Reverse Regression: A Method for Joint Analysis of Multiple Endpoints in Randomized Clinical Trials

Zhiwei Zhang

U.S. Food and Drug Administration

Supplementary Material

This appendix consists of (1) examples of $t(y)$ motivated by parametric models for (F,G) and (2) proofs of theoretical results stated in the main text.

S1 Examples of $t(y)$ Motivated by Parametric Models for (F,G)

We restrict attention to a single endpoint in this section. If we assume $F=N(\theta_{01},\theta_{02})$ and $G=N(\theta_{11},\theta_{12})$, then the reverse regression model involves y and y^2 with respective regression coefficients $\theta_{11}\theta_{12}^{-1}-\theta_{01}\theta_{02}^{-1}$ and $2^{-1}(\theta_{02}^{-1}-\theta_{12}^{-1})$. If we assume, in addition, that $\theta_{02}=\theta_{12}=\theta_2$, then the y^2 term is not needed and the regression coefficient for y becomes $(\theta_{11}-\theta_{01})\theta_2^{-1}$, which equals 0 if and only if $F=G$. As another example, consider the following gamma model:

$$p_\theta(y)=p_{\theta_1,\theta_2}(y)=\frac{y^{\theta_1-1}\exp(-y/\theta_2)}{\Gamma(\theta_1)\theta_2^{\theta_1}},\qquad y>0.$$

If $f=p_{\theta_{01},\theta_{02}}$ and $g=p_{\theta_{11},\theta_{12}}$, then the corresponding reverse regression involves $\log y$ and y, with coefficients $\theta_{11}-\theta_{01}$ and $\theta_{02}^{-1}-\theta_{12}^{-1}$, respectively. The $\log y$ term can be omitted if we assume $\theta_{01}=\theta_{11}$ (i.e., common shape), while the y term is unnecessary if $\theta_{02}=\theta_{12}$ (i.e., common scale). In the latter case, the reverse regression takes the same form as in the normal case with equal variance.

For a categorical outcome, the reverse regression typically involves a collection of dummy variables unless additional structure is imposed. For example, one could set the right side of (2) to $\alpha+\sum_{k=1}^K\beta_k 1_{y=k}$ if Y takes its value from $\{0,\dots,K\}$ for some $K\ge 1$. If the values of Y follow a natural order, it may be appropriate to specify a linear association structure as in Agresti (1990, Section 8.1). Let $v_0\le\cdots\le v_K$ be given; then linear association between Z and Y means that the following log-odds ratio is linear in some unknown parameter β:

$$\log\frac{\mathrm{P}[Y=k|Z=1]\,\mathrm{P}[Y=j|Z=0]}{\mathrm{P}[Y=k|Z=0]\,\mathrm{P}[Y=j|Z=1]}=\beta(v_k-v_j),\qquad 0\le j<k\le K.$$

A simple and intuitive characterization of the above is available through the corresponding reverse regression model:

$$\text{logit}(\text{P}[Z = 1|Y]) = \alpha + \beta V,$$

where $V = \sum_{k=1}^{K} v_k 1_{Y=k}$. Thus, in the present setting, imposing a linear association structure amounts to assigning a numerical score to each level of Y and treating it as a continuous variable in the logistic regression.

S2 Proofs

In the proofs of Theorems 1–3, we shall use the subscript 0 to denote the true value of a parameter and abbreviate $s_0 = s(\cdot, \cdot; \alpha_0, \beta_0)$ and $I_0 = I_{\alpha_0,\beta_0}$. We also write $\mathbb{P}_0$ for the true distribution of (Z, Y), $\mathbb{P}_n$ for the empirical distribution, and $\mathbb{Q}_n = \sqrt{n}(\mathbb{P}_n - \mathbb{P}_0)$ for the empirical process based on the (Z_i, Y_i), $i = 1, \ldots, n$.

