Supporting Information for Incorporating baseline covariates to validate surrogate endpoints with a constant biomarker under control arm

Web Appendix A

Imputation algorithm details

For drawing the model parameters and potential outcomes, we consider the full data likelihood as follows, first considering the case with no covariates in general:

$$(2\pi)^{-n/2}|\Sigma|^{-1/2}\exp\left(-\frac{1}{2}\begin{pmatrix} S(1)-\mu_{S1} \\ T(0)-\mu_{T0} \\ T(1)-\mu_{T1} \end{pmatrix}^T \Sigma^{-1}\begin{pmatrix} S(1)-\mu_{S1} \\ T(0)-\mu_{T0} \\ T(1)-\mu_{T1} \end{pmatrix}\right)$$

As the conjugate, posterior distributions can be written in closed-form for most identified parameters, we use Markov Chain Monte Carlo (MCMC) methods. Let subscript l denote the l^{th} iteration of the Gibbs sampler, μ generally denote the set of mean parameters in the model and QRQ denote the set of variance and correlation parameters in the model.

$$T_i(0)\left|\begin{matrix} S_i(1), T_i(1), \\ \mu^{l-1}, R^{l-1}, Q^{l-1} \end{matrix}\right. \qquad \begin{matrix} S_i(1) \\ T_i(1) \end{matrix}\left|\begin{matrix} T_i(0), \mu^{l-1}, \\ R^{l-1}, Q^{l-1} \end{matrix}\right.$$

During each iteration of the MCMC, we impute the missing potential outcome under treatment $z=0$ for those who we observed an outcome under treatment $z=1$ and vice versa. After the outcomes (denoted in general as Y) are imputed, we draw the mean, variance, and correlation parameters respectively:

For coefficients, $\mu|\cdot \sim$ Matrix Normal $\big((X^TX+\Lambda_0)(X^TY), X^TX+\Lambda_0, \Sigma\big)$ for prior matrix Λ_0

$\sigma_Y|\cdot \propto \sigma_Y^{-n}\exp(-\frac{1}{2}\Sigma_{i=1}^n(Y_i-\mu)(QRQ)^{-1}(Y_i-\mu)^T)$

For $j = T, 10, 11,\ \ \rho_j|\cdot \propto |R|^{-n/2}\exp(-\frac{1}{2}\Sigma_{i=1}^n(Y_i-\mu)(QRQ)^{-1}(Y_i-\mu)^T)$ for Uniform prior and within bounds determined by positive definiteness so the determinant is positive: $1-\rho_T^2+\rho_{10}^2 > \rho_{11}^2$.

The condition that $\rho_{10} = \rho_T \times \rho_{11}$ is found by either setting a term in the precision matrix to 0 or solving for the covariance of $S(1), T(0)|T(1)$ and setting this equal to 0. Since the outcomes are multivariate normal, the conditional covariance is equal to

$$\begin{pmatrix} \sigma_{S1}^2 - \sigma_{S1}^2\rho_{11} & \rho_{10}\sigma_{T0}\sigma_{S1} - \sigma_{S1}\rho_{11}\sigma_{T0}\rho_T \\ \rho_{10}\sigma_{T0}\sigma_{S1} - \sigma_{T0}\rho_T\sigma_{S1}\rho_{11} & \sigma_{T0}^2 - \sigma_{T0}^2\rho_T^2 \end{pmatrix}.$$

Then we solve $\rho_{10}\sigma_{S1}\sigma_{T0} - \sigma_{S1}\rho_{11}\sigma_{T0}\rho_T = 0$. The same process holds for the conditional model for $\theta_{10} = \theta_T \times \theta_{11}$.

