Received: 21 December 2021 | Accepted: 10 October 2022

DOI: 10.1111/biom.13796

BIOMETRIC METHODOLOGY

Biometrics A JOURNAL OF THE INTERNATIONAL BIOMETRIC SOCIETY WILEY

Estimating the area under the ROC curve when transporting a prediction model to a target population

Bing Li[1] | **Constantine Gatsonis**[1] | **Issa J. Dahabreh**[2,3] | **Jon A. Steingrimsson**[1]

[1]Department of Biostatistics, Brown University, Providence, Rhode Island, USA

[2]CAUSALab, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA

[3]Departments of Epidemiology and Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA

Correspondence
Bing Li, Department of Biostatistics, Brown University, Providence, RI, USA.
Email: bing_li1@brown.edu

Funding information
Patient-Centered Outcomes Research Institute, Grant/Award Numbers: ME-1502-27794, ME-2019C3-17875; National Cancer Institute, Grant/Award Numbers: U10CA180794, U10CA180820; National Institute of General Medical Sciences, Grant/Award Number: U54GM115677

Abstract

We propose methods for estimating the area under the receiver operating characteristic (ROC) curve (AUC) of a prediction model in a target population that differs from the source population that provided the data used for original model development. If covariates that are associated with model performance, as measured by the AUC, have a different distribution in the source and target populations, then AUC estimators that only use data from the source population will not reflect model performance in the target population. Here, we provide identification results for the AUC in the target population when outcome and covariate data are available from the sample of the source population, but only covariate data are available from the sample of the target population. In this setting, we propose three estimators for the AUC in the target population and show that they are consistent and asymptotically normal. We evaluate the finite-sample performance of the estimators using simulations and use them to estimate the AUC in a nationally representative target population from the National Health and Nutrition Examination Survey for a lung cancer risk prediction model developed using source population data from the National Lung Screening Trial.

KEYWORDS

covariate shift, domain adaptation, importance weighting, model performance, prediction models, transportability, U-processes

1 | INTRODUCTION

Users of prediction models are typically interested in deploying the models in some target population of substantive interest. Yet, in most cases, the data used to develop the models are sampled from a source population that is not the same as the target population (e.g., from a different health-care system or geographic region). Consequently, the samples from the source and target population typically have different underlying distributions. In particular, prediction error modifiers, that is, variables that affect model performance for a given prediction model and a measure of its performance (Steingrimsson et al., 2021), are likely to have a different distribution across the two populations. When that is the case, measures of model performance estimated in data from the source population will not reflect model performance in the target population.

Here, we consider the estimation of measures of model performance in the target population, when covariate data are available from the samples of both the source and target population, but outcome data are only available from the sample of the source population. Outcome data from the target population are often unavailable (e.g., because they require the use of specialized tests) or unsuitable for model assessment (e.g., when follow-up duration is inadequate to capture the events of interest or when outcomes are not ascertained using the same definitions as in the source

population). In such cases, we need to combine information from the source and target populations to estimate model performance in the latter; estimating measures of model performance in the target population and quantifying uncertainty about them requires "transportability" methods appropriate for measures of model performance (Steingrimsson et al., 2021).

Transportability methods for classification accuracy and loss-based model performance measures (e.g., mean squared error) have been studied in the machine learning literature on domain adaption (Ben-David et al., 2007, 2010), covariate shift (Shimodaira, 2000; Sugiyama et al., 2007, 2013), and transfer learning (Long et al., 2015). Ben-David et al. (2007, 2010) developed uniform convergence bounds on the target population misclassification error and Sugiyama et al. (2007, 2013) proposed reweighting observations to obtain an unbiased estimator for the target population risk (expected loss). In the statistical literature, research on transportability methods has focused on average treatment effect estimation (Cole & Stuart, 2010; Dahabreh et al., 2019, 2020; Lu et al., 2019; Westreich et al., 2017), a task that involves different estimands, identifiability conditions, and estimation procedures compared to transportability methods for assessing model performance.

The classification accuracy of prediction models is commonly evaluated using the area under the receiver operating characteristic curve, often referred to as the "area under the curve" (AUC) (Bamber, 1975; Hanley & McNeil, 1982; McNeil & Hanley, 1984). The AUC can be interpreted as the probability that a randomly sampled observation with the outcome has a higher predicted value than a randomly sampled observation without the outcome. As a bipartite ranking function, the AUC is defined on pairs of observations, rather than a single (independent) observation, complicating any theoretical study of AUC estimators (DeLong et al., 1988; Wieand et al., 1989; Zhou et al., 2009). The only work on estimating the AUCs in a target population that we are aware of (Agarwal et al., 2005; Usunier et al., 2005), developed generalization bounds, not estimation procedures, under the rather implausible assumption that the joint distribution of the outcome and covariates is the same in the target and source population. Despite the frequent use of AUCs for evaluation of diagnostic tests or prediction models, no method has been proposed for estimating the AUC in a target population that has a different data distribution compared to the source population.