Proof of Theorem 1

Let us write

$$\begin{aligned}
&\sqrt{n}(\widehat{F} - F_0)h \\
&= \sqrt{n}\left[\mathbb{P}_n \frac{h(Y)}{1 - \widehat{\pi} + \widehat{\pi}\exp\{\widehat{\alpha}^* + \widehat{\beta}^{\text{T}} t(Y)\}} - \mathbb{P}_0 \frac{h(Y)}{1 - \pi_0 + \pi_0 \exp\{\alpha_0^* + \beta_0^{\text{T}} t(Y)\}}\right] \\
&= \sqrt{n}[\mathbb{P}_n \psi_{\widehat{\pi},\widehat{\alpha},\widehat{\beta}}(Y) - \mathbb{P}_0 \psi_{\pi_0,\alpha_0,\beta_0}(Y)] \\
&= \mathbb{Q}_n \psi_{\widehat{\pi},\widehat{\alpha},\widehat{\beta}}(Y) + \sqrt{n}\mathbb{P}_0[\psi_{\widehat{\pi},\widehat{\alpha},\widehat{\beta}}(Y) - \psi_{\pi_0,\alpha_0,\beta_0}(Y)],
\end{aligned}$$

where

$$\psi_{\pi,\alpha,\beta}(y) = (1 - \pi)^{-1} h(y)[1 - \text{logit}^{-1}\{\alpha + \beta^{\text{T}} t(y)\}].$$

It follows from Lemma 2.6.18 of van der Vaart and Wellner (1996) that the class of functions $\psi_{\pi,\alpha,\beta}$ with (π, α, β) ranging over a neighborhood of $(\pi_0, \alpha_0, \beta_0)$ is VC-subgraph and hence Donsker. By the dominated convergence theorem, the map $(\pi, \alpha, \beta) \mapsto \psi_{\pi,\alpha,\beta} \in L_2(\mathbb{P}_0)$ is continuous at $(\pi_0, \alpha_0, \beta_0)$. Further, by the continuous mapping theorem, $\|\psi_{\widehat{\pi},\widehat{\alpha},\widehat{\beta}} - \psi_{\pi_0,\alpha_0,\beta_0}\|_2 = o_p(1)$, where the L_2-norm is evaluated under the true distribution of (Z, Y) with (π, α, β) regarded as an index. Now it follows from theorem 19.24 of van der Vaart (1998) that

$$\mathbb{Q}_n \psi_{\widehat{\pi},\widehat{\alpha},\widehat{\beta}} = \mathbb{Q}_n \psi_{\pi_0,\alpha_0,\beta_0} + o_p(1). \tag{S2.1}$$

Again by the dominated convergence theorem, the map $(\pi, \alpha, \beta) \mapsto \mathbb{P}_0 \psi(\pi, \alpha, \beta)$ is differentiable at $(\pi_0, \alpha_0, \beta_0)$ with derivative $(1-\pi_0)^{-1}(F_0 h, -a^{\text{T}})^{\text{T}}$, where $a^{\text{T}} = \text{E}\{\text{E}[Z|Y](1 - \text{E}[Z|Y])h(Y)(1, t(Y)^{\text{T}})\}$. Hence

$$\sqrt{n}\mathbb{P}_0[\psi_{\widehat{\pi},\widehat{\alpha},\widehat{\beta}} - \psi_{\pi_0,\alpha_0,\beta_0}] = \frac{(F_0 h)\mathbb{Q}_n Z}{1 - \pi_0} - \frac{\sqrt{n} a^{\text{T}}}{1 - \pi_0} \begin{pmatrix} \widehat{\alpha} - \alpha_0 \\ \widehat{\beta} - \beta_0 \end{pmatrix} + o_p(1) \tag{S2.2}$$

by the delta method. Combining (S2.1), (S2.2) and (6) yields

$$\begin{aligned}&\sqrt{n}(\widehat{F}-F_0)h\\ &=(1-\pi_0)^{-1}\mathbb{Q}_n\left\{(1-\mathrm{E}[Z|Y])h(Y)-a^{\mathrm{T}}I_0^{-1}s_0(Z,Y)-(F_0h)(1-Z)\right\}+o_p(1). \quad \text{(S2.3)}\end{aligned}$$

Similarly, we have

$$\sqrt{n}(\widehat{G}-G_0)h=\pi_0^{-1}\mathbb{Q}_n\left\{\mathrm{E}[Z|Y]h(Y)+a^{\mathrm{T}}I_0^{-1}s_0(Z,Y)-(G_0h)Z\right\}+o_p(1). \quad \text{(S2.4)}$$

Subtracting (S2.3) from (S2.4) shows that $(\widehat{G}-\widehat{F})h$ is asymptotically linear with influence function U_1+U_2, where U_1 and U_2 are defined in the theorem. The asymptotic variance of $(\widehat{G}-\widehat{F})h$ is given by $\mathrm{var}(U_1)+\mathrm{var}(U_2)$ because U_1 and U_2 are uncorrelated.

