Web Appendix

Study Designs for Extending Causal Inferences from a Randomized Trial to a Target Population

Issa J. Dahabreh[1-4], Sebastien J-P.A. Haneuse[5], James M. Robins[4,5], Sarah E. Robertson[1,2], Ashley L. Buchanan[6], Elizabeth A. Stuart[7], and Miguel A. Hernán[4,5,8]

[1]Center for Evidence Synthesis in Health, Brown University School of Public Health, Providence, RI

[2]Department of Health Services, Policy & Practice, School of Public Health, Brown University, Providence, RI

[3]Department of Epidemiology, School of Public Health, Brown University, Providence, RI

[4]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA

[5]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA

[6]Department of Pharmacy Practice, College of Pharmacy, University of Rhode Island, RI

[7]Departments of Mental Health, Biostatistics, and Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD

[8]Harvard-MIT Division of Health Sciences and Technology, Boston, MA

Sunday 29th November, 2020

Table of Contents

Web Appendix 1 The nested trial design with subsampling using covariate-dependent sampling probabilities

As noted in the main text, a more general version of the nested trial design assumes that the sampling probabilities for non-randomized individuals depend on baseline auxiliary covariates [1]. Let $X = (X_1, X_2)$, where X_1 represents baseline auxiliary covariates that are available on all members of the actual population in which the trial is nested, and X_2 represents covariates that are only measured among randomized individuals $(S = 1)$ and sampled non-randomized individuals $(S = 0, D = 1)$.

The identifiability conditions and identification results remain the same as in the main text; but the sampling properties of this design are

$$\begin{aligned}
&\Pr[D = 1|S = 1] = 1, \text{ and} \\
&\Pr[D = 1|X, A, Y, S = 0] = \Pr[D = 1|X_1, S = 0] \equiv c(X_1),
\end{aligned}$$

where $0 < c(X_1) \leq 1$ is a known function that only depends on X_1, allowing the sampling of non-randomized individuals to depend on the auxiliary covariates that are available from all members of the actual population.

Web Appendix 2 Identification of the probability of trial participation in the nested trial design with subsampling (simple random sampling)

In this section we present identification results for the marginal and conditional probabilities of trial participation in nested trial designs with subsampling when a simple random sample of non-randomized individuals is obtained.

Identification of the marginal probability of trial participation

Using the definition of conditional probability and re-arranging,

$$\Pr[S=1] = \frac{\Pr[S=1|D=1]\Pr[D=1]}{\Pr[D=1|S=1]} \text{ and } \Pr[S=0] = \frac{\Pr[S=0|D=1]\Pr[D=1]}{\Pr[D=1|S=0]}.$$

Taking the ratio of the above expressions and exploiting the sampling properties for non-nested trial designs,

$$\begin{aligned}\frac{\Pr[S=1]}{\Pr[S=0]} &= \frac{\Pr[S=1|D=1]}{\Pr[S=0|D=1]} \times \frac{\Pr[D=1|S=0]}{\Pr[D=1|S=1]} \\ &= \frac{\Pr[S=1|D=1]}{\Pr[S=0|D=1]} \times c.\end{aligned}$$

With a bit of algebra, the above expression can be re-arranged to show that

$$\Pr[S=1] = \left\{ 1 + \frac{\Pr[S=0|D=1]}{\Pr[S=1|D=1]} \times c^{-1} \right\}^{-1}.$$

By setting $c=1$ we see that in the nested-trial design with a census of non-randomized individuals $\Pr[S=1] = \Pr[S=1|D=1]$.

Identification of the conditional probability of trial participation

The argument for the conditional probability is parallel to the one presented above for the marginal probability. Using properties of conditional probability,

$$\Pr[S=1|X] = \frac{\Pr[S=1|X,D=1]\Pr[D=1|X]}{\Pr[D=1|X,S=1]} \text{ and}$$

$$\Pr[S=0|X] = \frac{\Pr[S=0|X,D=1]\Pr[D=1|X]}{\Pr[D=1|X,S=0]}.$$

Taking the ratio of the above expressions and exploiting the sampling properties for non-nested trial designs,

$$\begin{aligned}\frac{\Pr[S=1|X]}{\Pr[S=0|X]} &= \frac{\Pr[S=1|X,D=1]}{\Pr[S=0|X,D=1]} \times \frac{\Pr[D=1|X,S=0]}{\Pr[D=1|X,S=1]} \\ &= \frac{\Pr[S=1|X,D=1]}{\Pr[S=0|X,D=1]} \times c.\end{aligned}$$