Marginalization: Note the integral in equation 3 can be replaced with summation over the support of X for discrete covariates. Let $F_n(x)$ be the empirical distribution function of x, so for each value of s, equation 3 can be approximated by

$$\int_X \frac{(\gamma_{0,C} + \gamma_{1,C}s)\frac{1}{\epsilon_{S1}\sqrt{2\pi}}\exp(-\frac{1}{2\epsilon_{S1}^2}(s - (1\ x)^T(\omega_1\ \omega_2)^2)dF_n(x)}{f(s)}$$

Let $X_k, k = 1, ..., n$ denote the discrete values of X and let γ_{0,C_k} be the value of $\gamma_{0,C}$ at X_k. Then for a fixed s, we calculate $w_k = \exp(-\frac{1}{2\hat{\sigma}_{S1}^2}(s - (1\ X_k)^T(\hat{\omega_1}\ \hat{\omega_2}))^2)$ for each value of X_k, rescale w_k such that $\Sigma_k w_k = 1$, and calculate $\Sigma_x(\gamma_{0,C_k} + \gamma_{1,C}s)w_k$. This can be solved in closed form when X is normally distributed.

Web Appendix B

The four trivariate normal distributions and corresponding surrogacy quantities in the main text are derived from the four-dimensional, joint normal distribution:

$$\begin{pmatrix} S(1) \\ T(0) \\ T(1) \\ X \end{pmatrix} \sim N\left(\begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \end{pmatrix}, \begin{pmatrix} \sigma_{S1}^2 & \sigma_{S1}\sigma_{T0}\rho_{10} & \sigma_{S1}\sigma_{T1}\rho_{11} & \sigma_{S1}\sigma_X\rho_{1X} \\ & \sigma_{T0}^2 & \sigma_{T0}\sigma_{T1}\rho_T & \sigma_{T0}\sigma_X\rho_{X0} \\ & & \sigma_{T1}^2 & \sigma_{T1}\sigma_X\rho_{X1} \\ & & & \sigma_X^2 \end{pmatrix}\right)$$

1. $$\begin{pmatrix} S(1) \\ T(0) \\ T(1) \end{pmatrix} \sim N\left(\begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{pmatrix}, \begin{pmatrix} \sigma_{S1}^2 & \rho_{10}\sigma_{S1}\sigma_{T0} & \rho_{11}\sigma_{S1}\sigma_{T1} \\ & \sigma_{T0}^2 & \rho_T\sigma_{T0}\sigma_{T1} \\ & & \sigma_{T1}^2 \end{pmatrix}\right)$$

2. $$\left(\begin{array}{c|c} S(1) & X \\ T(0) & \\ T(1) & \end{array}\right) \sim N\left(\begin{pmatrix} \omega_1 + \omega_2 X \\ \omega_3 + \omega_4 X \\ \omega_5 + \omega_6 X \end{pmatrix}, \begin{pmatrix} \epsilon_{S1}^2 & \epsilon_{S1}\epsilon_{T0}\theta_{10} & \epsilon_{S1}\epsilon_{T1}\theta_{11} \\ & \epsilon_{T0}^2 & \epsilon_{T0}\epsilon_{T1}\theta_T \\ & & \epsilon_3^2 \end{pmatrix}\right)$$

$\gamma_{1,OC} = \frac{\epsilon_3\theta_{11} - \epsilon_2\theta_{10}}{\epsilon_1}$, $\gamma_{0,OC} = (\omega_5 + \omega_6 X - \omega_3 - \omega_4 X) - \gamma_{1,OC}(\omega_1 + \omega_2 X)$

$\gamma_{1,O} = \frac{\sigma_{T1}\rho_{11} - \sigma_{T0}\rho_{10}}{\sigma_{S1}}$, $\gamma_{0,O} = (\delta_3 - \delta_2) - \gamma_{1,O}\delta_1$

3. $$\begin{pmatrix} S(1) \\ T^D(0) \\ T^D(1) \end{pmatrix} \sim N\left(\begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix}, \begin{pmatrix} \tau_{S1}^2 & \tau_{S1}\tau_{T0}\pi_{10} & \tau_{S1}\tau_{T1}\pi_{11} \\ & \tau_{T0}^2 & \tau_{T0}\tau_{T1}\pi_T \\ & & \tau_{T1}^2 \end{pmatrix}\right)$$

4. $$\left(\begin{array}{c|c} S(1) & X \\ T^D(0) & \\ T^D(1) & \end{array}\right) \sim N\left(\begin{pmatrix} \phi_1 + \phi_2 X \\ \phi_3 + \phi_4 X \\ \phi_5 + \phi_6 X \end{pmatrix}, \begin{pmatrix} \xi_{S1}^2 & \xi_{S1}\xi_{T0}\psi_{10} & \xi_{S1}\xi_{T1}\psi_{11} \\ & \xi_{T0}^2 & \xi_{T0}\xi_{T1}\psi_T \\ & & \xi_{T1}^2 \end{pmatrix}\right)$$

