Supporting Information for

Using a Surrogate with Heterogeneous Utility to Test for a Treatment Effect

by Parast, Cai, and Tian

Appendix A

Discrete Example

Let Y denote the primary outcome and S denote the surrogate marker. We use potential outcomes notation where each person has a potential $\{Y^{(1)}, Y^{(0)}, S^{(1)}, S^{(0)}\}$ where $Y^{(g)}$ and $S^{(g)}$ are the outcome and surrogate when the patient receives treatment g. Our main quantity of interest is the treatment effect on the primary outcome quantified as $\Delta \equiv E(Y^{(1)} - Y^{(0)}) = E(Y^{(1)}) - E(Y^{(0)})$. The earlier treatment effect incorporating S information is defined in the main text as

$$\Delta_P = \int \mu_0^p(s) dF^{(1)}(s) - \int \mu_0^p(s) dF^{(0)}(s) \tag{1}$$

where $\mu_0^p(s) \equiv E(Y^{(0p)} = y|S^{(0p)} = s)$. In this example, we will have heterogeneity in the utility of the surrogate with respect to gender. Consider our prior study, which we refer to as Study A in this example, and is shown in Figure A1. The Study A sample is 50% female and 50% male. For all individuals, $(S^{(1)}, S^{(0)})$ are independent of gender, and $\left\{E(S^{(1)}), E(S^{(0)})\right\} = (10, 5)$. For females, $E(Y^{(1)} \mid S^{(1)} = s) = 3 + 5s$ and $E(Y^{(0)} \mid S^{(0)} = s) = 1 + 3S$. It can be shown that for females, $\Delta = 53 - 16 = 37$ and $\Delta_P = 15$. The proportion of the treatment effect on the primary outcome that is explained by the surrogate among females is thus 15/37=41%, which would not be considered as a strong surrogacy. For males, $E(Y^{(1)} \mid S^{(1)} = s) = 15s$ and $E(Y^{(0)} \mid S^{(0)} = s) = 14.8S$. It can be shown that for males, $(\Delta, \Delta_P) = (76, 74)$ and the proportion explained by the surrogate marker is 97% among males, representing strong surrogacy.

To calculate Δ_P for a future study, let's consider the conditional mean that is central to this calculation, $\mu_0^p(s) = E(Y^{(0p)} = y|S^{(0p)} = s)$ where the superscript p indicates that this is referring to the prior study, i.e., study A. In this example, this would be $\mu_0^p(s) = 0.5 \times (1+3s) + 0.5 \times 14.8s = 8.9s + 0.5$. Now assume our current study is Study B shown in Figure A1 which is 95% female and 5% male. Importantly, the joint distributions of $(Y^{(1)}, Y^{(0)}, S^{(1)}, S^{(0)})$ in males and females remain as described above; the only difference is the distribution of gender. The treatment effect, Δ in this new study is $0.95 \times 37 + 0.05 \times 76 = 38.95$. If one were to calculate Δ_P not accounting for this known heterogeneity in the utility of the surrogate, the quantity obtained would be $\Delta_P = 8.9 \times 10 + 0.5 - 8.9 \times 5 - 0.5 = 44.5$, recalling that $E(S^{(1)}) = 10$ and $E(S^{(0)}) = 5$ for all individuals in both studies. However, using our proposed approach which does account for heterogeneity, we use Δ_H as the earlier treatment effect, defined in the main text as:

$$\Delta_H = \int \mu_0^p(s, w) dF^{(1)}(s, w) - \int \mu_0^p(s, w) dF^{(0)}(s, w).$$

Thus, $\Delta_H = 95\% \times (1+3\times 10) + 5\% \times (14.8 \times 10) - 95\% \times (1+3\times 5) - 5\% \times (14.8 \times 5) = 17.95$. Therefore $\Delta_H < \Delta < \Delta_P$ and Δ_P no longer retains the property of providing a lower bound on the treatment effect on Y.

Now we consider a study, labeled Study C in Figure A1, which is 95% males and 5% females. Using similar calculations, we can show that $\Delta = 74.05$, $\Delta_P = 44.05$ and $\Delta_H = 71.05$. Thus, in this case, Δ_H will provide better lower bound for Δ and the test based on Δ_H is expected to be more powerful than that based on Δ_P. The discrete case, as illustrated in this example, is relatively straightforward in terms of how to go about calculating the needed quantities separately by group and appropriately accounting for the different distribution in the new study. The continuous baseline covariate case, however, is more complex, and our Appendix C presents an example such that even if the prior and current studies have the same distribution for covariates, Δ_P may still fail to be a valid lower bound for Δ.