The above expression can be re-arranged to show that

$$\Pr[S=1|X] = \left\{1 + \frac{\Pr[S=0|X,D=1]}{\Pr[S=1|X,D=1]} \times c^{-1}\right\}^{-1}.$$

By setting $c = 1$ we see that in the nested-trial design with a census of non-randomized individuals $\Pr[S=1|X] = \Pr[S=1|X,D=1]$.

Web Appendix 3 Identification of the probability of trial participation in the nested trial design with subsampling (covariate-dependent sampling)

In this section we present identification results for the marginal and conditional probabilities of participation in nested trial designs with subsampling when the sampling of non-randomized individuals uses covariate-dependent sampling probabilities. The sampling properties of this design are given in Web Appendix 1.

Identification of the marginal probability of trial participation

Using the results from Web Appendix 2, under covariate-dependent sampling of non-randomized individuals, it is still true that

$$\begin{aligned}
\frac{\Pr[S=1]}{\Pr[S=0]} &= \frac{\Pr[S=1|D=1]}{\Pr[S=0|D=1]} \times \frac{\Pr[D=1|S=0]}{\Pr[D=1|S=1]} \\
&= \frac{\Pr[S=1|D=1]}{\Pr[S=0|D=1]} \times \Pr[D=1|S=0],
\end{aligned}$$

where the second equality follows from the sampling properties of the design.

We will now show that $\Pr[D=1|S=0]$ is identifiable in this design, using the known sampling probabilities $c(X_1)$ among the non-randomized individuals. To see that, consider the expectation of the inverse of the sampling probabilities, among sampled non-randomized individuals:

$$\begin{aligned}
\mathrm{E}\left[\frac{1}{c(X_1)}\middle|S=0,D=1\right] &= \mathrm{E}\left[\frac{1}{\Pr[D=1|X_1,S=0]}\middle|S=0,D=1\right] \\
&= \frac{1}{\Pr[S=0,D=1]}\,\mathrm{E}\left[\frac{I(S=0,D=1)}{\Pr[D=1|X_1,S=0]}\right] \\
&= \frac{1}{\Pr[S=0,D=1]}\,\mathrm{E}\left[\frac{\Pr[S=0,D=1|X_1]}{\Pr[D=1|X_1,S=0]}\right] \\
&= \frac{1}{\Pr[S=0,D=1]}\,\mathrm{E}\left[\Pr[S=0|X_1]\right] \\
&= \frac{\Pr[S=0]}{\Pr[S=0,D=1]} \\
&= \frac{1}{\Pr[D=1|S=0]}.
\end{aligned}$$

Using this result, we can write

$$\frac{\Pr[S=1]}{\Pr[S=0]} = \frac{\Pr[S=1|D=1]}{\Pr[S=0|D=1]} \times \left\{ \mathrm{E}\left[\frac{1}{c(X_1)} \Big| S=0, D=1 \right] \right\}^{-1},$$

which establishes the identifiability of the marginal odds of trial participation in the nested trial design with covariate-dependent sampling of non-randomized individuals because all quantities on the right-hand-side of the above equation are conditional on $D = 1$.

Last, with a bit of algebra, we obtain the following expression for the marginal probability of trial participation:

$$\Pr[S=1] = \left\{ 1 + \frac{\Pr[S=0|D=1]}{\Pr[S=1|D=1]} \times \mathrm{E}\left[\frac{1}{c(X_1)} \Big| S=0, D=1 \right] \right\}.$$

Identification of the conditional probability of trial participation

Using an argument similar to the one in Web Appendix 2, but applied to the case of covariate-dependent sampling of non-randomized individuals, we obtain

$$\Pr[S=1|X] = \left\{ 1 + \frac{\Pr[S=0|X, D=1]}{\Pr[S=1|X, D=1]} \times \frac{1}{c(X_1)} \right\}^{-1}.$$

This result establishes the identifiability of the conditional probability of trial participation in nested trial designs with covariate-dependent sampling of non-randomized individuals because the inverse of the conditional odds of trial participation in the sampled data, $\dfrac{\Pr[S=0|X, D=1]}{\Pr[S=1|X, D=1]}$, are identifiable, and $c(X_1)$ is known by design.