In this paper, we provide identifiability results for the AUC in the target population, when covariate data are available from the samples of both the source and target population, but outcome data are only available from the sample of the source population. We develop inverse-odds weighting (IOW), outcome model-based, and doubly robust (DR) estimators for the AUC in the target population allowing for differences in the covariate distribution between the target and source population. We show that the estimators are $\sqrt{n}$ consistent and asymptotically normal and evaluate their finite-sample performance using simulations. We apply the methods to estimate the AUC of a lung cancer risk prediction model built using data from the National Lung Screening Trial (NLST) (Team, 2011) and using target population data from the subset of the National Health and Nutrition Examination Survey (NHANES) that met the eligibility criteria for NLST. As for most clinical trials, the NLST participants were not selected from a population via any kind of formal sampling, and thus, it is unclear what population they represent. In contrast, the NHANES is specifically designed to be representative of noninstitutionalized U.S. adults to whom a model developed using NLST data may be applied.

2 | DATA AND TARGET PARAMETERS

Let Y be a binary outcome; $\boldsymbol{X}$ a covariate vector taking values in $\mathcal{X}$; and S an indicator of an observation being in the source population (i.e., $S = 1$ if from source population and $S = 0$ if from target population). In our setting, covariate and outcome data are available from a random sample of the source population, $\{(\boldsymbol{X}_i, S_i = 1, Y_i) : i = 1, \ldots, n_1\}$; in addition, covariate data (but not outcome data) are available from a separately obtained random sample from the target population, $\{(\boldsymbol{X}_i, S_i = 0) : i = 1, \ldots, n_0\}$. Such a "nonnested" design (Dahabreh et al., 2021; Steingrimsson et al., 2021), where the source population data and the target population data are sampled separately, is commonly used in studies examining the performance of a prediction model in a new target population. We combine the data from the source and target population to form a "composite" dataset $\mathcal{O} = \{\boldsymbol{O}_i = (\boldsymbol{X}_i, S_i, S_i \times Y_i), i = 1, \ldots, n = n_1 + n_0\}$ that consists of the sample from the source population and the sample from the target population (Dahabreh et al., 2021) and denote the sample size of this "composite" dataset by n. We use the notation $S \times Y$ to indicate that Y is only available in the source population ($S = 1$).

We model the data as if there is a single (near-)infinite superpopulation, and the samples from the source and the target population come from nonoverlapping strata of the superpopulation, with unknown sampling probabilities. We do not assume that the source population sample is obtained through a formal sampling process (i.e., the source population data can be obtained from a nonprobability sample (Chen et al., 2020)). Nevertheless, we assume that the sample from the source population is obtained

from some, potentially not fully characterized, stratum of the superpopulation. In contrast, we assume that the sample from the target population is a random sample of a well-defined stratum of the superpopulation that is of substantive interest. The superpopulation approach is commonly applied when analyzing study data from randomized or observational studies (Robins, 1988) and is natural for transportability analyses (Dahabreh & Hernán, 2019; Dahabreh et al., 2021). The sampling framework relates, but is not identical, to the framework for inference from nonprobability samples proposed by Elliott and Valliant (2017).

Let $\boldsymbol{X}^*$ be a vector that contains a subset of the covariates in $\boldsymbol{X}$ and let $g(\boldsymbol{X}^*;\boldsymbol{\beta})$ be a prediction model for $\Pr[Y = 1|S = 1,\boldsymbol{X}^*]$ indexed by some parameter $\boldsymbol{\beta}$. We wish to estimate the AUC of $g(\boldsymbol{X}^*;\boldsymbol{\beta})$ in the target population. We assume that the model $g(\boldsymbol{X}^*;\boldsymbol{\beta})$ is fit on data that are independent of $\mathcal{O}$. This setup accommodates a number of different scenarios for estimating the AUCs in a target population that occur in practice: evaluating the performance of an existing prediction model (e.g., the PLCOm2012 model for lung cancer risk (Tammemägi et al., 2013)) in a new target population; examining the performance of a novel biomarker in the target population; or internally validating a newly developed prediction model using a hold-out test dataset or cross-validation.