Appendix B

As noted in this text, Assumptions $(C1)-(C3)$ together guarantee that $E(Y^{(1)} \mid W=w) \geq E(Y^{(0)} \mid W=w)$, for all w in the support of W. This result is due to the derivation:

$$
\begin{aligned}
\Delta(w) =& E(Y^{(1)} \mid W=w) - E(Y^{(0)} \mid W=w) \\
=& \int_s E(Y^{(1)} \mid S^{(1)}=s, W=w) dF^{(1)}(s \mid w) - \int_s E(Y^{(0)} \mid S^{(0)}=s, W=w) dF^{(0)}(s \mid w) \\
\geq& \int_s E(Y^{(0)} \mid S^{(0)}=s, W=w) dF_1(s \mid w) - \int_s E(Y^{(0)} \mid S^{(0)}=s, W=w) dF^{(0)}(s \mid w) \\
=& \int_s E(Y^{(0)} \mid S^{(0)}=s, W=w) d\left\{F^{(1)}(s \mid w) - F^{(0)}(s \mid w)\right\} \\
=& \int_s \left\{F^{(0)}(s \mid w) - F^{(1)}(s \mid w)\right\} \frac{\partial E(Y^{(0)} \mid S^{(0)}=s, W=w)}{\partial s} ds \geq 0,
\end{aligned}
$$

where $F^{(g)}(s \mid w) = P(S^{(g)} \leq s | W=w), g=0,1$. That is, while treatment effect heterogeneity is allowed, the directions of the conditional average treatment effect among subgroups of patients with $W=w$ need to be consistent. One important implication is that under the null $H_0 : \Delta = E\{\Delta(W)\} = 0$, i.e., no average treatment effect, the conditional average treatment effect $\Delta(w) = 0$ for all w as well. Furthermore, from the derivation, it is clear that $\Delta(w) = 0$ if and only if both

1. $F^{(1)}(s \mid w) = F^{(0)}(s \mid w)$, i.e., $P(S^{(1)} > s | W=w) = P(S^{(0)} > s | W=w)$ and
2. $E(Y^{(1)} | S^{(1)}=s, W=w) = E(Y^{(0)} | S^{(0)}=s, W=w)$.

Specifically, $\Delta(w) = 0$ implies that there is no treatment effect on the distribution of the surrogate marker in the subgroup of patients with $W=w$. In summary, under Assumptions (C1)-(C3)

$$\Delta = 0 \Rightarrow \Delta(w) = 0 \Rightarrow S^{(1)} \mid W=w \sim S^{(0)} \mid W=w.$$

This relationship allows us to test the common null $H_0 : \Delta = 0$ via testing a seemingly more restrictive null that $S^{(1)} \mid W = w \sim S^{(0)} \mid W = w$, for all w in the support of W.

For $(C2)$ and $(C3)$, if the primary outcome or surrogate are such that lower values are "better", one can simply define the outcome/surrogate as $-X$ where X is the initial value.

Assumptions $(C5) - (C6)$ are not required for the validity of the testing procedure proposed in the next section in that the p-value under the null follows a uniform distribution even without them, but it allows us to estimate a lower bound of the average treatment effect, Δ, and construct the corresponding test statistic.

Under the following additional assumptions:

(C7) $Y^{(1)} \perp S^{(0)} | S^{(1)}, W$ and $Y^{(0)} \perp S^{(1)} | S^{(0)}, W$;

(C8) $Y^{(1p)} \perp S^{(0p)} | S^{(1p)}, W^p$ and $Y^{(0p)} \perp S^{(1p)} | S^{(0p)}, W^p$,

the treatment effect on the surrogate marker defined in Section **??** and on the primary outcome can be interpreted within a causal framework: the proposed test statistic is an estimate of the portion of the treatment effect on the primary outcome attributable to the treatment effect on the surrogate marker. Otherwise, the proposed treatment effect on the surrogate marker can always serve as a lower bound for the average treatment effect on Y and can be used in practice without assuming them.

To summarize, Assumptions $(C1) - (C4)$ are needed for the validity of the proposed testing procedure, Assumptions $(C5) - (C6)$ allow us to interpret the test statistic based on he surrogate marker and baseline covariate only as a "conservative" estimator (or a lower bound) of the average treatment effect on the primary outcome, and causal interpretation of the lower is possible under additional assumptions $(C7) - (C8)$.

Appendix C

To estimate Δ using the primary outcome (gold standard) we use $\widehat{\Delta} = n_1^{-1}\sum_{i=1}^{n_1} Y_{1i} - n_0^{-1}\sum_{i=1}^{n_0} Y_{0i}$ and conduct a t-test to test $H_0 : \Delta = 0$.