A note about the identification of counterfactual quantities

To obtain identification results analogous to those in the main text, for the counterfactual quantities of interest, $\mathrm{E}[Y^a]$ and $\mathrm{E}[Y^a|S=0]$, in the nested trial design with subsampling using covariate-dependent sampling probabilities, in addition to the identifiability of the probability of trial participation, we need to also establish the identifiability of the distribution of covariates in the target population and in the non-randomized subset of that population.[1] Because the identifiability of the distribution of the covariates in the randomized subset of the population follows immediately from the sampling properties, and $\Pr[S=1]$ is also identifiable as argued earlier in this section of the Appendix, the main issue is the identifiability of the distribution of covariates among the non-randomized subset of the target population.

[1] An alternative identification approach for $\mathrm{E}[Y^a]$ in the nested trial design with subsampling using covariate-dependent sampling probabilities is presented in [1].

Let $x = (x_1, x_2)$ be a realization of $X = (X_1, X_2)$. With little loss of generality, we will show that the density function $f(x|S = 0)$ of the baseline covariates in the non-randomized subset of the target population is indeed identifiable as

$$f(x|S = 0) = \frac{1}{c(x_1) \times \theta} f(x|S = 0, D = 1),$$

where $\theta = \mathrm{E}\left[\frac{1}{c(X_1)}\middle| S = 0, D = 1\right]$ and all quantities on the right hand side are known by design or conditioned on the subset of sampled non-randomized individuals, $(S = 0, D = 1)$.

As shown earlier in this section of the Appendix, $\Pr[D = 1|S = 0] = \left\{\mathrm{E}\left[\frac{1}{c(X_1)}\middle| S = 0, D = 1\right]\right\}^{-1}$; thus, $\theta = 1/\Pr[D = 1|S = 0]$. Using this result, the definition of $c(X_1)$, and the sampling properties of the design, we can write

$$\frac{1}{c(x_1) \times \theta} = \frac{\Pr[D = 1|S = 0]}{\Pr[D = 1|X_1 = x_1, S = 0]} = \frac{\Pr[D = 1|S = 0]}{\Pr[D = 1|X = x, S = 0]}.$$

It follows that

$$\begin{aligned}
\frac{1}{c(x_1) \times \theta} f(x|S = 0, D = 1) &= \frac{\Pr[D = 1|S = 0]}{\Pr[D = 1|X = x, S = 0]} \times f(x|S = 0, D = 1) \\
&= \frac{\Pr[D = 1|S = 0]}{\Pr[D = 1|X = x, S = 0]} \times \frac{f(x|S = 0)\Pr[D = 1|X = x, S = 0]}{\Pr[D = 1|S = 0]} \\
&= f(x|S = 0),
\end{aligned}$$

which establishes the desired result.

Web Appendix 4 Estimating the conditional probability of trial participation in the nested trial design with subsampling (simple random sampling)

We sketch the proof for the convergence in probability of the estimators for the conditional probability of trial participation described in the main text, without delving into the technical conditions needed to make the arguments more rigorous.

We assume a model $p(X;\gamma)$ for $\Pr[S=1|X]$, where γ is a finite-dimensional parameter. Consider the likelihood function for the nested trial design with a census of the actual population,

$$\mathscr{L}(\gamma)=\prod_{i=1}^{n}\big[p(X_i;\gamma)\big]^{S_i}\big[1-p(X_i;\gamma)\big]^{1-S_i},$$

and, the pseudo-likelihood function for the nested trial design with known sampling probability of the non-randomized individuals,

$$\mathscr{L}_{\text{W}}(\gamma)=\prod_{i=1}^{n}\big[p(X_i;\gamma)\big]^{S_iD_i}\big[1-p(X_i;\gamma)\big]^{[(1-S_i)D_i]/c}.$$

For $\mathscr{L}(\gamma)$, the sample size-normalized objective function to be maximized is

$$\widehat{\ell}(\gamma)=\frac{1}{n}\sum_{i=1}^{n}\Big\{S_i\log p(X_i;\gamma)+(1-S_i)\log\big[1-p(X_i;\gamma)\big]\Big\}.$$

