)\}^T$, where

$$S_{\xi_\eta}(y,x,z) = \frac{\partial}{\partial \xi_\eta} \log\{p_{Y,X|Z}(y,x|z;\theta_0,\xi_{\eta 0},\xi_{\psi 0})\},$$

and similarly for $S_{\xi_\psi}(Y,X,Z)$. Thus, the parametric submodel nuisance tangent space is made up of elements $\{B_1 S_{\xi_\eta}(Y,X,Z) + B_2 S_{\xi_\psi}(Y,X,Z)\}$, where B_1 and B_2 are conformable matrices.

Denote the nuisance tangent space for the semiparametric model $p_{Y|Z}(y|z;\theta,\eta)$ by Λ_η. By definition, any element spanned by the parametric-submodel nuisance score vector $S_{\xi_\eta}^*(Y,Z)$, where $S_{\xi_\eta}^*(y,z) = \dfrac{\partial}{\partial \xi_\eta} \log\{p_{Y|Z}(y|z;\theta_0,\xi_{\eta 0}\}$, is an element of Λ_η.

From (B.1), we get

$$\log\left\{\int p_{Y,X|Z}(y,x|z;\theta,\xi_\eta,\xi_\psi)dx\right\} = \log\left\{p_{Y|Z}(y|z;\theta,\xi_\eta)\right\} \tag{B.3}$$

so that

$$\frac{\partial}{\partial \xi_\eta}\log\left\{\int p_{Y,X|Z}(y,x|z;\theta_0,\xi_{\eta 0},\xi_{\psi 0})dx\right\} = \frac{\partial}{\partial \xi_\eta}\log\left\{p_{Y|Z}(y|z;\theta_0,\xi_{\eta 0})\right\}$$

and thus

$$B_1\frac{\partial}{\partial \xi_\eta}\log\left\{\int p_{Y,X|Z}(y,x|z;\theta_0,\xi_{\eta 0},\xi_{\psi 0})dx\right\} = B_1\frac{\partial}{\partial \xi_\eta}\log\left\{p_{Y|Z}(y|z;\theta_0,\xi_{\eta 0})\right\} = B_1 S_{\xi_\eta}^*(y,z).$$

Under regularity conditions, the left-hand side of the above equation is equal to

$$
\begin{aligned}
& B_1 \frac{\int \frac{\partial}{\partial \xi_\eta} p_{Y,X|Z}(y,x|z;\theta_0,\xi_{\eta 0},\xi_{\psi 0})dx}{\int p_{Y,X|Z}(y,x|z;\theta_0,\xi_{\eta 0},\xi_{\psi 0})dx} \\
&= B_1 \frac{\int \frac{\partial}{\partial \xi_\eta} \log\{p_{Y,X|Z}(y,x|z;\theta_0,\xi_{\eta 0},\xi_{\psi 0})\} p_{Y,X|Z}(y,x|z;\theta_0,\xi_{\eta 0},\xi_{\psi 0})dx}{\int p_{Y,X|Z}(y,x|z;\theta_0,\xi_{\eta 0},\xi_{\psi 0})dx} \\
&= B_1 E\{S_{\xi_\eta}(Y,X,Z)|Y=y,Z=z\}.
\end{aligned}
$$

Thus, $B_1 E\{S_{\xi_\eta}(Y,X,Z)|Y,Z\} = B_1 S^*_{\xi_\eta}(Y,Z) \in \Lambda_\eta$.

Similarly, taking the derivative of both sides of (B.3) with respect to ξ_ψ and evaluating them at the truth, we get $\frac{\partial}{\partial \xi_\psi} \log\left\{\int p_{Y,X|Z}(y,x|z;\theta_0,\xi_{\eta 0},\xi_{\psi 0})dx\right\} = 0$, which leads to $B_2 E\{S_{\xi_\psi}(Y,X,Z)|Y,Z\} = 0$.

Combining the arguments above, it follows that any element in the submodel nuisance tangent space, $h(Y,X,Z) = B_1 S_{\xi_\eta}(Y,X,Z) + B_2 S_{\xi_\psi}(Y,X,Z)$, must satisfy the condition

$$E\{h(Y,X,Z)|Y,Z\} \in \Lambda_\eta.$$

From (B.2), we get

$$\log\left\{\int p_{Y,X|Z}(y,x|z;\theta,\xi_\eta,\xi_\psi)dy\right\} = \log\{p_X(x;\theta,\xi_\eta,\xi_\psi)\} \tag{B.4}$$