To estimate $\widetilde{\Delta}_P$, we use the nonparametric estimation approach of Parast et al. (2019) by estimating $\mu_0^p(s)$ as

$$\widehat{\mu}_0^p(s) = \frac{\sum_{i=1}^{n_0^p} K_{h_4}(S_{0i}^p - s)Y_{0i}^p}{\sum_{i=1}^{n_0^p} K_{h_4}(S_{0i}^p - s)},$$

and then estimate $\widetilde{\Delta}_P$ as

$$\widehat{\Delta}_P = n_1^{-1}\sum_{i=1}^{n_1} \widehat{\mu}_0^p(S_{1i}) - n_0^{-1}\sum_{i=1}^{n_0} \widehat{\mu}_0^p(S_{0i}).$$

Note that this estimate only uses S data from the current study (no Y data from the current study) and S, Y data from the previous study in group $Z = 0$ only. To obtain an estimate for the standard error of $\widehat{\Delta}_P$, σ_P, we simply take the empirical standard deviation of the transformed surrogate i.e., let $\widetilde{Y}_{gi} = \widehat{\mu}_0^p(S_{gi})$, and then $\widehat{\sigma}_P = \widehat{var}(\widetilde{Y}_{1i})/n_1 + \widehat{var}(\widetilde{Y}_{0i})/n_0$ where $\widehat{var}$ indicates the empirical variance. This alternative testing procedure would then use the test statistic $Z_P = \widehat{\Delta}_P/\widehat{\sigma}_P$ and reject the null hypothesis when $|Z_P| > \Phi^{-1}(1-\alpha/2)$.

Importantly, one may also consider simply using the surrogate markers measured in the current study and define $\Delta_M = E(S^{(1)}) - E(S^{(0)})$ and conduct a t-test of $H_{0M} : \Delta_M = 0$. The disadvantage of this approach is that there is no way to relate Δ_M and Δ i.e., the estimate of Δ_M does not give any helpful information about the magnitude of Δ. In addition, this approach does not take advantage of information from the previous study nor does it account for heterogeneity in the utility of the surrogate marker. For these reasons, we do not compare our approach to this test.

Appendix D

Our proposed estimator for $\widetilde{\Delta}_H$ is

$$\widehat{\Delta}_H = \frac{1}{n}\left\{\sum_{i=1}^{n_0}\left[\widehat{m}_1(W_{0i};\widehat{\mu}_0^p) - \widehat{m}_0(W_{0i};\widehat{\mu}_0^p)\right] + \sum_{i=1}^{n_1}\left[\widehat{m}_1(W_{1i};\widehat{\mu}_0^p) - \widehat{m}_0(W_{1i};\widehat{\mu}_0^p)\right]\right\}.$$

Let $\widetilde{\mu}_g = E\left\{\widehat{\mu}_0^p(S^{(g)}, W) \mid \widehat{\mu}_0^p\right\}, g = 0, 1$. It is obvious that $\widetilde{\Delta}_H = \widetilde{\mu}_1 - \widetilde{\mu}_0$. Also, let $m_g(w;\widehat{\mu}_0^p) = E\left\{\widehat{\mu}_0^p(S^{(g)}, W) \mid W = w\right\}$.

In this section, we only consider the randomness in the current study, i.e., the probability measure is conditional on $\widehat{\mu}_0^p(\cdot,\cdot)$. Now consider the centered term

$$\begin{aligned}&\frac{1}{n}\sum_{g=0}^{1}\sum_{j=1}^{n_g}\widehat{m}_1(W_{gj};\widehat{\mu}_0^p) - \widetilde{\mu}_1\\ =&\frac{1}{n}\sum_{g=0}^{1}\sum_{j=1}^{n_g}\left[n_1^{-1}\sum_{i=1}^{n_1}\frac{K_h(W_{1i} - W_{gj})\widetilde{S}_{1i}}{\widehat{f}_1(W_{gj})}\right] - \widetilde{\mu}_1,\end{aligned}$$