Provided the technical conditions for the uniform law of large numbers hold, the above objective function converges uniformly in probability, in the sense of the definition in Section 2.1 of [2], to

$$\ell_0(\gamma)=\text{E}\Big[S\log p(X;\gamma)+(1-S)\log\big[1-p(X;\gamma)\big]\Big].$$

By Theorem 2.1 of [2], if $\ell_0(\gamma)$ is uniquely maximized at γ_0, the parameter space is compact, and $\ell_0(\gamma)$ is continuous, then the estimator $\widehat{\gamma}$ obtained by maximizing $\widehat{\ell}(\gamma)$, converges in probability to γ_0, that is, $\widehat{\gamma}\xrightarrow{p}\gamma_0$.

For $\mathscr{L}_{\text{W}}(\gamma)$, the sample size-normalized objective function to be maximized is

$$\widehat{\ell}_{\text{W}}(\gamma)=\frac{1}{n}\sum_{i=1}^{n}\left\{S_iD_i\log p(X_i;\gamma)+\frac{(1-S_i)D_i}{c}\log\big[1-p(X_i;\gamma)\big]\right\}.$$

Because c is assumed to be bounded away from 0, and provided the technical conditions for the uniform weak law of large numbers hold, the above objective function converges uniformly in probability to

$$\ell_{\mathrm{w}0}(\gamma) = \mathrm{E}\left[SD \log p(X;\gamma) + \frac{(1-S)D}{c} \log\left[1 - p(X;\gamma)\right]\right].$$

We will now show that $\ell_0(\gamma) = \ell_{\mathrm{w}0}(\gamma)$.

By design, if $S = 1$, then $SD = 1$; if $S = 0$, then $SD = 0$. Thus, to establish the result we only need to show that

$$\mathrm{E}\left[(1-S)\log[1 - p(X;\gamma)\right] = \mathrm{E}\left[\frac{(1-S)D}{c} \log\left[1 - p(X;\gamma)\right]\right].$$

Starting from the right-hand-side,

$$\begin{aligned}
\mathrm{E}\left[\frac{(1-S)D}{c} \log\left[1 - p(X;\gamma)\right]\right] &= \mathrm{E}\left[\log\left[1 - p(X;\gamma)\right] \middle| S = 0, D = 1\right] \frac{\Pr[S = 0, D = 1]}{c} \\
&= \mathrm{E}\left[\log\left[1 - p(X;\gamma)\right] \middle| S = 0\right] \Pr[S = 0] \\
&= \mathrm{E}\left[(1-S)\log\left[1 - p(X;\gamma)\right]\right],
\end{aligned}$$

which establishes the result.

Because $\ell_0(\gamma) = \ell_{\mathrm{w}0}(\gamma)$, it follows that the maximizer of $\widehat{\ell}_{\mathrm{w}}(\gamma)$, $\widehat{\gamma}_{\mathrm{w}}$, converges in probability to γ_0, that is, $\widehat{\gamma}_{\mathrm{w}} \xrightarrow{p} \gamma_0$.

To obtain the asymptotic distribution of the estimators, we need additional technical conditions as in Theorem 3.1 of [2]; provided these conditions hold, $\widehat{\gamma}$ and $\widehat{\gamma}_{\mathrm{w}}$ are asymptotically normal.

Web Appendix 5 Estimating the conditional probability of trial participation in the nested trial design with subsampling (covariate-dependent sampling)

As in the previous section, we assume a model $p(X;\gamma)$ for $\Pr[S=1|X]$, where γ is a finite-dimensional parameter. The weighted pseudo-likelihood function becomes

$$\mathscr{L}_{\mathrm{W}}^{*}(\gamma)=\prod_{i=1}^{n}\big[p(X_i;\gamma)\big]^{S_iD_i}\big[1-p(X_i;\gamma)\big]^{[(1-S_i)D_i]/c(X_{1i})}.$$