$$
\begin{aligned}
& \frac{\partial}{\partial \xi_\eta} \log\left\{\int p_{Y,X|Z}(y,x|z;\theta_0,\xi_{\eta 0},\xi_{\psi 0})dy\right\} = \frac{\partial}{\partial \xi_\eta} \log\{p_X(x;\theta_0,\xi_{\eta 0},\xi_{\psi 0})\} \\
& \quad \frac{\int \frac{\partial}{\partial \xi_\eta} \log\{p_{Y,X|Z}(y,x|z;\theta_0,\xi_{\eta 0},\xi_{\psi 0})\} p_{Y,X|Z}(y,x|z;\theta_0,\xi_{\eta 0},\xi_{\psi 0})dy}{\int p_{Y,X|Z}(y,x|z;\theta_0,\xi_{\eta 0},\xi_{\psi 0})dy} = \frac{\partial}{\partial \xi_\eta} \log\{p_X(x;\theta_0,\xi_{\eta 0},\xi_{\psi 0})\},
\end{aligned}
$$

and $E\{S_{\xi_\eta}(Y,X,Z)|X,Z\} = S^*_{\xi_\eta}(X)$, where $S^*_{\xi_\eta}(x) = \frac{\partial}{\partial \xi_\eta} \log\{p_X(x;\theta_0,\xi_{\eta 0},\xi_{\psi 0})\}$. $S^*_{\xi_\eta}(X)$ has expectation 0 by the following argument. We have $\int p_X(x;\theta,\xi_\eta,\xi_\psi)dx = 1$, which implies that

$$\frac{\partial}{\partial \xi_\eta} \int p_X(x; \theta_0, \xi_{\eta 0}, \xi_{\psi 0}) dx = \int \frac{\partial}{\partial \xi_\eta} p_X(x; \theta_0, \xi_{\eta 0}, \xi_{\psi 0}) dx = 0,$$

so that

$$\int \frac{\partial}{\partial \xi_\eta} \log\{p_X(x; \theta_0, \xi_{\eta 0}, \xi_{\psi 0})\} p_X(x; \theta_0, \xi_{\eta 0}, \xi_{\psi 0}) dx = 0,$$

which implies that $E\{S^*_{\xi_\eta}(X)\} = 0$, as required.

Similarly, taking the derivative of both sides of (B.4) with respect to ξ_ψ and evaluating them at the truth, we get $E\{S_{\xi_\psi}(Y, X, Z)|X, Z\} = S^*_{\xi_\psi}(X)$, where

$$S^*_{\xi_\psi}(x) = \frac{\partial}{\partial \xi_\psi} \log\left\{p_X(x; \theta_0, \xi_{\eta 0}, \xi_{\psi 0})\right\},$$

and $E\{S^*_{\xi_\psi}(X)\} = 0$.

Therefore, if we define Λ_x as the space of all mean zero functions of X, i.e., $\Lambda_x = \{h(X) : E\{h(X)\} = 0\}$, then any element in the submodel nuisance tangent space, $h(Y, X, Z) = B_1 S_{\xi_\eta}(Y, X, Z) + B_2 S_{\xi_\psi}(Y, X, Z)$, must also satisfy the condition:

$$E\{h(Y, X, Z)|X, Z\} \in \Lambda_x.$$

To summarize, we have demonstrated that any element $h(Y, X, Z)$ that is spanned by the score vector $\{S^T_{\xi_\eta}(Y, X, Z), S^T_{\xi_\psi}(Y, X, Z)\}^T$ must satisfy

$$\text{(a).} \quad E\{h(Y, X, Z)|Y, Z\} \in \Lambda_\eta, \tag{B.5}$$

$$\text{(b).} \quad E\{h(Y, X, Z)|X, Z\} \in \Lambda_x. \tag{B.6}$$

With these relationships in mind, we conjecture that the nuisance tangent space consists of all functions $h(Y, X, Z)$ satisfying conditions (a) and (b) given in (B.5) and (B.6). We denote such as a space by $\Lambda^{(conj)}$. We can easily show that the space of functions satisfying (B.5) is given by $\Lambda_\eta + \Lambda_1$, where

$$\Lambda_1 = \{h_1(Y, X, Z) : \quad E\{h_1(Y, X, Z|Y, Z)\} = 0\}, \tag{B.7}$$

and the space of functions satisfying (B.6) is given by $\Lambda_x + \Lambda_2$, where

$$\Lambda_2 = \{h_2(Y, X, Z) : \quad E\{h_2(Y, X, Z|X, Z)\} = 0\}. \tag{B.8}$$

Consequently, the conjectured nuisance tangent space is $\Lambda^{(conj)} = (\Lambda_\eta + \Lambda_1) \cap (\Lambda_x + \Lambda_2)$. We have already proven that any element in a parametric submodel nuisance tangent space must belong to $\Lambda^{(conj)}$; in addition, it can be shown that the space $\Lambda^{(conj)}$ is closed. Therefore, the nuisance tangent space $\Lambda \subset \Lambda^{(conj)}$.

