Appendix

Identification of the average treatment effect

If the causal quantity of interest is the average treatment effect in the target population (instead of each potential outcome mean), we can replace identifiability condition (4) in the main text with

(4*) *Generalizability in measure*,

$$\mathrm{E}[Y^1|X, S=1] - \mathrm{E}[Y^0|X, S=1] = \mathrm{E}[Y^1|X] - \mathrm{E}[Y^0|X],$$

Generalizability in measure can be interpreted as the absence of effect measure modification by S on the additive scale, conditional on baseline covariates.

This condition suffices to identify $\mathrm{E}[Y^1 - Y^0]$ because

$$\begin{aligned}
\mathrm{E}[Y^1 - Y^0] &= \mathrm{E}\big[\mathrm{E}[Y^1 - Y^0|X]\big] \\
&= \mathrm{E}\big[\mathrm{E}[Y^1|X] - \mathrm{E}[Y^0|X]\big] \\
&= \mathrm{E}\big[\mathrm{E}[Y^1|X, S=1] - \mathrm{E}[Y^0|X, S=1]\big] \\
&= \mathrm{E}\big[\mathrm{E}[Y^1|X, S=1, A=1]\big] - \mathrm{E}\left[\mathrm{E}[Y^0|X, S=1, A=0]\right] \\
&= \mathrm{E}\big[\mathrm{E}[Y|X, S=1, A=1]\big] - \mathrm{E}\left[\mathrm{E}[Y|X, S=1, A=0]\right],
\end{aligned}$$

where the third equality follows from generalizability in measure.

Normalized weights

Identity (3), in the main text, holds because

$$\begin{aligned}
\mathrm{E}\left[\frac{I(S=1, A=a)}{\Pr[S=1, A=a|X]}\right] &= \mathrm{E}\left[\mathrm{E}\left[\frac{I(S=1, A=a)}{\Pr[S=1, A=a|X]}\middle|X\right]\right] \\
&= \mathrm{E}\left[\frac{\mathrm{E}[I(S=1, A=a)|X]}{\Pr[S=1, A=a|X]}\right] \\
&= 1.
\end{aligned}$$

g-formula and IP weighting estimators

Recall that we want to estimate $\psi(a) = \mathrm{E}\big[\mathrm{E}[Y|X, S=1, A=a]\big]$, or, for discrete X,

$$\psi(a) = \sum_x \mathrm{E}[Y|X=x, S=1, A=a] \Pr[X=x]. \tag{A.1}$$

Finite sample equivalence of g-formula and IP weighting estimators: Using equation (A.1), by nonparametrically estimating the conditional expectation of the outcome and the probability mass function of X, the g-formula-based estimator can be written as

$$\begin{aligned}
\widehat{\psi}_{\mathrm{G}}(a) &= \sum_x \widehat{\mathrm{E}}[Y|X=x, S=1, A=a] \widehat{\Pr}[X=x] \\
&= \sum_x \frac{\sum_{i=1}^m Y_i I(S_i=1, A_i=a, X_i=x)}{\sum_{i=1}^m I(S_i=1, A_i=a, X_i=x)} \times \frac{\sum_{i=1}^m I(X_i=x)}{m},
\end{aligned}$$

where we have used the nonparametric frequency estimators

$$\widehat{\mathrm{E}}[Y|X=x, S=1, A=a] = \frac{\sum_{i=1}^m Y_i I(S_i=1, A_i=a, X_i=x)}{\sum_{i=1}^m I(S_i=1, A_i=a, X_i=x)} \text{ and}$$

$$\widehat{\Pr}[X=x] = \frac{\sum_{i=1}^m I(X_i=x)}{m}.$$

We now show that, when the conditional outcome mean, and the probability of trial participation and treatment are nonparametrically estimated, the proposed IP weighting estimator is equivalent

to the g-formula estimator:

$$\begin{aligned}
\widehat{\psi}^{\text{HT}}_{\text{IPW}}(a) &= \frac{1}{m}\sum_{i=1}^{m}\frac{Y_i I(S_i=1,A_i=a)}{\widehat{p}(X_i)\widehat{e}_a(X_i)} \\
&= \frac{1}{m}\sum_{i=1}^{m}\sum_{x}\frac{Y_i I(S_i=1,A_i=a,X_i=x)}{\widehat{p}(x)\widehat{e}_a(x)} \\
&= \frac{1}{m}\sum_{i=1}^{m}\sum_{x}\frac{Y_i I(S_i=1,A_i=a,X_i=x)\sum_{j=1}^{m} I(X_j=x)}{\sum_{j=1}^{m} I(S_j=1,A_j=a,X_j=x)} \\
&= \sum_{x}\frac{\sum_{i=1}^{m} Y_i I(S_i=1,A_i=a,X_i=x)}{\sum_{i=1}^{m} I(S_i=1,A_i=a,X_i=x)} \times \frac{\sum_{i=1}^{m} I(X_i=x)}{m} \\
&= \widehat{\psi}_{\text{G}}(a).
\end{aligned}$$

In the above derivation, we have used the nonparametric frequency estimators

$$\widehat{p}(x) = \frac{\sum_{i=1}^{m} I(S_i=1,X_i=x)}{\sum_{i=1}^{m} I(X_i=x)} \text{ and } \widehat{e}_a(x) = \frac{\sum_{i=1}^{m} I(S_i=1,A_i=a,X_i=x)}{\sum_{i=1}^{m} I(S_i=1,X_i=x)},$$

so that

$$\left\{\widehat{p}(x)\widehat{e}_a(x)\right\}^{-1} = \frac{\sum_{i=1}^{m} I(X_i=x)}{\sum_{i=1}^{m} I(S_i=1,A_i=a,X_i=x)}.$$

Normalizing the weights to sum to 1: When $p(X)$ and $e_a(X)$ are nonparametric estimators for $\Pr[S=1|X]$ and $\Pr[A=a|X,S=1]$, respectively, the sum of the weights in each treatment group in the trial equals the total number of observations,

$$\sum_{i=1}^{m}\frac{I(S_i=1,A_i=a)}{\widehat{p}(X_i)\widehat{e}_a(X_i)} = m.$$

It immediately follows that the IP weighting estimator in equation (4) can be written as

$$\widehat{\psi}^{\text{HT}}_{\text{IPW}}(a) = \left\{\sum_{i=1}^{m}\frac{I(S_i=1,A_i=a)}{\widehat{p}(X_i)\widehat{e}_a(X_i)}\right\}^{-1}\sum_{i=1}^{m}\frac{Y_i I(S_i=1,A_i=a)}{\widehat{p}(X_i)\widehat{e}_a(X_i)} = \widehat{\psi}^{\text{Hájek}}_{\text{IPW}}(a).$$

When do $\widehat{\psi}_{\mathbf{G}}(a)$ and $\widehat{\psi}^{\mathbf{part}}_{\mathbf{IPW}}(a)$ produce identical results?

To see that, when the probability of trial participation is nonparametrically estimated, discrepancies between $\widehat{\psi}_{\mathrm{G}}(a)$ and $\widehat{\psi}^{\mathrm{part}}_{\mathrm{IPW}}(a)$ are due to baseline covariate imbalances among trial participants, we examine when these two estimators are equivalent. It turns out that the two estimators produce identical estimates when

$$\frac{\sum_{i=1}^{m} I(S_i = 1, A_i = a, X_i = x)}{\widehat{\Pr}[X = x, S = 1]} = \widehat{c},$$

where $\widehat{c}$ is a function of the observed data that does not depend on x (see the Appendix of [7] for the derivation of this condition). We can re-write the above condition as

$$\frac{\sum_{i=1}^{m} I(S_i = 1, A_i = a, X_i = x)}{\sum_{i=1}^{m} I(S_i = 1, X_i = x)} = \frac{\widehat{c}}{m}, \text{ or } \widehat{\Pr}[A = a | X = x, S = 1] = \frac{\widehat{c}}{m}.$$

Because the condition has to hold for each covariate pattern x, it implies that the empirical probability of treatment among trial participants does not depend on X, or, in other words, that X is *perfectly balanced* among treatment groups in the trial. Such perfect balance does not typically occur in finite samples from marginally randomized trials due to sampling variability. We can *induce perfect balance*, however, by re-weighing the trial data using a nonparametric estimate of the probability of receiving the treatment actually received, as is done in our IP weighting estimators.