The target parameter of interest is the AUC in the target population defined as

$$\tau_0 = \mathrm{E}\left[I\left(g(\boldsymbol{X}_i^*;\boldsymbol{\beta}) > g(\boldsymbol{X}_j^*;\boldsymbol{\beta})\right)|Y_i = 1, Y_j = 0, S_i = 0, S_j = 0\right], \tag{1}$$

where i is a random observation from the target population that has the outcome and j is a random observation from the target population without the outcome and the expectation is over the population underlying the target population sample.

An alternative sampling design is a "nested" design (Dahabreh et al., 2021), where the source population is embedded within a larger target population. For nested designs, a natural target parameter is

$$\tau_{0,nested} = \mathrm{E}\left[I\left(g(\boldsymbol{X}_i^*;\boldsymbol{\beta}) > g(\boldsymbol{X}_j^*;\boldsymbol{\beta})\right)|Y_i = 1, Y_j = 0\right], \tag{2}$$

where the expectation is over the population underlying the target population sample. In Supplementary Web Appendix A, we provide additional details on the nested sampling design and show that the target parameter is identifiable and propose two estimators for $\tau_{0,nested}$. In the remainder of the main text of the manuscript, we focus on nonnested designs because they are more commonly encountered in prediction modeling practice.

3 | IDENTIFICATION

The following conditions are sufficient for identifying the AUC in the target population using the observed data:

A1. Positivity conditions: (i) $\Pr[S = 1|\boldsymbol{X} = \boldsymbol{x}] > 0$, for all $\boldsymbol{x}$ that have positive density in the target population, $f_{\boldsymbol{X},S}(\boldsymbol{X} = \boldsymbol{x}, S = 0) \neq 0$; (ii) $\mathrm{E}[\Pr[Y = 1|\boldsymbol{X}_i, S = 1](1 - \Pr[Y = 1|\boldsymbol{X}_j, S = 1])|S_i = 0, S_j = 0] > 0$, where i is a random observation from the target population that has the outcome and j is a random observation from the target population without the outcome.

A2. Conditional independence of the outcome and data source: $Y \perp\!\!\!\perp S|\boldsymbol{X}$.

Positivity condition A1 (i) states that all covariate patterns that have a positive density in the target population have a positive probability of appearing in the sample from the source population. This implies that the source population has at least as broad of a spectrum of participants as the target population. Positivity condition A1 (ii) is a mild condition ensuring that there is sufficient variability in the outcomes to allow for estimating AUC in the target population. The conditional independence condition (A2) implies that for every $\boldsymbol{x}$ with positive density in the target population $f_{\boldsymbol{X},S}(\boldsymbol{X} = \boldsymbol{x}, S = 0) \neq 0$, $\Pr[Y = 1|\boldsymbol{X} = \boldsymbol{x}, S = 1] = \Pr[Y = 1|\boldsymbol{X} = \boldsymbol{x}, S = 0]$, which is untestable using the observed data in our setup. Informally, we assume that the "true" relationship between Y and $\boldsymbol{X}$ is the same in both populations, but we do not assume that the model $g(\boldsymbol{X}^*;\boldsymbol{\beta})$ is correctly specified. This implies that the methods we develop in Section 4 can estimate the AUC in the target population of both correctly and incorrectly specified models. Last, because $\boldsymbol{X}^*$ may contain a proper subset of the covariates in $\boldsymbol{X}$, we accommodate cases where the set of covariates needed to satisfy the conditional exchangeability assumption is richer than the set of covariates included in the prediction model.

The following theorem states that, if conditions A1 and A2 hold, the AUC in the target population is identifiable using outcome and covariate information from a sample from the source population, but only covariate information from a sample from the target population.

Theorem 1. *Under identifiability conditions A1 and A2, the AUC in the target population is identified by the observed data functional*

$$\tau_0 = \frac{\mathrm{E}[w(\boldsymbol{X}_k,\boldsymbol{X}_l)I(g(\boldsymbol{X}_k^*;\boldsymbol{\beta}) > g(\boldsymbol{X}_l^*;\boldsymbol{\beta}), Y_k = 1, Y_l = 0)|S_k = 1, S_l = 1]}{\mathrm{E}[w(\boldsymbol{X}_k,\boldsymbol{X}_l)I(Y_k = 1, Y_l = 0)|S_k = 1, S_l = 1]}. \tag{3}$$