To prove that $\Lambda^{(conj)}$ is truly the nuisance tangent space, we need to show that any element in $\Lambda^{(conj)}$ can be represented as some element or a limit of elements from some parametric submodel nuisance tangent spaces. Consider some arbitrary bounded element $h(Y, X, Z) \in \Lambda^{(conj)}$; namely,

$$h(Y, X, Z) = h_\eta(Y, Z) + h_1(Y, X, Z) = h_x(X) + h_2(Y, X, Z) \tag{B.9}$$

for some $h_\eta(Y, Z) \in \Lambda_\eta$, $h_1(Y, X, Z) \in \Lambda_1$, $h_x(X) \in \Lambda_x$, and $h_2(Y, X, Z) \in \Lambda_2$. We will construct the parametric submodels in three steps.

Step 1. Because $h_\eta(Y, Z) \in \Lambda_\eta$, $h_\eta(Y, Z)$ is either the corresponding score vector of some parametric submodel $p_{Y|Z}(y, |z; \theta_0, \xi_\eta)$ or the limiting score vector of a sequence of parametric submodels $p_{Y|Z}(y, |z; \theta_0, \xi_{\eta j})$. For simplicity, we first assume the former case; that is,

$$\frac{\partial}{\partial \xi_\eta} \log\{p_{Y|Z}(y|z; \theta_0, \xi_{\eta 0})\} = h_\eta(y, z).$$

Without loss of generality, we can assume that this submodel contains the truth when $\xi_\eta = \xi_{\eta 0} = 0$. From here on θ is taken to equal θ_0 (the truth) and hence will be suppressed in the notation.

We begin by considering an approximation to the parametric-submodel given by

$$p^*_{Y,X|Z}(y, x|z, \xi_\eta) = p_{0Y,X|Z}(y, x|z)[1 + \xi_\eta^T\{h_\eta(y, z) + h_1(y, x, z)\}]. \tag{B.10}$$

This is a proper density function as long as ξ_η is chosen sufficiently close to 0, and it contains the truth if ξ_η is chosen to be 0. By construction, the score vector is given by $\{h_\eta(Y,Z)+h_1(Y,X,Z)\}$. We may show that the submodel given by (B.10) satisfies condition (ii) as follows:

$$\begin{aligned}\int p^*_{Y,X|Z}(y,x|z,\xi_\eta)dy &= \int p_{0Y,X|Z}(y,x|z)[1+\xi_\eta^T\{h_\eta(y,z)+h_1(y,x,z)\}]dy\\ &= \int p_{0Y,X|Z}(y,x|z)[1+\xi_\eta^T\{h_x(x)+h_2(y,x,z)\}]dy\\ &= p_{0X}(x)+\xi_\eta^T h_x(x)p_{0X}(x)+\xi_\eta^T p_{0X}(x)\int h_2(y,x,z)p_{0Y|X,Z}(y|x,z)dy\\ &= p_{0X}(x)\{1+\xi_\eta^T h_X(x)\},\end{aligned}$$

where the last equality follows because, by definition, the conditional expectation of $h_2(Y,X,Z)$ given (X,Z) is 0. This argument shows that X and Z are independent by this submodel. Therefore, this submodel satisfies condition (ii).

However, this submodel does not satisfy condition (i), because

$$\begin{aligned}\int p^*_{Y,X|Z}(y,x|z,\xi_\eta)dx &= \int p_{0Y,X|Z}(y,x|z)[1+\xi_\eta^T\{h_\eta(y,z)+h_1(y,x,z)\}]dx\\ &= p_{0Y|Z}(y|z)+p_{0Y|Z}\xi_\eta^T[h_\eta(y,z)+E\{h_1(Y,X,Z)|Y=y,Z=z\}]\\ &= p_{0Y|Z}(y|z)\{1+\xi_\eta^T h_\eta(y,z)\}\neq p_{Y|Z}(y|z,\xi_\eta).\end{aligned}$$