Proof of Theorem 2

For estimating $(G_0-F_0)h$, substituting π_0 and/or λ_0 in (5) amounts to adding a constant multiple of Z to the influence function. Therefore it suffices to show that Z is uncorrelated with the influence function for $(\widehat{G}-\widehat{F})h$ (i.e., U_1+U_2). To this end, we write

$$\langle Z,U_1\rangle_2=\langle \mathrm{E}[Z|Y],U_1\rangle_2=\mathrm{E}\{\mathrm{E}[Z|Y]U_1\}, \quad \text{(S2.5)}$$

where $\langle\cdot,\cdot\rangle_2$ denotes inner product in $L_2(\mathbb{P}_0)$. Next, define the operator A by

$$Ah'=\mathrm{E}\{(Z-\mathrm{E}[Z|Y])^2h'(Y)(1,t(Y)^{\mathrm{T}})\}^{\mathrm{T}}=\langle (Z-\mathrm{E}[Z|Y])h'(Y),s_0(Z,Y)\rangle_2,$$

where the inner product is taken elementwise. Note that a in the proof of Theorem 1 is just Ah. Obviously, A is linear. Furthermore, for any constant c, we have

$$\begin{aligned}(Ac)^{\mathrm{T}}I_0^{-1}s_0&=\langle c(Z-\mathrm{E}[Z|Y]),s_0^{\mathrm{T}}\rangle_2I_0^{-1}s_0=\langle ce_1^{\mathrm{T}}s_0,s_0^{\mathrm{T}}\rangle_2I_0^{-1}s_0\\ &=ce_1^{\mathrm{T}}\langle s_0,s_0^{\mathrm{T}}\rangle_2I_0^{-1}s_0=ce_1^{\mathrm{T}}I_0I_0^{-1}s_0=ce_1^{\mathrm{T}}s_0=c(Z-\mathrm{E}[Z|Y]),\end{aligned}$$

where $e_1=(1,0,\dots,0)^{\mathrm{T}}$. Now we can write $U_2=a_0^{\mathrm{T}}I_0^{-1}s_0$ where $a_0=Ah_0$ and

$$h_0=\frac{h}{\pi_0(1-\pi_0)}-\frac{G_0h}{\pi_0}-\frac{F_0h}{1-\pi_0}=\frac{h-G_0h}{\pi_0}+\frac{h-F_0h}{1-\pi_0}.$$

It follows that

$$\begin{aligned}\langle Z,U_2\rangle_2&=\langle Z-\mathrm{E}[Z|Y],U_2\rangle_2=\langle e_1^{\mathrm{T}}s_0,a_0^{\mathrm{T}}I_0^{-1}s_0\rangle_2=e_1^{\mathrm{T}}\langle s_0,s_0^{\mathrm{T}}\rangle_2I_0^{-1}a_0=e_1^{\mathrm{T}}I_0I_0^{-1}a_0\\ &=e_1^{\mathrm{T}}a_0=\langle (Z-\mathrm{E}[Z|Y])h_0,e_1^{\mathrm{T}}s_0\rangle_2=\mathrm{E}\{(Z-\mathrm{E}[Z|Y])^2h_0(Y)\}\\ &=\mathrm{E}\{\mathrm{E}[Z|Y](1-\mathrm{E}[Z|Y])h_0(Y)\}.\end{aligned} \quad \text{(S2.6)}$$

It is now straightforward to verify that (S2.5) and (S2.6) cancel each other, proving the theorem.

Proof of Theorem 3

We shall borrow notation and results from the previous proofs. It can be argued as in the proof of Theorem 1 that $(\widetilde{G}-\widetilde{F})h$ is asymptotically linear with influence function U_1+U_2' with $U_2'=(Z-\mathrm{E}[Z|Y])h_0(Y)$. Since U_1 and U_2' are uncorrelated, the asymptotic variance of $(\widetilde{G}-\widetilde{F})h$ is given by $\mathrm{var}(U_1)+\mathrm{var}(U_2')$. Recall that the asymptotic variance of $(\widehat{G}-\widehat{F})h$ is $\mathrm{var}(U_1)+\mathrm{var}(U_2)$ with $U_2=a_0^{\mathrm{T}}I_0^{-1}s_0$. Thus it suffices to show that $\mathrm{var}(U_2)\le\mathrm{var}(U_2')$. To this end, we write

$$\mathrm{var}(U_2)=a_0^{\mathrm{T}}I_0^{-1}\,\mathrm{var}(s_0)I_0^{-1}a_0=a_0^{\mathrm{T}}I_0^{-1}I_0I_0^{-1}a_0=a_0^{\mathrm{T}}I_0^{-1}a_0=\|I_0^{-1/2}a_0\|^2.$$