Here, the subscripts k and l denote a random pair of observations from the source population and for a pair of covariate vectors $\boldsymbol{X}^\dagger$ and $\boldsymbol{X}^\ddagger$, we define

$$w(\boldsymbol{X}^\dagger, \boldsymbol{X}^\ddagger) = \frac{\Pr[S=0|\boldsymbol{X}^\dagger]\Pr[S=0|\boldsymbol{X}^\ddagger]}{\Pr[S=1|\boldsymbol{X}^\dagger]\Pr[S=1|\boldsymbol{X}^\ddagger]}. \quad (4)$$

Furthermore, τ_0 can be equivalently written as

$$\tau_0 = \frac{\mathrm{E}[m(\boldsymbol{X}_i, \boldsymbol{X}_j) I(g(\boldsymbol{X}_i^*;\boldsymbol{\beta}) > g(\boldsymbol{X}_j^*;\boldsymbol{\beta}))|S_i=0, S_j=0]}{\mathrm{E}[m(\boldsymbol{X}_i, \boldsymbol{X}_j)|S_i=0, S_j=0]}, \quad (5)$$

where the subscripts i and j denote a random pair of observations from the target population and for a pair of covariate vectors $\boldsymbol{X}^\dagger$ and $\boldsymbol{X}^\ddagger$, we define

$$m(\boldsymbol{X}^\dagger, \boldsymbol{X}^\ddagger) = \Pr[Y=1|S=1, \boldsymbol{X}^\dagger]\Pr[Y=0|S=1, \boldsymbol{X}^\ddagger]. \quad (6)$$

It is likely that the target and source population data are two separate samples from their underlying populations with unknown sampling fractions. Under this "nonnested" sampling design, $\Pr[S=1]$, $\Pr[S=1|\boldsymbol{X}]$, and $w(\boldsymbol{X}_k, \boldsymbol{X}_l)$ are unidentifiable as they depend on the unknown sampling fractions. Nevertheless, the AUC in the target population is still identifiable. Specifically, using D to denote inclusion in the composite dataset (either from the source or the target population) and define

$$w_D(\boldsymbol{X}^\dagger, \boldsymbol{X}^\ddagger) = \frac{\Pr[S=0|\boldsymbol{X}^\dagger, D=1]\Pr[S=0|\boldsymbol{X}^\ddagger, D=1]}{\Pr[S=1|\boldsymbol{X}^\dagger, D=1]\Pr[S=1|\boldsymbol{X}^\ddagger, D=1]}, \quad (7)$$

we show in Supplementary Web Appendix B that

$$w_D(\boldsymbol{X}^\dagger, \boldsymbol{X}^\ddagger) \propto w(\boldsymbol{X}^\dagger, \boldsymbol{X}^\ddagger). \quad (8)$$

This proportionality result indicates that we can use the identifiable inverse-odds weights among sampled observations $w_D(\boldsymbol{X}_k, \boldsymbol{X}_l)$ in place of the unidentifiable $w(\boldsymbol{X}_k, \boldsymbol{X}_l)$, in expression (3). More generally, this implies that the AUC in the target population is identifiable under the biased sampling design induced by the "nonnested" sampling and the expectations in expressions (3) and (5) can be interpreted as integrating with respect to densities from the biased sampling model (see Dahabreh et al., 2019 and Dahabreh et al., 2021 for more discussion on the biased sampling model).

4 | ESTIMATION OF AUC IN THE TARGET POPULATION

Expression (3) suggests the IOW estimator

$$\widehat{\tau}_{\mathrm{IOW}} = \frac{\sum_{i\neq j} w(\boldsymbol{X}_i, \boldsymbol{X}_j; \widehat{\boldsymbol{\gamma}}) I(g(\boldsymbol{X}_i^*;\boldsymbol{\beta}) > g(\boldsymbol{X}_j^*;\boldsymbol{\beta}), Y_i=1, Y_j=0, S_i=1, S_j=1)}{\sum_{i\neq j} w(\boldsymbol{X}_i, \boldsymbol{X}_j; \widehat{\boldsymbol{\gamma}}) I(Y_i=1, Y_j=0, S_i=1, S_j=1)}, \quad (9)$$

where the estimator for the inverse-odds weight of a pair of observations is

$$w(\boldsymbol{X}_i, \boldsymbol{X}_j; \widehat{\boldsymbol{\gamma}}) = \frac{[1-\pi(\boldsymbol{X}_i; \widehat{\boldsymbol{\gamma}})][1-\pi(\boldsymbol{X}_j; \widehat{\boldsymbol{\gamma}})]}{\pi(\boldsymbol{X}_i; \widehat{\boldsymbol{\gamma}})\pi(\boldsymbol{X}_j; \widehat{\boldsymbol{\gamma}})}. \quad (10)$$