Step 2. In order to derive a model that satisfies conditions (i) and (ii) while still leading to the same score vector, we consider the following construction. Take the random vector (V,X,Z) which has density $p^*_{Y,X|Z}(v,x|z,\xi_\eta)$ as defined by (B.10). The idea is to perturb the random variable V slightly to ensure that the transformed random variable Y has conditional density $p_{Y|Z}(y|z,\xi_\eta)$ while not affecting the independence of X and Z or the score vector. Toward that end, define

$$(Y,X,Z)=\{G(V,Z,\xi_\eta),X,Z\}, \tag{B.11}$$

where $G\{V,Z,\xi_\eta\} = F_2^{-1}\{F_1(V|Z,\xi_\eta)|Z,\xi_\eta\}$; $F_1(y|Z,\xi_\eta)$ is the cdf for $p_{0Y|Z}(y|z)\{1+\xi_\eta^T h_\eta(y,z)\}$; and $F_2(y|Z,\xi_\eta)$ is the cdf for $p_{Y|Z}(y|z,\xi_\eta)$. By construction, the conditional distribution of Y given Z is $p_{Y|Z}(y|z,\xi_\eta)$, and the conditional density of X given Z does not change, i.e., $X \perp\!\!\!\perp Z$. Therefore, by construction, this submodel satisfies conditions (i) and (ii). In addition, when $\xi_\eta = 0$, $G\{V,Z,\xi_\eta = 0\} = V$, and $p_{Y,X|Z}(y,x|z,\xi_\eta = 0) = p_{0Y,X|Z}(y,x|z)$; i.e., contains the truth.

Next, we derive the density of $(Y,X|Z)$, i.e., $p_{Y,X|Z}(y,x|z,\xi_\eta)$, and show that the score vector of this density is still $\{h_\eta(Y,Z)+h_1(Y,X,Z)\}$ as required. As $Y = F_2^{-1}\{F_1(V|Z,\xi_\eta)|Z,\xi_\eta\}$, we obtain $V = F_1^{-1}\{F_2(Y|Z,\xi_\eta)|Z,\xi_\eta\}$. Consequently,

$$\frac{dV}{dY} = \frac{p_{Y|Z}(y|z,\xi_\eta)}{p_{0Y|Z}(v|z)\{1+\xi_\eta^T h_\eta(v,z)\}}.$$

Using the change of variable formula, the density of $(Y,X|Z)$ is

$$\begin{aligned} p_{Y,X|Z}(y,x|z,\xi_\eta) &= p_{0Y,X|Z}(v,x|z)[1+\xi_\eta^T\{h_\eta(v,z)+h_1(v,x,z)\}]\frac{dV}{dY} \\ &= p_{0Y|Z}(v|z)p_{0X|Y,Z}(x|v,z)[1+\xi_\eta^T\{h_\eta(v,z)+h_1(v,x,z)\}\frac{p_{Y|Z}(y|z,\xi_\eta)}{p_{0Y|Z}(v|z)\{1+\xi_\eta^T h_\eta(v,z)\}} \\ &= p_{0X|Y,Z}(x|v,z)p_{Y|Z}(y|z,\xi_\eta)\frac{1+\xi_\eta^T\{h_\eta(v,z)+h_1(v,x,z)\}}{\{1+\xi_\eta^T h_\eta(v,z)\}}, \end{aligned} \tag{B.12}$$

where $v = F_1^{-1}\{F_2(y|z,\xi_\eta)|z,\xi_\eta\}$.

Now, we will derive the score vector of $p_{Y,X|Z}(y,x|z,\xi_\eta)$.

$$\begin{aligned} \frac{\partial}{\partial\xi_\eta}\log\{p_{Y,X|Z}(y,x|z,\xi_{\eta 0})\} &= \frac{\frac{\partial}{\partial v}\{p_{0X|Y,Z}(x|v,z)\}\cdot\frac{\partial v}{\partial\xi_\eta}|_{\xi_\eta=0}}{p_{0X|Y,Z}(x|y,z)} + \frac{\partial}{\partial\xi_\eta}\log\{p_{Y|Z}(y|z,\xi_{\eta 0})\} \\ &+ \frac{h_\eta(v,z)+h_1(v,x,z)+\xi_\eta^T\frac{\partial}{\partial\xi_\eta}\{h_\eta(v,z)+h_1(v,x,z)\}}{1+\xi_\eta^T\{h_\eta(v,z)+h_1(v,x,z)\}}|_{\xi_\eta=0} - \frac{h_\eta(v,z)+\xi_\eta^T\frac{\partial h_\eta(v,z)}{\partial\xi_\eta}}{1+\xi_\eta^T h_\eta(v,z)}|_{\xi_\eta=0} \end{aligned}$$

$$= \frac{\frac{\partial}{\partial v}\{p_{0X|Y,Z}(x|v,z)\} \cdot \frac{\partial v}{\partial \xi_\eta}|_{\xi_\eta=0}}{p_{0X|Y,Z}(x|y,z)} + h_\eta(y,z) + h_\eta(y,z) + h_1(y,x,z) - h_\eta(y,z)$$
$$= h_\eta(y,z) + h_1(y,x,z) + \frac{\frac{\partial}{\partial v}\{p_{0X|Y,Z}(x|v,z)\} \cdot \frac{\partial v}{\partial \xi_\eta}|_{\xi_\eta=0}}{p_{0X|Y,Z}(x|y,z)}.$$