which is

$$
\begin{aligned}
&\frac{1}{nn_1}\sum_{j=1}^{n_0}\sum_{i=1}^{n_1}\frac{K_h(W_{1i}-W_{0j})\widetilde{S}_{1i}}{\widehat{f}_1(W_{0j})}+\frac{1}{n}\sum_{i=1}^{n_1}\left[\frac{1}{n_1}\sum_{j=1}^{n_1}\frac{K_h(W_{1i}-W_{1j})}{\widehat{f}_1(W_{1j})}\right]\widetilde{S}_{1i}-\widetilde{\mu}_1\\
=&\frac{1}{nn_1}\sum_{j=1}^{n_0}\sum_{i=1}^{n_1}\frac{K_h(W_{1i}-W_{0j})\widetilde{S}_{1i}}{\widehat{f}_1(W_{0j})}+\frac{1}{n}\sum_{i=1}^{n_1}\left[\frac{1}{n_1}\sum_{j=1}^{n_1}K_h(W_{1i}-W_{1j})\right]\frac{\widetilde{S}_{1i}}{\widehat{f}_1(W_{1i})}-\widetilde{\mu}_1+O_p(h^2)\\
=&\frac{n_0}{nn_1}\sum_{i=1}^{n_1}\frac{\widehat{f}_0(W_{1i})}{\widehat{f}_1(W_{1i})}\widetilde{S}_{1i}+\frac{1}{n}\sum_{i=1}^{n_1}\widetilde{S}_{1i}-\widetilde{\mu}_1+O_p(h^2)\\
=&\frac{1}{n_1}\sum_{i=1}^{n_1}(\widetilde{S}_{1i}-\widetilde{\mu}_1)+\frac{n_0}{nn_1}\sum_{i=1}^{n_1}\frac{\widehat{f}_0(W_{1i})-\widehat{f}_1(W_{1i})}{\widehat{f}_1(W_{1i})}\widetilde{S}_{1i}+O_p(h^2)\\
=&\frac{1}{n_1}\sum_{i=1}^{n_1}(\widetilde{S}_{1i}-\widetilde{\mu}_1)+\frac{n_0}{nn_1}\sum_{i=1}^{n_1}\left[\frac{1}{n_0}\sum_{j=1}^{n_0}K_h(W_{0j}-W_{1i})-\frac{1}{n_1}\sum_{j=1}^{n_1}K_h(W_{1j}-W_{1i})\right]\frac{\widetilde{S}_{1i}}{f_1(W_{1i})}+O_p(h^2)\\
=&\frac{1}{n_1}\sum_{i=1}^{n_1}(\widetilde{S}_{1i}-\widetilde{\mu}_1)+\pi_0\left[\frac{1}{n_0}\sum_{i=1}^{n_0}\widehat{m}_1(W_{0i};\widehat{\mu}_0^p)-\frac{1}{n_1}\sum_{i=1}^{n_1}\widehat{m}_1(W_{1i};\widehat{\mu}_0^p)\right]+O_p(h^2)\\
=&\frac{1}{n_1}\sum_{i=1}^{n_1}(\widetilde{S}_{1i}-\widetilde{\mu}_1)+\pi_0\left[\frac{1}{n_0}\sum_{i=1}^{n_0}m_1(W_{0i};\widehat{\mu}_0^p)-\frac{1}{n_1}\sum_{i=1}^{n_1}m_1(W_{1i};\widehat{\mu}_0^p)\right]\\
&+\pi_0\left[\frac{1}{n_0}\sum_{i=1}^{n_0}\left(\widehat{m}_1(W_{0i};\widehat{\mu}_0^p)-m_1(W_{0i};\widehat{\mu}_0^p)\right)-\frac{1}{n_1}\sum_{i=1}^{n_1}\left(\widehat{m}_1(W_{1i};\widehat{\mu}_0^p)-m_1(W_{1i};\widehat{\mu}_0^p)\right)\right]+O_p(h^2)
\end{aligned}
$$

where $\pi_g = n_g/n$ and $\widehat{f}_1(w)$ is the nonparametric estimator for the density function of W based on observations in treatment group 1. Now, consider the expansion

$$
\widehat{m}_1(w;\widehat{\mu}_0^p)-m_1(w;\widehat{\mu}_0^p)=\frac{1}{n_1}\sum_{i=1}^{n_1}K_h(W_{1i}-w)\left\{\widetilde{S}_{1i}-m_1(W_{1i};\widehat{\mu}_0^p)\right\}+O_p\left(h^2+\frac{\log(n_1)}{n_1h}\right)
$$

uniform in w. Therefore,

$$\begin{aligned}
&\frac{1}{n_0}\sum_{j=1}^{n_0}\{\widehat{m}_1(W_{0j};\widehat{\mu}_0^p)-m_1(W_{0j};\widehat{\mu}_0^p)\}\\
=&\frac{1}{n_1n_0}\sum_{j=1}^{n_0}\sum_{i=1}^{n_1}K_h(W_{1i}-W_{0j})\left\{\widetilde{S}_{1i}-m_1(W_{1i};\widehat{\mu}_0^p)\right\}+O_p\left(h^2+\frac{\log(n_1)}{n_1h}\right)\\
=&\frac{1}{n_1}\sum_{i=1}^{n_0}\widehat{f}_0(W_{1i})\left\{\widetilde{S}_{1i}-m_1(W_{1i};\widehat{\mu}_0^p)\right\}+O_p\left(h^2+\frac{\log(n_1)}{n_1h}\right)\\
=&\frac{1}{n_1}\sum_{i=1}^{n_0}f_0(W_{1i})\left\{\widetilde{S}_{1i}-m_1(W_{1i};\widehat{\mu}_0^p)\right\}+O_p\left(h^2+\frac{\log(n_1)}{n_1h}\right)+o_p\left(\frac{1}{\sqrt{n_1}}\right)
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\frac{1}{n_1}\sum_{i=1}^{n_1}(\widehat{m}_1(W_{1i};\widehat{\mu}_0^p)-m_1(W_{1i};\widehat{\mu}_0^p))\\
=&\frac{1}{n_1}\sum_{i=1}^{n_0}f_0(W_{1i})\left\{\widetilde{S}_{1i}-m_1(W_{1i};\widehat{\mu}_0^p)\right\}+O_p\left(h^2+\frac{\log(n_1)}{n_1h}\right)+o_p\left(\frac{1}{\sqrt{n_0}}\right),
\end{aligned}$$