Here, $\pi(\boldsymbol{X}; \widehat{\boldsymbol{\gamma}})$ is an estimator based on a model for the conditional expectation of the source population indicator indexed by a parameter $\boldsymbol{\gamma}$. We will use $\boldsymbol{\gamma}_0$ to denote the "true" value of $\boldsymbol{\gamma}$. We assume that the parameter space of $\boldsymbol{\gamma}$, denoted by $\boldsymbol{\Gamma}$, is a bounded Euclidean space and $\inf_{\boldsymbol{x}\in\mathcal{X},\boldsymbol{\gamma}\in\boldsymbol{\Gamma}} \pi(\boldsymbol{x};\boldsymbol{\gamma}) > 0$, $\sup_{\boldsymbol{x}\in\mathcal{X},\boldsymbol{\gamma}\in\boldsymbol{\Gamma}} \pi(\boldsymbol{x};\boldsymbol{\gamma}) < 1$. The IOW estimator $\widehat{\tau}_{\mathrm{IOW}}$ is constructed by weighting pairs of observations according to their odds of being from the target population. As we show in Section 5, consistency of $\widehat{\tau}_{\mathrm{IOW}}$ relies on correctly specifying the model $\pi(\boldsymbol{X};\boldsymbol{\gamma})$.

Another estimator, based on expression (5), is

$$\widehat{\tau}_{\mathrm{OM}} = \frac{\sum_{i\neq j} m(\boldsymbol{X}_i, \boldsymbol{X}_j; \widehat{\boldsymbol{\alpha}}) I(g(\boldsymbol{X}_i^*;\boldsymbol{\beta}) > g(\boldsymbol{X}_j^*;\boldsymbol{\beta}), S_i=0, S_j=0)}{\sum_{i\neq j} m(\boldsymbol{X}_i, \boldsymbol{X}_j; \widehat{\boldsymbol{\alpha}}) I(S_i=0, S_j=0)}, \quad (11)$$

where $m(\boldsymbol{X}_i, \boldsymbol{X}_j; \widehat{\boldsymbol{\alpha}}) = q(\boldsymbol{X}_i; \widehat{\boldsymbol{\alpha}})(1 - q(\boldsymbol{X}_j; \widehat{\boldsymbol{\alpha}}))$. Here, $q(\boldsymbol{X}; \widehat{\boldsymbol{\alpha}})$ is an estimator based on an outcome model $q(\boldsymbol{X};\boldsymbol{\alpha})$ for $\Pr[Y=1|\boldsymbol{X}, S=1]$, indexed by a parameter $\boldsymbol{\alpha}$. As above, we will use $\boldsymbol{\alpha}_0$ to denote the "true" value of $\boldsymbol{\alpha}$. We assume that the parameter space of $\boldsymbol{\alpha}$, denoted by $\mathcal{A}$, is a bounded Euclidean space and $\inf_{\boldsymbol{x}\in\mathcal{X},\boldsymbol{\alpha}\in\mathcal{A}} q(\boldsymbol{x};\boldsymbol{\alpha}) > 0$, $\sup_{\boldsymbol{x}\in\mathcal{X},\boldsymbol{\alpha}\in\mathcal{A}} q(\boldsymbol{x};\boldsymbol{\alpha}) < 1$. The validity of this condition depends on the complexity of the outcome model. When complex data-adaptive models violate that condition, less complex models that provide more smoothing can be useful (Cole & Hernán, 2008; Petersen et al., 2012). As shown in Section 5, consistency of the outcome model estimator $\widehat{\tau}_{\mathrm{OM}}$ relies on correctly specifying the outcome model $q(\boldsymbol{X};\boldsymbol{\alpha})$.

Now we describe a DR estimator that combines a model for the probability of being from the source population, $\pi(\boldsymbol{X};\boldsymbol{\gamma})$, with a model for the outcome, $q(\boldsymbol{X};\boldsymbol{\alpha})$. For a function $k(\boldsymbol{O}_i, \boldsymbol{O}_j)$, define

$$d^{\mathrm{IOW}}(\boldsymbol{O}_i, \boldsymbol{O}_j; \boldsymbol{\gamma}, k(\boldsymbol{O}_i, \boldsymbol{O}_j)) = w(\boldsymbol{X}_i, \boldsymbol{X}_j; \boldsymbol{\gamma}) I(Y_i=1, Y_j=0, S_i=1, S_j=1) k(\boldsymbol{O}_i, \boldsymbol{O}_j), \quad (12)$$

which is the contribution for the pair of observations (i, j) in the sum for the numerator of the IOW estimator if $k(\boldsymbol{O}_i, \boldsymbol{O}_j) = I(g(\boldsymbol{X}_i^*;\boldsymbol{\beta}) > g(\boldsymbol{X}_j^*;\boldsymbol{\beta}))$, and is that for the denominator if $k(\boldsymbol{O}_i, \boldsymbol{O}_j) = 1$.