In the above argument, we have used the facts that when $\xi_\eta = 0$, $v = y$, and that $\frac{\partial}{\partial \xi_\eta}\log\{p_{Y|Z}(y|z,\xi_{\eta 0})\} = h_\eta(y,z)$. Note that if $\frac{\partial v}{\partial \xi_\eta}|_{\xi_\eta=0} = 0$, then the score vector is $h_\eta(y,z) + h_1(y,x,z)$ as needed. So in the following we will show that $\frac{\partial v}{\partial \xi_\eta}|_{\xi_\eta=0} = 0$.

First note, under suitable regularity conditions,

$$\frac{\partial F_2(y|z,\xi_\eta)}{\partial \xi_\eta}|_{\xi_\eta=0} = \frac{\partial}{\partial \xi_\eta}\int_{-\infty}^{y} p_{Y|Z}(u|z,\xi_\eta)du|_{\xi_\eta=0} = \int_{-\infty}^{y} \frac{\partial}{\partial \xi_\eta} p_{Y|Z}(u|z,\xi_\eta)|_{\xi_\eta=0}\, du$$
$$= \int_{-\infty}^{y} p_{0Y|Z}(u|z)\frac{\partial}{\partial \xi_\eta}\log\{p_{Y|Z}(u|z,\xi_\eta)\}|_{\xi_\eta=0}du = \int_{-\infty}^{y} p_{0Y|Z}(u|z)h_\eta(u,z)du.$$

Similarly,

$$\frac{\partial F_1(y|z,\xi_\eta)}{\partial \xi_\eta}|_{\xi_\eta=0} = \int_{-\infty}^{y} \frac{\partial}{\partial \xi_\eta} p_{0Y|Z}(u|z)\{1 + \xi_\eta^T h_\eta(u|z)\}|_{\xi_\eta=0}\, du = \int_{-\infty}^{y} p_{0Y|Z}(u|z)h_\eta(u,z)\, du.$$

Consequently, $\frac{\partial F_1(y|z,\xi_\eta)}{\partial \xi_\eta}|_{\xi_\eta=0} = \frac{\partial F_2(y|z,\xi_\eta)}{\partial \xi_\eta}|_{\xi_\eta=0}$.

By construction, $y = F_2^{-1}\{F_1(v|z,\xi_\eta)|z,\xi_\eta\}$. Thus $F_2(y|z,\xi_\eta) = F_1(v|z,\xi_\eta)$, and it follows that

$$\frac{\partial F_2(y|z,\xi_\eta)}{\partial \xi_\eta}|_{\xi_\eta=0} = \frac{\partial F_1(v|z,\xi_\eta)}{\partial v} \cdot \frac{\partial v}{\partial \xi_\eta}|_{\xi_\eta=0} + \frac{\partial F_1(v|z,\xi_\eta)}{\partial \xi_\eta}|_{\xi_\eta=0}.$$

Notice that, when $\xi_\eta = 0$, we have $v = y$, and $\frac{F_2(y|z,\xi_\eta)}{\xi_\eta}|_{\xi_\eta=0} = \frac{F_1(y|z,\xi_\eta)}{\xi_\eta}|_{\xi_\eta=0}$, so that

$$\frac{\partial F_1(v|z,\xi_\eta)}{\partial v} \cdot \frac{\partial v}{\partial \xi_\eta}|_{\xi_\eta=0} = 0,$$

and it follows that

$$p_{0Y|Z}(y|z) \cdot \frac{\partial v}{\partial \xi_\eta}|_{\xi_\eta=0} = 0, \quad \text{which implies} \quad \frac{\partial v}{\partial \xi_\eta}|_{\xi_\eta=0} = 0.$$

Therefore, we have constructed a submodel given by (B.12) that satisfies conditions (i) and (ii). In addition, this submodel has a score vector equal to $h_\eta(Y,Z) + h_1(Y,X,Z) = h_x(X) + h_2(Y,X,Z)$, which is arbitrarily chosen from the conjecture space.