and

$$\sqrt{n}\left[\frac{1}{n_0}\sum_{i=1}^{n_0}(\widehat{m}_1(W_{0i};\widehat{\mu}_0^p)-m_1(W_{0i};\widehat{\mu}_0^p))-\frac{1}{n_1}\sum_{i=1}^{n_1}(\widehat{m}_1(W_{1i};\widehat{\mu}_0^p)-m_1(W_{1i};\widehat{\mu}_0^p))\right] \quad (2)$$

$$=O_p\left(\sqrt{n_1}h^2+\frac{\log(n_1)}{\sqrt{n_1}h}\right)+o_p(1). \quad (3)$$

Therefore, when $h=O(n_1^{-\delta}),\delta\in(1/4,1/2)$, the right hand side of (3) becomes $o_p(1)$, and thus

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{g=0}^{1}\sum_{j=1}^{n_g}\widehat{m}_1(W_{gj};\widehat{\mu}_0^p)-\widetilde{\mu}_1\\
=&\frac{\sqrt{n}}{n_1}\sum_{i=1}^{n_1}(\widetilde{S}_{1i}-\widetilde{\mu}_1)+\pi_0\left[\frac{\sqrt{n}}{n_0}\sum_{j=1}^{n_0}m_1(W_{0j};\widehat{\mu}_0^p)-\frac{\sqrt{n}}{n_1}\sum_{j=1}^{n_1}m_1(W_{1j};\widehat{\mu}_0^p)\right]+o_p(1).
\end{aligned}$$

Finally, we have

$$
\begin{aligned}
&\sqrt{n}\left\{\widehat{\Delta}_H - \widetilde{\Delta}_H\right\} \\
=&\frac{\sqrt{n}}{n_1}\sum_{i=1}^{n_1}(\widetilde{S}_{1i} - \widetilde{\mu}_1) + \pi_0\left[\frac{\sqrt{n}}{n_0}\sum_{i=1}^{n_0} m_1(W_{0i};\widehat{\mu}_0^p) - \frac{\sqrt{n}}{n_1}\sum_{i=1}^{n_1} m_1(W_{1i};\widehat{\mu}_0^p)\right] \\
&- \frac{\sqrt{n}}{n_0}\sum_{i=1}^{n_0}(\widetilde{S}_{0i} - \widetilde{\mu}_0) + \pi_1\left[\frac{\sqrt{n}}{n_1}\sum_{i=1}^{n_1} m_0(W_{1i};\widehat{\mu}_0^p) - \frac{\sqrt{n}}{n_0}\sum_{i=1}^{n_0} m_0(W_{0i};\widehat{\mu}_0^p)\right] + o_p(1) \\
=&\frac{\sqrt{n}}{n_1}\sum_{i=1}^{n_1}\Big(\widetilde{S}_{1i} - \pi_0 m_1(W_{1i};\widehat{\mu}_0^p) - \pi_1 m_0(W_{1i};\widehat{\mu}_0^p) - \pi_1(\widetilde{\mu}_1 - \widetilde{\mu}_0)\Big) \\
&- \frac{\sqrt{n}}{n_0}\sum_{i=1}^{n_0}\Big(\widetilde{S}_{0i} - \pi_0 m_1(W_{0i};\widehat{\mu}_0^p) - \pi_1 m_0(W_{0i};\widehat{\mu}_0^p) - \pi_0(\widetilde{\mu}_1 - \widetilde{\mu}_0)\Big) + o_p(1),
\end{aligned}
$$

which converges weakly to a mean zero Gaussian distribution with a variance of

$$
\frac{1}{\pi_1}E\left\{\widetilde{S}_{1i} - \pi_0 m_1(W_{1i};\widehat{\mu}_0^p) - \pi_1 m_0(W_{1i};\widehat{\mu}_0^p) - \pi_1\widetilde{\Delta}_H\right\}^2
$$

$$
+\frac{1}{\pi_0}E\left\{\widetilde{S}_{0i} - \pi_0 m_1(W_{0i};\widehat{\mu}_0^p) - \pi_1 m_0(W_{0i};\widehat{\mu}_0^p) - \pi_0\widetilde{\Delta}_H\right\}^2.
$$