Note that the only difference between $\mathscr{L}_{\mathrm{W}}^{*}(\gamma)$ and $\mathscr{L}_{\mathrm{W}}(\gamma)$ is that the weights in the former depend on X_1. The sample size-normalized objective function to be maximized is

$$\widehat{\ell}_{\mathrm{W}}^{*}(\gamma)=\frac{1}{n}\sum_{i=1}^{n}\left\{S_iD_i\log p(X_i;\gamma)+\frac{(1-S_i)D_i}{c(X_{1i})}\log\big[1-p(X_i;\gamma)\big]\right\}.$$

Because $c(X_1)$ is assumed to be bounded away from 0, and provided the technical conditions for the uniform weak law of large numbers hold, the above objective function converges uniformly in probability to

$$\ell_{\mathrm{W0}}^{*}(\gamma)=\mathrm{E}\left[SD\log p(X;\gamma)+\frac{(1-S)D}{c(X_1)}\log\big[1-p(X;\gamma)\big]\right].$$

We will now show that $\ell_0(\gamma)=\ell_{\mathrm{W0}}^{*}(\gamma)$.

By design, if $S=1$, then $SD=1$; if $S=0$, then $SD=0$. Thus, to establish the result we only need to show that

$$\mathrm{E}\Big[(1-S)\log\big[1-p(X;\gamma)\big]\Big]=\mathrm{E}\left[\frac{(1-S)D}{c(X_1)}\log\big[1-p(X;\gamma)\big]\right].$$

Starting from the right-hand-side,

$$\begin{aligned}
\mathrm{E}\left[\frac{(1-S)D}{c(X_1)}\log\big[1-p(X;\gamma)\big]\right] &= \mathrm{E}\left[\mathrm{E}\left[\frac{(1-S)D}{c(X_1)}\log\big[1-p(X;\gamma)\big]\Big|X_1\right]\right] \\
&= \mathrm{E}\left[\frac{1}{c(X_1)}\,\mathrm{E}\big[(1-S)D\log\big[1-p(X;\gamma)\big]\big|X_1\big]\right] \\
&= \mathrm{E}\left[\frac{\Pr[S=0,D=1|X_1]}{c(X_1)}\,\mathrm{E}\big[\log\big[1-p(X;\gamma)\big]\big|X_1,S=0,D=1\big]\right] \\
&= \mathrm{E}\left[\Pr[S=0|X_1]\,\mathrm{E}\big[\log\big[1-p(X;\gamma)\big]\big|X_1,S=0,D=1\big]\right] \\
&= \mathrm{E}\left[\Pr[S=0|X_1]\,\mathrm{E}\big[\log\big[1-p(X;\gamma)\big]\big|X_1,S=0\big]\right] \\
&= \mathrm{E}\left[\mathrm{E}\big[(1-S)\log\big[1-p(X;\gamma)\big]\big|X_1\big]\right] \\
&= \mathrm{E}\Big[(1-S)\log\big[1-p(X;\gamma)\big]\Big],
\end{aligned}$$

which establishes the result.

Because $\ell_0(\gamma)=\ell^*_{\mathrm{W}0}(\gamma)$, it follows that the maximizer of $\widehat{\ell}^*_{\mathrm{W}}(\gamma)$, $\widehat{\gamma}^*_{\mathrm{W}}$, converges in probability to γ_0, that is, $\widehat{\gamma}^*_{\mathrm{W}} \xrightarrow{p} \gamma_0$.

In practical terms, this result suggests that the conditional probability of trial participation in the target population can be estimated using a weighted regression of S on X among sampled patients, using weights equal to 1 for randomized patients (all of whom are sampled); $1/c(X_1)$ for sampled non-randomized individuals; 0 for non-sampled non-randomized individuals.

As above, provided the technical conditions of Theorem 3.1 of [2] hold, $\widehat{\gamma}^*_{\mathrm{W}}$ is asymptotically normal.

Web Appendix References

[1] I. J. Dahabreh, M. A. Hernán, S. E. Robertson, A. Buchanan, and J. A. Steingrimsson, "Generalizing trial findings in nested trial designs with sub-sampling of non-randomized individuals," *arXiv preprint arXiv:1902.06080v2*, 2019 (accessed: 11/03/2020).

[2] W. K. Newey and D. McFadden, "Large sample estimation and hypothesis testing," in *Handbook of econometrics, volume 4* (R. F. Engle and D. L. McFadden, eds.), pp. 2111–2245, Amsterdam, The Netherlands: Elsevier B.V., 1994.