Denote $b=I_0^{-1/2}a_0=\langle(Z-\mathrm{E}[Z|Y])h_0,I_0^{-1/2}s_0\rangle_2=\langle U_2',I_0^{-1/2}s_0\rangle_2$. Then

$$\begin{aligned}\|b\|^2=b^{\mathrm{T}}b=b^{\mathrm{T}}\langle U_2',I_0^{-1/2}s_0\rangle_2&=\langle U_2',b^{\mathrm{T}}I_0^{-1/2}s_0\rangle_2\le\|U_2'\|_2\|b^{\mathrm{T}}I_0^{-1/2}s_0\|_2\\&=\mathrm{sd}(U_2')\,\mathrm{sd}(b^{\mathrm{T}}I_0^{-1/2}s_0)=\mathrm{sd}(U_2')(b^{\mathrm{T}}I_0^{-1/2}I_0I_0^{-1/2}b)^{1/2}=\mathrm{sd}(U_2')\|b\| \qquad \text{(S2.7)}\end{aligned}$$

by the Cauchy-Schwartz inequality. It follows that $\mathrm{var}(U_2)=\|b\|^2\le\mathrm{var}(U_2')$.

We now show that $\mathrm{var}(U_2)=\mathrm{var}(U_2')$ if and only if $h\in\mathrm{lin}\{1,t\}$, where $\mathrm{lin}\{\cdot\}$ denotes linear span. First, suppose $h\in\mathrm{lin}\{1,t\}$ so that $h_0\in\mathrm{lin}\{1,t\}$. Then there exists a vector b^* such that $U_2'=(Z-\mathrm{E}[Z|Y])h_0=b^{*\mathrm{T}}I_0^{-1/2}s_0$. By definition, $b=\langle U_2',I_0^{-1/2}s_0\rangle_2=I_0^{-1/2}I_0I_0^{-1/2}b^*=b^*$. Hence $U_2'=b^{\mathrm{T}}I_0^{-1/2}s_0$ and equality in (S2.7) holds. Conversely, suppose $\mathrm{var}(U_2)=\mathrm{var}(U_2')$. In the trivial case that $b=0$ (which implies $0=\mathrm{var}(U_2)=\mathrm{var}(U_2')$), we have $0=U_2'=(Z-\mathrm{E}[Z|Y])h_0(Y)$ almost surely. It follows that, almost surely, $h_0=0$, h is a constant, and hence $h\in\mathrm{lin}\{1,t\}$. So assume $b\neq0$; then equality in (S2.7) implies that

$$U_2'=cb^{\mathrm{T}}I_0^{-1/2}s_0 \qquad \text{(S2.8)}$$

for some constant c. (In fact, $c=1$ because $b=\langle U_2',I_0^{-1/2}s_0\rangle_2=cb$.) Because $Z-\mathrm{E}[Z|Y]\neq0$, (S2.8) implies that $h_0=b^{\mathrm{T}}I_0^{-1/2}(1,t^{\mathrm{T}})^{\mathrm{T}}\in\mathrm{lin}\{1,t\}$ and hence $h\in\mathrm{lin}\{1,t\}$. This completes the proof.

Proof of Theorem 4

Because stochastic ordering is preserved under a monotone transformation, we may assume without loss of generality that each t_j is identity. It suffices to show that F_1 is stochastically smaller than G_1 under the given conditions. Let $y_1^*\in\mathbb{R}$ be given. If $F_1(y_1^*)=0$ or 1, then $G_1(y_1^*)=F_1(y_1^*)$ because F and G are assumed to have the same support. Otherwise, we write

$$G_1(y_1^*)/F_1(y_1^*)=\mathrm{E}[g(Y)/f(Y)|Y_{[1]}\le y_1^*,Z=0]. \qquad \text{(S2.9)}$$

On the other hand, we have

$$\begin{aligned}1&=\mathrm{E}[g(Y)/f(Y)|Z=0]\\&=\mathrm{P}[Y_{[1]}\le y_1^*|Z=0]\,\mathrm{E}[g(Y)/f(Y)|Y_{[1]}\le y_1^*,Z=0]\\&\quad+\mathrm{P}[Y_{[1]}>y_1^*|Z=0]\,\mathrm{E}[g(Y)/f(Y)|Y_{[1]}>y_1^*,Z=0].\end{aligned}$$