Complete simulation study results

eAppendix Table 1: Simulation results comparing estimators using parametric models.

$\Pr[S=1]$	m	Estimator	$\psi(1)$		$\psi(0)$		$\psi(1)-\psi(0)$	
			Bias	Variance	Bias	Variance	Bias	Variance
0.05	2000	$\widehat{\psi}_{\mathrm{G}}(a)$	0.0022	0.0604	0.0031	0.0614	-0.0009	0.1240
		$\widehat{\psi}^{\mathrm{HT}}_{\mathrm{IPW}}(a)$	-0.0357	1.0953	0.0287	1.3603	-0.0645	2.4094
		$\widehat{\psi}^{\mathrm{Hájek}}_{\mathrm{IPW}}(a)$	-0.1436	0.3063	0.1491	0.3208	-0.2926	0.6113
		$\widehat{\psi}^{\mathrm{part}}_{\mathrm{IPW}}(a)$	-0.1484	0.3219	0.1595	0.3346	-0.3079	0.5938
		$\widehat{\psi}_{\mathrm{AIPW}}(a)$	-0.0014	0.1147	0.0041	0.1233	-0.0055	0.2426
	5000	$\widehat{\psi}_{\mathrm{G}}(a)$	-0.0009	0.0229	-0.0009	0.0228	-0.0000	0.0469
		$\widehat{\psi}^{\mathrm{HT}}_{\mathrm{IPW}}(a)$	0.0008	0.7642	0.0166	0.4519	-0.0158	1.2112
		$\widehat{\psi}^{\mathrm{Hájek}}_{\mathrm{IPW}}(a)$	-0.0657	0.1890	0.0687	0.1722	-0.1344	0.3630
		$\widehat{\psi}^{\mathrm{part}}_{\mathrm{IPW}}(a)$	-0.0706	0.1931	0.0732	0.1772	-0.1438	0.3523
		$\widehat{\psi}_{\mathrm{AIPW}}(a)$	-0.0006	0.0537	-0.0017	0.0462	0.0010	0.1016
	10,000	$\widehat{\psi}_{\mathrm{G}}(a)$	-0.0004	0.0113	0.0001	0.0111	-0.0005	0.0229
		$\widehat{\psi}^{\mathrm{HT}}_{\mathrm{IPW}}(a)$	-0.0175	0.2131	0.0051	0.2484	-0.0226	0.4509
		$\widehat{\psi}^{\mathrm{Hájek}}_{\mathrm{IPW}}(a)$	-0.0440	0.1030	0.0370	0.1065	-0.0810	0.2062
		$\widehat{\psi}^{\mathrm{part}}_{\mathrm{IPW}}(a)$	-0.0462	0.1050	0.0410	0.1096	-0.0873	0.2019
		$\widehat{\psi}_{\mathrm{AIPW}}(a)$	0.0001	0.0234	0.0014	0.0241	-0.0014	0.0477
0.1	2000	$\widehat{\psi}_{\mathrm{G}}(a)$	-0.0028	0.0270	0.0008	0.0260	-0.0036	0.0551
		$\widehat{\psi}^{\mathrm{HT}}_{\mathrm{IPW}}(a)$	-0.0224	0.4550	0.0195	0.5051	-0.0419	0.9349
		$\widehat{\psi}^{\mathrm{Hájek}}_{\mathrm{IPW}}(a)$	-0.0741	0.1752	0.0735	0.1797	-0.1476	0.3467
		$\widehat{\psi}^{\mathrm{part}}_{\mathrm{IPW}}(a)$	-0.0792	0.1847	0.0795	0.1850	-0.1587	0.3368
		$\widehat{\psi}_{\mathrm{AIPW}}(a)$	-0.0047	0.0497	-0.0001	0.0510	-0.0046	0.1021
	5000	$\widehat{\psi}_{\mathrm{G}}(a)$	-0.0002	0.0103	0.0015	0.0106	-0.0017	0.0218