Similarly, define

$$d^{\mathrm{OM}}(\boldsymbol{O}_i, \boldsymbol{O}_j; \boldsymbol{\alpha}, k(\boldsymbol{O}_i, \boldsymbol{O}_j)) = m(\boldsymbol{X}_i, \boldsymbol{X}_j; \boldsymbol{\alpha}) I(S_i=0, S_j=0) k(\boldsymbol{O}_i, \boldsymbol{O}_j), \quad (13)$$

which is the contribution for the pair of observations (i, j) in the sum for the numerator of the outcome

model estimator, if $k(\boldsymbol{O}_i, \boldsymbol{O}_j) = I(g(\boldsymbol{X}_i^*; \boldsymbol{\beta}) > g(\boldsymbol{X}_j^*; \boldsymbol{\beta}))$, and is the contribution for the pair of observations (i, j) for the denominator of the outcome model estimator, if $k(\boldsymbol{O}_i, \boldsymbol{O}_j) = 1$.

Using this notation, the DR estimator is defined as

$$\widehat{\tau}_{\mathrm{DR}} = \frac{\sum_{i \neq j} d^{\mathrm{DR}}(\boldsymbol{O}_i, \boldsymbol{O}_j; \widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\alpha}}, k(\boldsymbol{O}_i, \boldsymbol{O}_j) = I(g(\boldsymbol{X}_i^*; \boldsymbol{\beta}) > g(\boldsymbol{X}_j^*; \boldsymbol{\beta})))}{\sum_{i \neq j} d^{\mathrm{DR}}(\boldsymbol{O}_i, \boldsymbol{O}_j; \widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\alpha}}, k(\boldsymbol{O}_i, \boldsymbol{O}_j) = 1)}, \tag{14}$$

where

$$\begin{aligned} d^{\mathrm{DR}}(\boldsymbol{O}_i, \boldsymbol{O}_j; \widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\alpha}}, k(\boldsymbol{O}_i, \boldsymbol{O}_j)) &= d^{\mathrm{IOW}}(\boldsymbol{O}_i, \boldsymbol{O}_j; \widehat{\boldsymbol{\gamma}}, k(\boldsymbol{O}_i, \boldsymbol{O}_j)) \\ &+ d^{\mathrm{OM}}(\boldsymbol{O}_i, \boldsymbol{O}_j; \widehat{\boldsymbol{\alpha}}, k(\boldsymbol{O}_i, \boldsymbol{O}_j)) \\ &- w(\boldsymbol{X}_i, \boldsymbol{X}_j; \widehat{\boldsymbol{\gamma}}) m(\boldsymbol{X}_i, \boldsymbol{X}_j; \widehat{\boldsymbol{\alpha}}) I(S_i \\ &= 1, S_j = 1) k(\boldsymbol{O}_i, \boldsymbol{O}_j). \end{aligned} \tag{15}$$

As we show in the next section, the DR estimator offers some robustness to misspecification of either $\pi(\boldsymbol{X}; \boldsymbol{\gamma})$ or $q(\boldsymbol{X}; \boldsymbol{\alpha})$ by providing "two opportunities" for valid inference.

5 | LARGE-SAMPLE PROPERTIES

5.1 | Consistency

The following theorem states that the DR estimator is consistent. A detailed proof is provided in Supplementary Web Appendix C.4.

Theorem 2 (Consistency of $\widehat{\tau}_{\mathrm{DR}}$). *Suppose that both* $\{\frac{1}{n}\sum_{i=1}^n \pi(\boldsymbol{X}_i; \boldsymbol{\gamma}), \boldsymbol{\gamma} \in \boldsymbol{\Gamma}\}$ *and* $\{\frac{1}{n}\sum_{i=1}^n q(\boldsymbol{O}_i; \boldsymbol{\alpha}), \boldsymbol{\alpha} \in \mathcal{A}\}$ *are stochastically equicontinuous and let* $\boldsymbol{\gamma}^*$ *and* $\boldsymbol{\alpha}^*$ *denote the asymptotic limits of* $\widehat{\boldsymbol{\gamma}}$ *and* $\widehat{\boldsymbol{\alpha}}$*, respectively. If either* $\pi(\boldsymbol{X}; \boldsymbol{\gamma}^*) = \Pr[S = 1|\boldsymbol{X}]$ *or* $q(\boldsymbol{X}; \boldsymbol{\alpha}^*) = \Pr[Y = 1|\boldsymbol{X}, S = 1]$*, then* $\widehat{\tau}_{DR}$ *converges in probability to the true AUC in the target population, that is,* $\widehat{\tau}_{DR} \xrightarrow{P} \tau_0$.