Hence (S2.9) will be ≤ 1 if and only if

$$\mathrm{E}[g(Y)/f(Y)|Y_{[1]} \leq y_1^*, Z=0] \leq \mathrm{E}[g(Y)/f(Y)|Y_{[1]} > y_1^*, Z=0]. \tag{S2.10}$$

To this end, we define

$$\begin{aligned} h_J(y) &= g(y)/f(y) = \exp\left(\alpha^* + \sum_{j=1}^{J} \beta_j y_j\right), \\ h_{J-1}(y_1,\ldots,y_{J-1}) &= \mathrm{E}[h_J(Y)|Y_{[1]}=y_1,\ldots,Y_{[J-1]}=y_{J-1},Z=0], \\ h_{J-2}(y_1,\ldots,y_{J-2}) &= \mathrm{E}[h_{J-1}(Y)|Y_{[1]}=y_1,\ldots,Y_{[J-2]}=y_{J-2},Z=0], \\ &\ldots, \\ h_1(y_1) &= \mathrm{E}[h_2(Y_{[1]},Y_{[2]})|Y_{[1]}=y_1,Z=0]. \end{aligned}$$

Because each β_j is nonnegative, h_J is increasing in each of its arguments. Next, the positive dependence assumption applied to $[Y_{[J]}|Y_{[1]},\ldots,Y_{[J-1]},Z=0]$, together with Lemma 1 below, implies that h_{J-1} is increasing in each argument. Repeat this argument $k-2$ more times to conclude that h_1 is increasing in its only argument. It follows that

$$\begin{aligned} \text{LHS of (S2.10)} &= \mathrm{E}[h_1(Y_{[1]})|Y_{[1]} \leq y_1^*, Z=0] \leq h_1(y_1^*) \\ &\leq \mathrm{E}[h_1(Y_{[1]})|Y_{[1]} > y_1^*, Z=0] = \text{RHS of (S2.10)}, \end{aligned}$$

and the proof is complete.

Lemma 1. *Let P_1 and P_2 be probability measures on $\mathbb{R}$ and let $h:\mathbb{R}\to[0,\infty)$ be increasing. If P_1 is stochastically smaller than P_2, then $P_1h \leq P_2h$.*

Proof. The result is immediate for $h = 1_{(x,\infty)}$. A limiting argument can be used to prove the result for $h = 1_{[x,\infty)}$ and hence for $h=\sum_{k=1}^K c_k 1_{[x_k,\infty)}$ with $c_k \geq 0$ for all k. A left-continuous increasing function h can be approximated by the sequence of functions

$$h_m = \sum_{k=0}^{2^{2m+1}} c_{m,k} 1_{[x_{m,k},\infty)},$$

where $x_{m,k} = -2^m + k2^{-m}$, $c_{m,0} = h(x_{m,0})$, $c_{m,k} = h(x_{m,k}) - h(x_{m,k-1})$, $k=1,\ldots,2^{2m+1}$, $m \geq 1$. Since h is nonnegative and increasing, each $c_{m,k}$ is nonnegative so $P_1 h_m \leq P_2 h_m$ for every m. Further, $h_m \uparrow h$ because h is assumed left-continuous. Hence, by the monotone convergence theorem,

$$P_1 h = \lim_{m\to\infty} P_1 h_m \leq \lim_{m\to\infty} P_2 h_m = P_2 h.$$

Finally, consider an increasing function h that is not left-continuous, and denote by h_- its left-hand limit. For $m \geq 1$, let $D_m = \{x \in [-2^m, 2^m] : h(x) - h_-(x) > 2^{-m}\}$; then D_m is finite for every m. Let $x'_{m,k}$, $k \geq 0$, denote the distinct values in $D_m \cup \{-2^m + k2^{-m} : k = 0,\ldots,2^{2m+1}\}$ arranged in ascending order. Let $c'_{m,0} = h(x'_{m,0})$, $c'_{m,k} = h(x'_{m,k}) - h(x'_{m,k-1})$ for $k \geq 1$, and

$$h'_m = \sum_{k\geq 0} c'_{m,k} 1_{[x'_{m,k},\infty)}.$$

This definition ensures that $h'_m \uparrow h$ even though h is not left-continuous. The result again follows from the monotone convergence theorem. $\square$

Bibliography

Agresti, A. (1990). *Categorical Data Analysis.* Wiley, New York.

van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge University Press, Cambridge.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics.* Springer-Verlag, New York.