		$\widehat{\psi}^{\mathrm{HT}}_{\mathrm{IPW}}(a)$	-0.0046	0.1843	0.0075	0.1986	-0.0121	0.3741
		$\widehat{\psi}^{\mathrm{Hájek}}_{\mathrm{IPW}}(a)$	-0.0292	0.0891	0.0322	0.0926	-0.0614	0.1782
		$\widehat{\psi}^{\mathrm{part}}_{\mathrm{IPW}}(a)$	-0.0320	0.0930	0.0354	0.0953	-0.0674	0.1751
		$\widehat{\psi}_{\mathrm{AIPW}}(a)$	0.0006	0.0203	0.0016	0.0212	-0.0010	0.0422
	10,000	$\widehat{\psi}_{\mathrm{G}}(a)$	0.0005	0.0051	-0.0004	0.0051	0.0009	0.0104
		$\widehat{\psi}^{\mathrm{HT}}_{\mathrm{IPW}}(a)$	-0.0049	0.0959	0.0024	0.0965	-0.0073	0.1908
		$\widehat{\psi}^{\mathrm{Hájek}}_{\mathrm{IPW}}(a)$	-0.0174	0.0507	0.0154	0.0514	-0.0329	0.1022
		$\widehat{\psi}^{\mathrm{part}}_{\mathrm{IPW}}(a)$	-0.0187	0.0531	0.0175	0.0534	-0.0361	0.1013
		$\widehat{\psi}_{\mathrm{AIPW}}(a)$	0.0005	0.0098	-0.0015	0.0103	0.0019	0.0206
0.2	2000	$\widehat{\psi}_{\mathrm{G}}(a)$	-0.0003	0.0112	0.0001	0.0115	-0.0004	0.0245
		$\widehat{\psi}^{\mathrm{HT}}_{\mathrm{IPW}}(a)$	-0.0089	0.1713	-0.0033	0.2195	-0.0055	0.3742
		$\widehat{\psi}^{\mathrm{Hájek}}_{\mathrm{IPW}}(a)$	-0.0298	0.0839	0.0228	0.0874	-0.0526	0.1640
		$\widehat{\psi}^{\mathrm{part}}_{\mathrm{IPW}}(a)$	-0.0325	0.0897	0.0259	0.0926	-0.0584	0.1607
		$\widehat{\psi}_{\mathrm{AIPW}}(a)$	-0.0014	0.0210	0.0013	0.0205	-0.0027	0.0430
	5000	$\widehat{\psi}_{\mathrm{G}}(a)$	-0.0006	0.0045	-0.0002	0.0046	-0.0004	0.0097
		$\widehat{\psi}^{\mathrm{HT}}_{\mathrm{IPW}}(a)$	-0.0032	0.0642	-0.0011	0.0695	-0.0021	0.1279
		$\widehat{\psi}^{\mathrm{Hájek}}_{\mathrm{IPW}}(a)$	-0.0121	0.0361	0.0086	0.0382	-0.0207	0.0718
		$\widehat{\psi}^{\mathrm{part}}_{\mathrm{IPW}}(a)$	-0.0131	0.0385	0.0104	0.0405	-0.0235	0.0709
		$\widehat{\psi}_{\mathrm{AIPW}}(a)$	-0.0012	0.0079	-0.0004	0.0080	-0.0008	0.0165
	10,000	$\widehat{\psi}_{\mathrm{G}}(a)$	-0.0001	0.0023	0.0000	0.0022	-0.0001	0.0048
		$\widehat{\psi}^{\mathrm{HT}}_{\mathrm{IPW}}(a)$	0.0007	0.0334	0.0015	0.0331	-0.0008	0.0644
		$\widehat{\psi}^{\mathrm{Hájek}}_{\mathrm{IPW}}(a)$	-0.0044	0.0201	0.0061	0.0199	-0.0104	0.0391
		$\widehat{\psi}^{\mathrm{part}}_{\mathrm{IPW}}(a)$	-0.0053	0.0211	0.0065	0.0213	-0.0119	0.0388
		$\widehat{\psi}_{\mathrm{AIPW}}(a)$	0.0003	0.0040	-0.0002	0.0039	0.0005	0.0081