$\gamma_{1,DC} = \frac{\xi_{T1}\psi_{11} - \xi_{T0}\psi_{10}}{\xi_{S1}}$, $\gamma_{0,DC} = (\phi_5 + \phi_6 X - \phi_3 - \phi_4 X) - \gamma_{1,DC}(\phi_1 + \phi_2 X)$

$\gamma_{1,D} = \frac{\tau_3\pi_{11} - \tau_2\pi_{10}}{\tau_1}$, $\gamma_{0,D} = (\eta_3 - \eta_2) - \gamma_{1,D}\eta_1$

The conditional quantities $\gamma_{1,OC}$ and $\gamma_{0,OC}$ can be calculated only from model 2, which can be subsequently marginalized over. The marginal quantities $\gamma_{1,O}$ and $\gamma_{0,O}$ are the same for models 1 and 2. Similarly, $\gamma_{1,DC}$ and $\gamma_{0,DC}$ from model 4 can be marginalized over to calculate the same $\gamma_{1,D}$ and $\gamma_{0,D}$ as in model 3.

Web Table 1

Setting	Fit Conditional Independence Assumption	$\gamma_{0,M}$ Est	Bias	SE	SD	$\gamma_{1,M}$ Est	Bias	SE	SD	True Distance from Cond. Ind. on fit scale
3A	$T^D(0) \perp S(1) \vert T^D(1)$	0.057	0.057	0.398	0.256	0.520	-0.030	0.175	0.093	0.002
3A	None	0.067	0.127	0.956	0.255	0.515	-0.065	0.468	0.093	0.002
4A	$T^D(0) \perp S(1) \vert T(1), X$	0.058	0.118	0.422	0.478	0.527	-0.053	0.177	0.088	0.000
4A	None	0.091	0.151	0.934	0.483	0.519	-0.061	0.455	0.090	0.000
1A	$T(0) \perp S(1) \vert T(1)$	0.068	0.068	0.414	0.303	0.516	-0.034	0.181	0.111	-0.098
1A	None	0.078	0.078	1.122	0.301	0.512	-0.038	0.552	0.110	-0.098
2A	$T(0) \perp S(1) \vert T(1), X$	0.044	0.044	0.426	0.308	0.525	-0.025	0.179	0.091	0.000
2A	None	0.080	0.080	0.989	0.493	0.522	-0.028	0.478	0.092	0.000
3B	$T^D(0) \perp S(1) \vert T^D(1)$	0.056	0.056	0.413	0.293	0.518	-0.032	0.181	0.108	-0.101
3B	None	0.066	0.066	1.117	0.291	0.514	-0.036	0.549	0.107	-0.101
4B	$T^D(0) \perp S(1) \vert T^D(1), X$	0.052	0.052	0.424	0.479	0.522	-0.028	0.180	0.084	0.000
4B	None	0.099	0.099	0.978	0.480	0.501	-0.049	0.471	0.085	0.000
1B	$T(0) \perp S(1) \vert T(1)$	0.058	0.058	0.398	0.256	0.519	-0.031	0.175	0.093	0.000
1B	None	0.067	0.067	0.955	0.255	0.516	-0.034	0.468	0.093	0.000
2B	$T(0) \perp S(1) \vert T(1), X$	0.058	0.058	0.422	0.478	0.527	-0.023	0.177	0.088	0.000
2B	None	0.082	0.082	0.998	0.479	0.517	-0.033	0.477	0.088	0.000
3C	$T^D(0) \perp S(1) \vert T^D(1)$	-0.974	0.046	0.399	0.286	0.534	-0.026	0.172	0.110	0.003
3C	None	-0.963	0.057	1.015	0.284	0.530	-0.030	0.497	0.109	0.003
4C	$T^D(0) \perp S(1) \vert T^D(1), X$	-0.944	0.076	0.422	0.478	0.529	-0.031	0.177	0.115	0.000
4C	None	-0.904	0.116	1.002	0.483	0.506	-0.054	0.478	0.109	0.000