Theorem 2 highlights that the DR estimator is consistent if at least one of the models $\pi(\boldsymbol{X}; \boldsymbol{\gamma})$ or $q(\boldsymbol{X}; \boldsymbol{\alpha})$ are correctly specified. Analogous consistency results for the IOW and outcome model estimators are provided in Supplementary Web Appendices C.2 and C.3, respectively. Consistency of all three estimators requires stochastic equicontinuity of the outcome model and/or the model for the conditional expectation of the source population indicator. Stochastic equicontinuity holds for generalized linear models or other models that belong to Lipschitz continuous function classes (see Supplementary Web Appendix C.1 for more details).

5.2 | Asymptotic normality

To derive the asymptotic distribution of the DR estimator, we write the estimator as the solution to an estimating equation. Define the U-process (Hoeffding, 1961; Mao, 2018) as

$$U_n(\tau, \boldsymbol{\gamma}, \boldsymbol{\alpha}) = \frac{1}{n(n-1)} \sum_{i \neq j} l(\boldsymbol{O}_i, \boldsymbol{O}_j; \tau, \boldsymbol{\gamma}, \boldsymbol{\alpha}), \tag{16}$$

where

$$\begin{aligned} l(\boldsymbol{O}_i, \boldsymbol{O}_j; \tau, \boldsymbol{\gamma}, \boldsymbol{\alpha}) &= d^{\mathrm{DR}}(\boldsymbol{O}_i, \boldsymbol{O}_j; \boldsymbol{\gamma}, \boldsymbol{\alpha}, 1)\tau - d^{\mathrm{DR}} \\ &(\boldsymbol{O}_i, \boldsymbol{O}_j; \boldsymbol{\gamma}, \boldsymbol{\alpha}, I(g(\boldsymbol{X}_i^*; \boldsymbol{\beta}) > g(\boldsymbol{X}_j^*; \boldsymbol{\beta}))). \end{aligned} \tag{17}$$

The DR estimator $\widehat{\tau}_{\mathrm{DR}}$ is obtained by solving the estimating equation $U_n(\tau, \widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\alpha}}) = 0$ for τ. To derive the asymptotic distribution of $\widehat{\tau}_{\mathrm{DR}}$, we assume that $\widehat{\boldsymbol{\gamma}}$ and $\widehat{\boldsymbol{\alpha}}$ are solutions to estimating equations. More precisely, we assume that $\widehat{\boldsymbol{\gamma}}$ is defined as the solution to the estimating equation $\sum_{i=1}^n h_1(\boldsymbol{O}_i; \boldsymbol{\gamma}) = 0$ and $\widehat{\boldsymbol{\alpha}}$ is defined as a solution to the estimating equation $\sum_{i=1}^n h_2(\boldsymbol{O}_i; \boldsymbol{\alpha}) = 0$. The following theorem states that the DR estimator is asymptotically normal; a proof is provided in Supplementary Web Appendix D.1.

Theorem 3 (Asymptotic distribution for the DR estimator). *Assume that* $\{\frac{1}{\sqrt{n}}\sum_{i=1}^n \pi(\boldsymbol{X}_i; \boldsymbol{\gamma}), \boldsymbol{\gamma} \in \boldsymbol{\Gamma}\}$ *and* $\{\frac{1}{\sqrt{n}}\sum_{i=1}^n q(\boldsymbol{O}_i; \boldsymbol{\alpha}), \boldsymbol{\alpha} \in \mathcal{A}\}$ *are both stochastically equicontinuous and assume that assumptions D2 and D3 in Lemma 3 in Supplementary Web Appendix D.1 are satisfied. Let* $\boldsymbol{\gamma}^*$ *and* $\boldsymbol{\alpha}^*$ *denote the asymptotic limits of* $\widehat{\boldsymbol{\gamma}}$ *and* $\widehat{\boldsymbol{\alpha}}$*, respectively. If either* $\pi(\boldsymbol{X}; \boldsymbol{\gamma}^*) = \Pr[S = 1|\boldsymbol{X}]$ *or* $q(\boldsymbol{X}; \boldsymbol{\alpha}^*) = \Pr[Y = 1|\boldsymbol{X}, S = 1]$*, then*