Therefore, the variance of $\widehat{\Delta}_H$ can be estimated as

$$
\widehat{\sigma}_H^2 = \frac{1}{n_1^2}\sum_{i=1}^{n_1}\Big(\widetilde{S}_{1i} - \pi_0\widehat{m}_1(W_{1i};\widehat{\mu}_0^p)) - \pi_1\widehat{m}_0(W_{1i};\widehat{\mu}_0^p) - \pi_1\widehat{\Delta}_H)\Big)^2
$$

$$
+\frac{1}{n_0^2}\sum_{i=1}^{n_0}\Big(\widetilde{S}_{0i} - \pi_0\widehat{m}_1(W_{0i};\widehat{\mu}_0^p) - \pi_1\widehat{m}_0(W_{0i};\widehat{\mu}_0^p) - \pi_0\widehat{\Delta}_H\Big)^2
$$

Next, we will derive the asymptotical distribution of $\sqrt{n}(\widehat{\Delta}_H^{AUG} - \widetilde{\Delta}_H)$. It is clear that

$$
\begin{aligned}
&\sqrt{n}(\widehat{\Delta}_H^{AUG} - \widetilde{\Delta}_H) \\
=&\frac{\sqrt{n}}{n_1}\sum_{i=1}^{n_1}\left\{\widetilde{S}_{1i} - \pi_0\widehat{m}_1(W_{1i};\widehat{\mu}_0^p) - \pi_1\widehat{m}_0(W_{1i};\widehat{\mu}_0^p) - \pi_1\widetilde{\Delta}_H\right\} \\
&-\frac{\sqrt{n}}{n_0}\sum_{i=1}^{n_1}\left\{\widetilde{S}_{0i} - \pi_0\widehat{m}_1(W_{0i};\widehat{\mu}_0^p) - \pi_1\widehat{m}_0(W_{0i};\widehat{\mu}_0^p) - \pi_0\widetilde{\Delta}_H\right\} \\
=&\frac{\sqrt{n}}{n_1}\sum_{i=1}^{n_1}\left\{\widetilde{S}_{1i} - \pi_0 m_1(W_{1i};\widehat{\mu}_0^p) - \pi_1 m_0(W_{1i};\widehat{\mu}_0^p) - \pi_1\widetilde{\Delta}_H\right\} \\
&-\frac{\sqrt{n}}{n_0}\sum_{i=1}^{n_1}\left\{\widetilde{S}_{0i} - \pi_0 m_1(W_{0i};\widehat{\mu}_0^p) - \pi_1 m_0(W_{0i};\widehat{\mu}_0^p) - \pi_0\widetilde{\Delta}_H\right\} \\
&-\sqrt{n}\left[\frac{\pi_0}{n_0}\sum_{i=1}^{n_0}(\widehat{m}_1(W_{0i};\widehat{\mu}_0^p) - m_1(W_{0i};\widehat{\mu}_0^p)) - \frac{\pi_0}{n_1}\sum_{i=1}^{n_1}(\widehat{m}_1(W_{1i};\widehat{\mu}_0^p) - m_1(W_{1i};\widehat{\mu}_0^p))\right] \\
&-\sqrt{n}\left[\frac{\pi_1}{n_1}\sum_{i=1}^{n_1}(\widehat{m}_1(W_{1i};\widehat{\mu}_0^p) - m_1(W_{1i};\widehat{\mu}_0^p)) - \frac{\pi_1}{n_0}\sum_{i=1}^{n_0}(\widehat{m}_0(W_{0i};\widehat{\mu}_0^p) - m_1(W_{0i};\widehat{\mu}_0^p))\right] \\
=&\frac{\sqrt{n}}{n_1}\sum_{i=1}^{n_1}\left\{\widetilde{S}_{1i} - \pi_0 m_1(W_{1i};\widehat{\mu}_0^p) - \pi_1 m_0(W_{1i};\widehat{\mu}_0^p) - \pi_1\widetilde{\Delta}_H\right\} \\
&-\frac{\sqrt{n}}{n_0}\sum_{i=1}^{n_1}\left\{\widetilde{S}_{0i} - \pi_0 m_1(W_{0i};\widehat{\mu}_0^p) - \pi_1 m_0(W_{0i};\widehat{\mu}_0^p) - \pi_0\widetilde{\Delta}_H\right\} + o_p(1) \\
=&\sqrt{n}(\widehat{\Delta}_H - \widetilde{\Delta}_H) + o_p(1).
\end{aligned}
$$