1C	$T(0) \perp S(1) \vert T(1)$	-0.905	0.095	0.418	0.266	0.501	-0.049	0.188	0.093	0.003
1C	None	-0.895	0.105	1.056	0.265	0.497	-0.053	0.520	0.092	0.003
2C	$T(0) \perp S(1) \vert T(1), X$	-0.944	0.056	0.422	0.478	0.529	-0.021	0.177	0.115	0.000
2C	None	-0.919	0.081	0.998	0.479	0.519	-0.031	0.477	0.116	0.000
3D	$T^D(0) \perp S(1) \vert T^D(1)$	-1.288	0.045	0.324	0.304	0.191	-0.029	0.130	0.094	0.000
3D	None	-1.288	0.045	1.567	0.301	0.193	-0.027	0.779	0.092	0.000
4D	$T^D(0) \perp S(1) \vert T^D(1), X$	-1.370	-0.037	0.292	0.513	0.231	0.011	0.103	0.136	0.000
4D	None	-1.321	0.012	1.145	0.513	0.209	-0.011	0.555	0.136	0.000
1D	$T(0) \perp S(1) \vert T(1)$	-1.257	0.093	0.353	0.360	0.177	-0.043	0.145	0.100	-0.094
1D	None	-1.261	0.089	1.996	0.357	0.181	-0.039	0.995	0.097	-0.094
2D	$T(0) \perp S(1) \vert T(1), X$	-1.371	-0.021	0.291	0.512	0.232	0.012	0.099	0.189	0.000
2D	None	-1.343	0.007	1.167	0.511	0.222	0.002	0.563	0.189	0.000
3E	$T^D(0) \perp S(1) \vert T^D(1)$	1.009	0.059	0.210	0.207	0.181	-0.129	0.116	0.105	0.415
3E	None	1.011	0.061	0.359	0.207	0.182	-0.128	0.564	0.102	0.415
4E	$T^D(0) \perp S(1) \vert T^D(1), X$	1.011	0.061	0.212	0.466	0.178	-0.132	0.118	0.106	0.414
4E	None	1.002	0.052	0.364	0.466	0.195	-0.115	0.565	0.104	0.414
1E	$T(0) \perp S(1) \vert T(1)$	1.012	0.063	0.219	0.233	0.181	-0.119	0.128	0.118	0.414
1E	None	1.014	0.065	0.392	0.232	0.183	-0.117	0.630	0.115	0.414
2E	$T(0) \perp S(1) \vert T(1), X$	1.011	0.062	0.212	0.466	0.178	-0.122	0.118	0.106	0.414
2E	None	1.002	0.053	0.364	0.466	0.195	-0.105	0.565	0.104	0.414
1F	$T(0) \perp S(1) \vert T(1)$	1.360	0.050	0.431	0.376	0.553	-0.027	0.187	0.148	0.183
1F	None	1.387	0.077	1.163	0.375	0.539	-0.041	0.568	0.148	0.183
2F	$T(0) \perp S(1) \vert T(1), X$	1.485	0.175	0.402	0.323	0.501	-0.079	0.170	0.222	0.000
2F	None	1.501	0.191	0.977	0.322	0.498	-0.082	0.475	0.218	0.000

Table 1: Simulation results demonstrating different definitions of the endpoint (settings 1-4) and different generating parameter values ($A - F$). Estimates and bias are shown here since true generating values differ across scenarios. These are shown in the main text Table 1.

Web Figure 1

Figure 1: Simulation results and sensitivity analysis of data example results over different values of θ_T using the observed data method and conditional independence assumption. The models are fit by either fixing the value of θ_T or placing the prior distribution over the parameter that we consider in the main text.