Step 3. Recall that in the above arguments, we have assumed that $h_\eta(Y,Z)$ is the score vector of some parametric submodel $p_{Y|Z}(y|z;\theta_0,\xi_\eta)$. Under this assumption, it has been demonstrated that a bounded element $h(Y,X,Z)$ can be represented as an element from some parametric submodel nuisance tangent space. More generally, $h_\eta(Y,Z)$ may be the limit of score vectors of a sequence of parametric submodels $p_{jY|Z}(y|z;\theta_0,\xi_{\eta j})$; i.e.,

$$\lim_{j\to\infty} \frac{\partial}{\partial \xi_\eta} \log\{p_{jY|Z}(y|z;\theta_0,\xi_{\eta 0})\} = h_\eta(y,z).$$

In this situation, almost identical arguments to those used above can be used to construct a sequence of submodels that satisfies condition (i) and (ii), and also the limit of the corresponding score vectors is $h_\eta(Y,Z)$.

Combining the above arguments, we have shown that any bounded element in $\Lambda^{(conj)}$ can be represented as some element or a limit of elements from some parametric submodel nuisance tangent spaces. As any element in $\Lambda^{(conj)}$ is either bounded or limit of bounded elements, it follows that the nuisance tangent space is

$$\Lambda = \Lambda^{(conj)} = (\Lambda_\eta + \Lambda_1) \cap (\Lambda_x + \Lambda_2). \tag{B.13}$$

As we argued earlier, estimating functions used to derive estimating equations that lead to semiparametric estimators for θ are orthogonal to the nuisance tangent space. Therefore, we now derive the orthogonal complement to the nuisance tangent space. To do so, we use result that that the orthogonal complement of the sum of two linear spaces is equal to the intersection of the orthogonal complements. That is, if H_1, H_2 are closed linear subspaces contained in the Hilbert space $\mathcal{H}$, then

$$(H_1 + H_2)^{\perp} = H_1^{\perp} \cap H_2^{\perp}, \tag{B.14}$$

We derived the nuisance tangent space Λ as given in (B.13). Therefore, using (B.14) we thus have that the orthogonal complement of the nuisance tangent space is

$$\Lambda^{\perp} = (\Lambda_{\eta}^{\perp} \cap \Lambda_1^{\perp}) + (\Lambda_x^{\perp} \cap \Lambda_2^{\perp}). \tag{B.15}$$

We examine the components making up the sum of spaces in (B.15) separately.

Complement of Λ_1. This space is given by

$$\Lambda_1^{\perp} = \{h(Y, Z), E\{h(Y, Z)\} = 0\}. \tag{B.16}$$

Proof: Suppose $E\{h(Y, Z)\} = 0$, and $h_1(Y, X, Z) \in \Lambda_1$, i.e., $E\{h_1(Y, X, Z|Y, Z)\} = 0$. Then

$$\begin{aligned} E\{h_1^T(Y, X, Z)h(Y, Z)\} &= E[E\{h_1^T(Y, X, Z)h(Y, Z)|Y, Z\}] \\ &= E[h^T(Y, Z)E\{h_1(Y, X, Z|Y, Z)\}] = 0. \end{aligned}$$

The last equality follows as $E\{h_1(Y, X, Z|Y, Z)\} = 0$ by assumption. Therefore, any mean-zero function of (Y, Z), $h(Y, Z)$, is orthogonal to Λ_1.

To finish the proof, we also must prove that any $h \in \mathcal{H}$ can be written as $h_1 + h_2$, where $h_1 \in \Lambda_1$, and $h_2 \in \{h(Y, Z), E\{h(Y, Z)\} = 0\}$. For any $h(Y, X, Z)$, construct $h_2(Y, Z) = E\{h(Y, X, Z)|Y, Z\}$, and $h_1(Y, X, Z) = h(Y, X, Z) - h_2(Y, Z)$. It is easy to verify that the constructed $h_1 \in \Lambda_1$ and $h_2 \in \{h(Y, Z), E\{h(Y, Z)\} = 0\}$.

Complement of $(\Lambda_{\eta}^{\perp} \cap \Lambda_1^{\perp})$. Because $\Lambda_1^{\perp} = \{h(Y, Z), E\{h(Y, Z)\} = 0\}$, then the space $(\Lambda_{\eta}^{\perp} \cap \Lambda_1^{\perp})$ consists of all elements which belong to the Hilbert space $\mathcal{H}_{Y,Z} = \{h(Y, Z) : E\{h(Y, Z)\} = 0\}$, i.e., all mean zero functions of (Y, Z), that are orthogonal to the nuisance tangent space Λ_η. This is precisely the orthogonal complement of the nuisance tangent space for the parametric submodel $p_{Y|Z}(y|z;\theta,\eta)$. Consequently, the space $(\Lambda_\eta^\perp \cap \Lambda_1^\perp)$ is the space from which estimating functions for θ are derived without the consideration of the auxiliary covariates. Therefore, we call this space the space of estimating functions $\mathcal{E} = (\Lambda_\eta^\perp \cap \Lambda_1^\perp)$.