Therefore, $\widehat{\Delta}_H^{AUG}$ and $\widehat{\Delta}_H$ are asymptotically equivalent. Furthermore, noting that

$$
\begin{aligned}
&\widetilde{S}_{1i} - \pi_0 m_1(W_{1i};\widehat{\mu}_0^p) - \pi_1 m_0(W_{1i};\widehat{\mu}_0^p) - \pi_1\widetilde{\Delta}_H \\
=&\left\{\widetilde{S}_{1i} - m_1(W_{1i};\widehat{\mu}_0^p)\right\} + \pi_1\left\{m_1(W_{1i};\widehat{\mu}_0^p) - m_0(W_{1i};\widehat{\mu}_0^p) - \widetilde{\Delta}_H\right\}
\end{aligned}
$$

and

$$
E\left[\left\{\widetilde{S}_{1i} - m_1(W_{1i};\widehat{\mu}_0^p)\right\}\left\{m_1(W_{1i};\widehat{\mu}_0^p) - m_0(W_{1i};\widehat{\mu}_0^p) - \widetilde{\Delta}_H\right\} \mid W_{1i}\right] = 0,
$$

we have

$$
\begin{aligned}
&E\left[\widetilde{S}_{1i} - \pi_0 m_1(W_{1i};\widehat{\mu}_0^p) - \pi_1 m_0(W_{1i};\widehat{\mu}_0^p) - \pi_1 \widetilde{\Delta}_H\right]^2 \\
=&E\left[\widetilde{S}_{1i} - m_1(W_{1i};\widehat{\mu}_0^p)\right]^2 + \pi_1^2 E\left[m_1(W_{1i};\widehat{\mu}_0^p) - m_0(W_{1i};\widehat{\mu}_0^p) - \widetilde{\Delta}_H\right]^2.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&E\left[\widetilde{S}_{0i} - \pi_0 m_1(W_{0i};\widehat{\mu}_0^p) - \pi_1 m_0(W_{0i};\widehat{\mu}_0^p) - \pi_0 \widetilde{\Delta}_H\right]^2 \\
=&E\left[\widetilde{S}_{0i} - m_0(W_{0i};\widehat{\mu}_0^p)\right]^2 + \pi_0^2 E\left[m_1(W_{0i};\widehat{\mu}_0^p) - m_0(W_{0i};\widehat{\mu}_0^p) - \widetilde{\Delta}_H\right]^2.
\end{aligned}
$$

Therefore, the variance of $\widehat{\Delta}_H^{(AUG)}$ can also be consistently estimated by

$$
\widehat{\sigma}^2_{AUG} = \frac{1}{n_1^2}\sum_{i=1}^{n_1}\left[\widehat{\mu}_0^{(p)}(S_{1i}, W_{1i}) - \widehat{m}_1(W_{1i};\widehat{\mu}_0^p)\right]^2 + \frac{1}{n_0^2}\sum_{i=1}^{n_0}\left[\widehat{\mu}_0^{(p)}(S_{0i}, W_{0i}) - \widehat{m}_0(W_{0i};\widehat{\mu}_0^p)\right]^2
$$

$$
+\frac{\pi_1^2}{n_1^2}\sum_{i=1}^{n_1}\left[\widehat{m}_1(W_{1i};\widehat{\mu}_0^p) - \widehat{m}_0(W_{1i};\widehat{\mu}_0^p) - \widehat{\Delta}_H\right]^2 + \frac{\pi_0^2}{n_0^2}\sum_{i=1}^{n_0}\left[\widehat{m}_1(W_{0i};\widehat{\mu}_0^p) - \widehat{m}_0(W_{0i};\widehat{\mu}_0^p) - \widehat{\Delta}_H\right]^2,
$$

and $\widehat{\Delta}_{(AUG)}/\widehat{\Delta}_H = 1 + o_p(1)$.

Appendix E

Here, we provide an example where there is heterogeneity in the utility of the surrogate and the W is distributed the *same* in the prior study and current study, but Δ_P still fails to provide a lower bound for Δ. In both the prior study and the current study, we assume that $\log(W) \sim \epsilon_W$, $S^{(g)} = W \times \exp(\delta_0 g + \epsilon_S)$, and $Y^{(g)} = S^{(g)}W, g \in \{0, 1\}$, where δ_0 is a positive constant, and ϵ_W and ϵ_S are two independent standard normals. It is obvious that

$\mu_0^p(s,w) = sw$ and

$$\begin{aligned}\Delta = \Delta_H =& E(S^{(1)}W) - E(S^{(0)}W) = E\left\{WE(S^{(1)} - S^{(0)} \mid W)\right\} \\ =& E\left\{W\left(\exp(0.5+\delta_0)W - \exp(0.5)W\right)\right\} = \exp\left(\frac{5}{2}\right)\left(\exp(\delta_0)-1\right).\end{aligned}$$