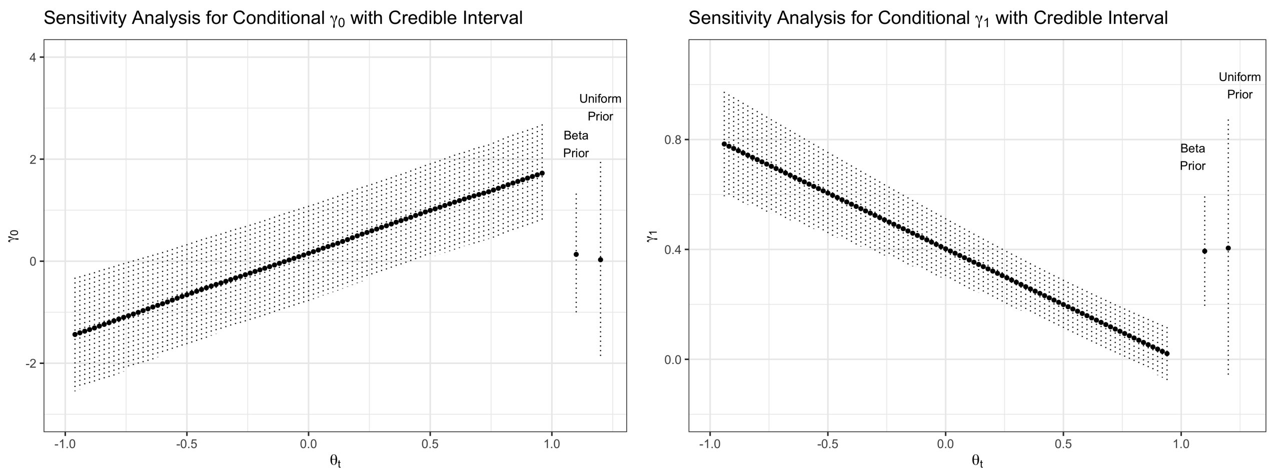
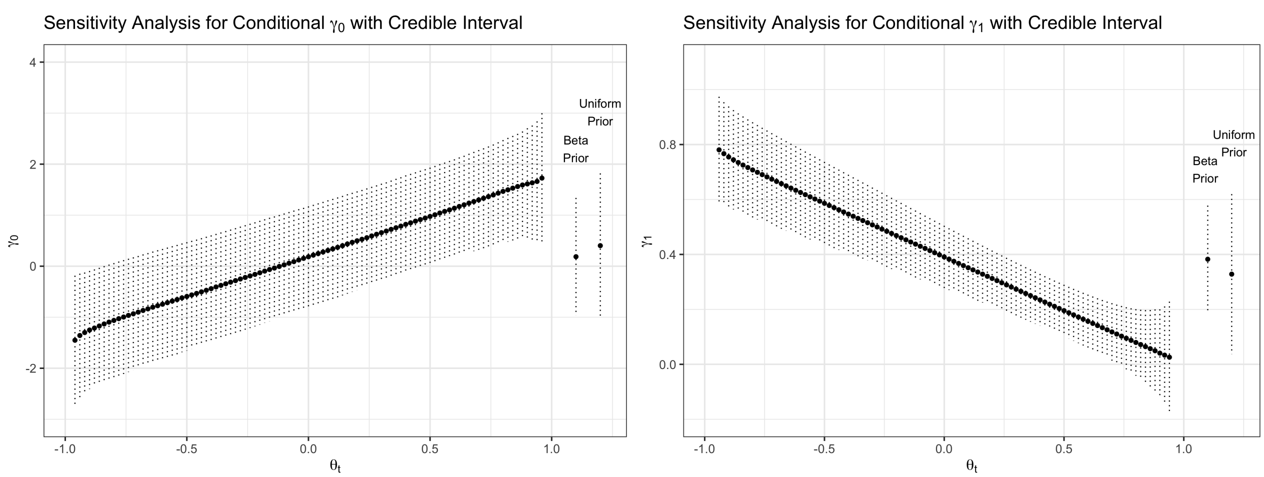


Figure 2: Simulation results and sensitivity analysis of data example results over different values of θ_T using the imputation method and conditional independence assumption.



The actual priors we used are Beta(2.7, 5) and Uniform(-1, 1). This is compared to fixing θ_T at possible values one at a time and repeating the process. These plots show that we would say $\gamma_0 > 0$ if $\theta_T > 0.48$, and $\gamma_1 > 0$ if $\theta_T < 0.75$ approximately. Also, we see that both the observed data only and imputation methods return nearly the same results (perhaps except at the tails where convergence at the boundaries of the parameter space may play a role).