Complement of $(\Lambda_x^\perp \cap \Lambda_2^\perp)$. The same techniques used to find the space $\Lambda_1^\perp$ may be used here to prove that $\Lambda_2^\perp = \{h(X,Z) : E\{h(X,Z)\} = 0\}]$. Consequently, $(\Lambda_x^\perp \cap \Lambda_2^\perp)$ consists of all mean-zero functions of X and Z that are orthogonal to functions of X. That is,

$$(\Lambda_x^\perp \cap \Lambda_2^\perp) = \{h(X,Z) : E\{h(X,Z)|X\} = 0\}. \tag{B.17}$$

Proof : if $E\{h(X,Z)|X\} = 0$, and $h_x(X) \in \Lambda_x$, then

$$E\{h^T(X,Z)h_x(X)\} = E[E\{h^T(X,Z)|X\}h_x(X)] = 0.$$

That is, $h(X,Z)$ is orthogonal to Λ_x. Moreover, any mean-zero functions of (X,Z) can be written as h_1+h_2, where $h_1 \in \Lambda_x$, and $h_2 \in \{h(X,Z) : E\{h(X,Z)|X\} = 0\}$. For any $h(X,Z)$ which has mean zero, construct $h_1(X) = E\{h(X,Z)|X\}$, and $h_2(X,Z) = h(X,Z) - h_1(X)$. Clearly $E\{h_1(X)\} = 0$, and $E\{h_2(X,Z)|X\} = 0$ as required.

We refer to this space as the Augmentation Space, denoted by $\mathcal{A}$. Therefore, we have shown that the space orthogonal to the nuisance tangent space is given by $\Lambda^\perp = \mathcal{E} + \mathcal{A}$. Thus, the class of estimating functions for θ (and hence β) based on all the data (Y,X,Z) lies in this space, so that an estimating function for θ may be written as the sum of an estimating function $m(Y,Z;\theta)$ based on (Y,Z) alone (an element of $\mathcal{E}$) and an element of $\mathcal{A}$.

Accordingly, we characterize elements of $\mathcal{A}$. All functions of X and Z may be written as $\sum_{g=1}^k I(Z=g)a_g(X)$ for arbitrary functions $a_g(X)$, $g = 1,\ldots,k$. Thus, we can write any function $h(X, Z)$ satisfying $E\{h(X, Z)|X\} = 0$ as

$$h(X, Z) = \sum_{g=1}^{k} I(Z = g)a_g(X) - E\left\{ \sum_{g=1}^{k} I(Z = g)a_g(X) \middle| X \right\} = \sum_{g=1}^{k} \{I(Z = g) - \pi_g\}a_g(X). \tag{B.18}$$

Thus the form of all estimating functions is

$$m(Y, Z; \theta) + \sum_{g=1}^{k} \{I(Z = g) - \pi_g\}a_g(X),$$

which may be written equivalently in the form given in (12) of the main paper.

Web Appendix C: Derivation of Optimal Estimating Function (14)

The choice of functions $a_g(X)$, $g = 1, \dots, k$ resulting in the optimal estimator, i.e., an estimator solving (12) in the main paper such that its variance is as small as possible, may be deduced from Theorem 4.5 of Tsiatis (2006). Alternatively, we derive such $a_g(X)$ directly. By the principles in Chapter 3 of Tsiatis (2006), the element of $\mathcal{E} + \mathcal{A}$ with smallest variance for a given $m(Y, Z; \theta) \in \mathcal{E}$ is the projection of $m(Y, Z; \theta)$ onto $\mathcal{A}$. Thus, we wish to find $a_g^*(X)$, $g = 1, \dots, k$, such that

$$E\left(\left[m(Y, Z; \theta) - \sum_{g=1}^{k} \{I(Z = g) - \pi_g\}a_g^*(X) \right] \left[\sum_{g=1}^{k} \{I(Z = g) - \pi_g\}a_g(X) \right] \right) = 0$$

for all $a_g(X)$, $g = 1, \dots, k$. Taking $a_g(X) = 0$ for $g \neq j$, we thus wish to find $a_g^*(X)$, $g = 1, \dots, k$, such that