Next, we have

$$\begin{aligned}\mu_0^p(s) =& E(WS^{(0)} \mid S^{(0)} = s) = sE(W^{(0)} \mid S^{(0)} = s) \\ =& s \times \exp\left(\frac{1}{4}\right)s^{\frac{1}{2}} = \exp\left(\frac{1}{4}\right)s^{\frac{3}{2}},\end{aligned}$$

and

$$\begin{aligned}\Delta_P =& E\left\{\left(S^{(1)}\right)^{\frac{3}{2}}\exp\left(\frac{1}{4}\right)\right\} - E\left\{\left(S^{(0)}\right)^{\frac{3}{2}}\exp\left(\frac{1}{4}\right)\right\} \\ =& \exp\left(\frac{5}{2}\right)\left(\frac{3\delta_0}{2} - 1\right).\end{aligned}$$

Consequently, in this setting, $\Delta_P > \Delta = \Delta_H$ even though the W has the same distribution in both studies.

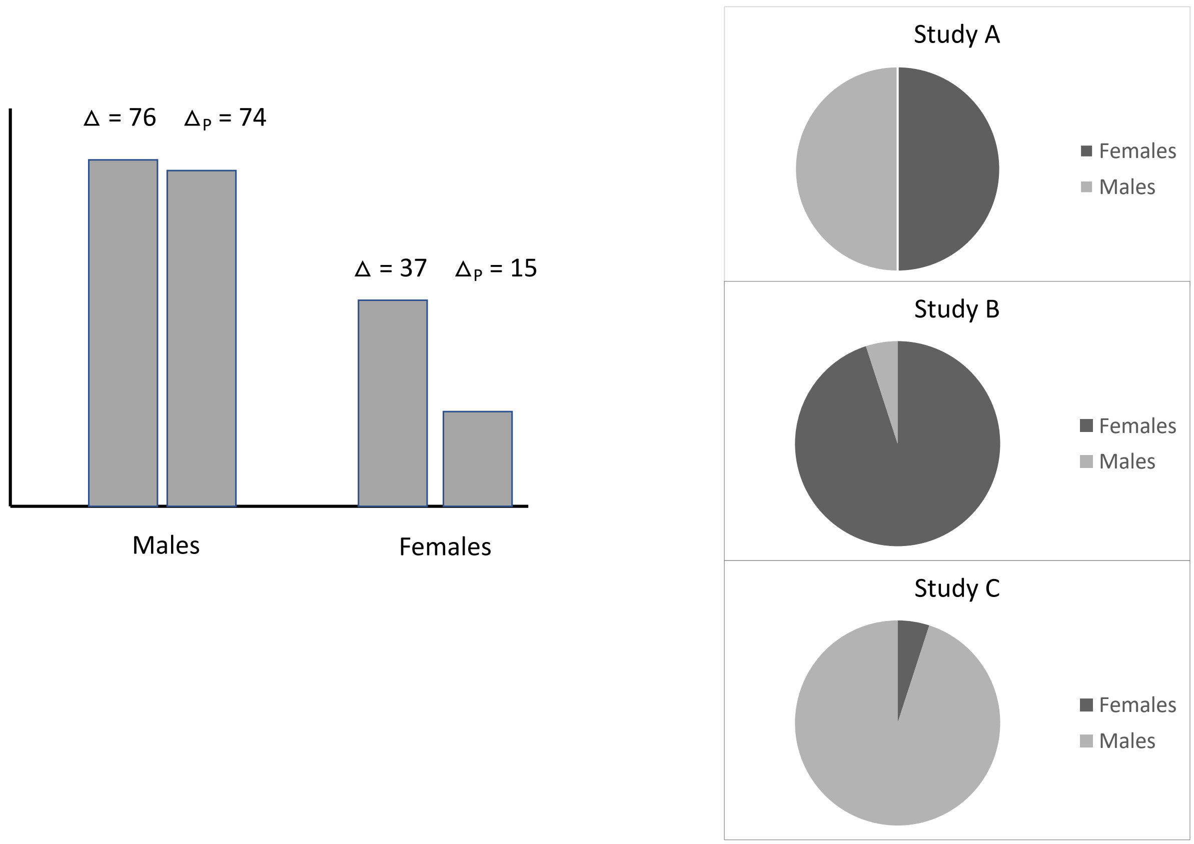

Figure A1: Discrete data example

References

Parast, L., Cai, T., and Tian, L. (2019). Using a surrogate marker for early testing of a treatment effect. *Biometrics* **75,** 1253–1263.