Web Table 2

Setting	Fit Conditional Independence Assumption	True Distribution	$\gamma_{0,M}$ Estimate	Bias	SE	SD	$\gamma_{1,M}$ Estimate	Bias	SE	SD
2A	$T(0) \perp S(1)\|T(1), X$	T	0.067	0.127	0.404	0.308	0.512	-0.068	0.171	0.124
2B	$T(0) \perp S(1)\|T(1), X$	T	0.067	0.067	0.404	0.308	0.512	-0.038	0.171	0.124
2C	$T(0) \perp S(1)\|T(1), X$	T	-0.937	0.063	0.402	0.312	0.513	-0.037	0.170	0.123
2D	$T(0) \perp S(1)\|T(1), X$	T	-1.371	-0.021	0.285	0.378	0.229	0.009	0.097	0.150
2E	$T(0) \perp S(1)\|T(1), X$	T	0.994	0.044	0.212	0.218	0.192	-0.118	0.090	0.153
2F	$T(0) \perp S(1)\|T(1), X$	T	1.452	0.140	0.416	0.339	0.508	-0.068	0.174	0.118
2A	$T(0) \perp S(1)\|T(1), X$	Gamma	0.133	0.193	0.396	0.252	0.481	-0.099	0.165	0.138
2B	$T(0) \perp S(1)\|T(1), X$	Gamma	0.133	0.133	0.397	0.252	0.485	-0.065	0.166	0.137
2C	$T(0) \perp S(1)\|T(1), X$	Gamma	-0.868	0.132	0.396	0.258	0.481	-0.069	0.165	0.138
2D	$T(0) \perp S(1)\|T(1), X$	Gamma	-1.311	0.039	0.280	0.308	0.200	-0.020	0.093	0.128
2E	$T(0) \perp S(1)\|T(1), X$	Gamma	1.015	0.065	0.210	0.194	0.164	-0.136	0.084	0.127
2F	$T(0) \perp S(1)\|T(1), X$	Gamma	1.475	0.163	0.401	0.306	0.488	-0.088	0.168	0.149

Table 2: Simulation results demonstrating effect of non-normal distributions when fitting conditional models that assume normality. Results are shown on the original scale and can be compared to Web Table 1.

Web Appendix C

Generative parameter values for plausible clinical trial data example: We generate the $S(1), T(0), T(1)$ outcomes at one time point and effects of age, A, and baseline, X.

$$\begin{pmatrix} X \\ A \end{pmatrix} \sim N\left(\begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_1 \\ & \sigma_2^2 \end{pmatrix}\right)$$

We chose values for the means and variances: $\delta_1 = 24, \delta_2 = 5$, $\sigma_1^2 = 1, \sigma_2^2 = 0.65$, $\rho_1 = 0.7$. We expect there to be a quadratic effect of age, so we generate

$$\left(\begin{array}{c|c} S(1) & X, \\ T(0) & A, \\ T(1) & A^2 \end{array}\right) \sim N\left(\begin{pmatrix} \phi_1 + \phi_2 X + \phi_3 A \\ \phi_4 + \phi_5 X + \phi_6 A + \phi_7 A^2 \\ \phi_8 + \phi_9 X + \phi_{10} A + \phi_{11} A^2 \end{pmatrix}, \begin{pmatrix} \xi_{S1}^2 & \xi_{S1}\xi_{T0}\psi_{10} & \xi_{S1}\xi_{T1}\psi_{11} \\ & \xi_{T0}^2 & \xi_{T0}\xi_{T1}\psi_T \\ & & \xi_{T1}^2 \end{pmatrix}\right)$$

$\phi_1 = 3.8, \phi_2 = 0, \phi_3 = 0$, $\phi_4 = 0.1, \phi_5 = 1.14, \phi_6 = 0.45, \phi_7 = -0.2$, $\phi_8 = 10.6, \phi_9 = 1.1$, $\phi_{10} = -1.15, \phi_{11} = -0.2, \xi_{S1}^2 = 1, \xi_{T0}^2 = 0.35, \xi_{T1}^2 = 0.35$, $\psi_{10} = 0.013, \psi_{11} = 0.65, \psi_T = 0.02$