$$E\left(\left[E\{m(Y, Z; \theta)|X, Z\} - \sum_{g=1}^{k} \{I(Z = g) - \pi_g\}a_g^*(X) \right] \{I(Z = j) - \pi_j\} \middle| X \right) = 0 \tag{C.1}$$

for each $j = 1, \dots, k$. It is straightforward to show that, writing $E\{m(Y, Z; \theta)|X, Z\} = \sum_{g=1}^{k} I(Z = g)E\{m(Y, g; \theta)|X, Z = g\}$, (C.1) implies that we must have

$$E\{m(Y, j; \theta)|X, Z = j\} - a_j^*(X) - \sum_{g=1}^{k} \left[E\{m(Y, g)|X, Z = g\} - a_g^*(X) \right] \pi_g = 0 \tag{C.2}$$

for all $j = 1, \ldots, k$. Expression (C.2) is satisfied when

$$a_g^*(X) = E\{m(Y, g; \theta)|X, Z = g\}, \quad g = 1, \ldots, k,$$

yielding the estimating function in (14) of the main paper.

Web Appendix D: Applications

Here, we give additional information and results for the two applications in Section 7 of the main paper.

The covariates in the analysis of the PURSUIT clinical trial data given in Section 7.1 of the main paper, are as follows: Age (years), height (cm), weight (kg), body mass index (kg/m^2), heart rate (beats per min), pulse (beats per min), gender (0=female, 1 = male), race (Caucasian, Black, Asian, Hispanic, Native American, Asiatic Indian, Other), geographic region (Eastern Europe, Western Europe, North America, Latin America), smoking status (current, former, never), diastolic and systolic blood pressure (mmHg), creatinine clearance (ml/min), rales (None, $\leq 1/3$, $\geq 1/3$), creatine kinase and creatine kinase-MB ratios, hours from symptoms to treatment, indicators of disease history (myocardial infarction at enrollment, prior myocardial infarction, ST depression, angina, diabetes, congestive heart failure, hypercholesterolemia, hypertension, renal insufficiency, peripheral vascular disease, family history of coronary artery disease) and treatment history indicators (percutaneous coronary intervention within 72 hours of randomization, previous percutaneous coronary angioplasty, calcium blockers, beta blockers, digoxin, nitroglycerin, coronary artery bypass graft).

For the ACTG 175 data in Section 7.2 of the main paper, we also carried out tests for each pairwise comparison of regimens $g = 2, 3, 4$ against the control treatment, $g = 1$, ZDV monotherapy. The three usual unadjusted Wald tests for these pairwise differences yield T_n for each comparison of 56.85, 19.22, and 20.55 for comparing groups 2, 3, and 4 against group 1, respectively; the corresponding proposed statistics $\widehat{T}_n^*$ are 98.27, 35.28, and 46.75. Of course, all test statistics reflect very strong evidence in favor of real differences in each case; however, notably, the augmented test statistics are much larger in each case.

The sample size in this trial was very large, so that all analyses are easily able to uncover treatment differences. To demonstrate that such results are possible in cases where the evidence is less clear-cut, we repeated these analyses on the data from a random subset of $n = 124$ subjects. The three pairwise unadjusted Wald test statistics are 2.23, 1.45, and 2.18, with p-values 0.07, 0.11, 0.07; the corresponding adjusted statistics are 4.77, 4.87, and 10.12, with p-values 0.01, 0.01, and < 0.001. Likewise, the three-degree-of-freedom unadjusted Wald and Kruskal-Wallis statistics (p-values) are 3.36 (0.34) and 2.35 (0.50), while the adjusted versions are 11.87 (0.01) and 4.59 (0.20). Although it is certainly not guaranteed that smaller p-values will be obtained for any given realization of data, these results demonstrate that the proposed adjustment methods are capable of effecting such improvements.

The 12 auxiliary covariates used in all analyses of these data, also reported in Tsiatis et al. (2007), are as follows: continuous variables: CD4 count (cells/mm^3), CD8 count (cells/mm^3), age (years), weight (kg), Karnofsky score (scale of 0-100), and indicator variables for hemophilia, homosexual activity, history of intravenous drug use, race (0=white, 1=non-white), gender (0=female), antiretroviral history (0=naive, 1=experienced), and symptomatic status (0=asymptomatic).

Additional References

Davidian, M., Tsiatis, A. A., and Leon, S. (2005). Semiparametric estimation of treatment effect in a pretest-posttest study with missing data (with Discussion). *Statistical Science* **20,** 261–301.