$$\sqrt{n}(\widehat{\tau}_{DR} - \tau_0) \xrightarrow{D} \mathcal{N}\left(0, Var\left[\rho(\boldsymbol{\gamma}^*, \boldsymbol{\alpha}^*)^{-1} Q(\boldsymbol{O}; \tau_0, \boldsymbol{\gamma}^*, \boldsymbol{\alpha}^*)\right]\right), \tag{18}$$

where $\xrightarrow{D}$ *denotes convergence in distribution. Here,*

$$\rho(\boldsymbol{\gamma}, \boldsymbol{\alpha}) = \mathrm{E}[d^{\mathrm{DR}}(\boldsymbol{O}_i, \boldsymbol{O}_j; \boldsymbol{\gamma}, \boldsymbol{\alpha}, 1)], \tag{19}$$

with

$$\begin{aligned} Q(\boldsymbol{O}_i; \tau, \boldsymbol{\gamma}, \boldsymbol{\alpha}) = &\left\{\frac{\partial \mathrm{E}[U_n(\tau, \boldsymbol{\gamma}, \boldsymbol{\alpha})]}{\partial \boldsymbol{\gamma}}\right\}\left\{\mathrm{E}\left[\frac{\partial h_1(\boldsymbol{O}_i; \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}\right]\right\}^{-1} h_1(\boldsymbol{O}_i; \boldsymbol{\gamma}) \\ &+ \left\{\frac{\partial \mathrm{E}[U_n(\tau, \boldsymbol{\gamma}, \boldsymbol{\alpha})]}{\partial \boldsymbol{\alpha}}\right\}\left\{\mathrm{E}\left[\frac{\partial h_2(\boldsymbol{O}_i; \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}\right]\right\}^{-1} h_2(\boldsymbol{O}_i; \boldsymbol{\alpha}) - l^{\dagger} \\ &(\boldsymbol{O}_i; \tau, \boldsymbol{\gamma}, \boldsymbol{\alpha}) \end{aligned} \tag{20}$$

and

$$l^{\dagger}(\boldsymbol{O}_i;\tau,\boldsymbol{\gamma},\boldsymbol{\alpha}) = \mathrm{E}[l(\boldsymbol{O}_i,\boldsymbol{O}_j;\tau,\boldsymbol{\gamma},\boldsymbol{\alpha})|\boldsymbol{O}_i] + \mathrm{E}[l(\boldsymbol{O}_j,\boldsymbol{O}_i;\tau,\boldsymbol{\gamma},\boldsymbol{\alpha})|\boldsymbol{O}_i]. \tag{21}$$

The variance of $\widehat{\tau}_{\mathrm{DR}}$ can be consistently estimated using the sample analog of the asymptotic variance expression in Theorem 3. If $\partial U_n(\tau,\boldsymbol{\gamma},\boldsymbol{\alpha})/\partial(\boldsymbol{\gamma},\boldsymbol{\alpha})$ converges uniformly to $\partial U(\tau,\boldsymbol{\gamma},\boldsymbol{\alpha})/\partial(\boldsymbol{\gamma},\boldsymbol{\alpha})$, then the variance of $\widehat{\tau}_{\mathrm{DR}}$ can be consistently estimated by $\frac{1}{n\widehat{\rho}^2}\sum_{k=1}^{n}\widehat{Q}_k^2$, where

$$\widehat{\rho} = \frac{1}{n(n-1)}\sum_{i\neq j} d^{\mathrm{DR}}(\boldsymbol{O}_i,\boldsymbol{O}_j;\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\gamma}},1) \tag{22}$$

and

$$\begin{aligned}\widehat{Q}_k =& \left\{\frac{1}{n(n-1)}\sum_{i\neq j}\left.\frac{\partial l(\boldsymbol{O}_i,\boldsymbol{O}_j;\widehat{\tau}_{\mathrm{DR}},\boldsymbol{\gamma},\widehat{\boldsymbol{\alpha}})}{\partial\boldsymbol{\gamma}}\right|_{\gamma=\widehat{\gamma}}\right\}\\ &\left\{\frac{1}{n}\sum_{i=1}^{n}\left.\frac{\partial h_1(\boldsymbol{O}_i;\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}}\right|_{\gamma=\widehat{\gamma}}\right\}^{-1} h_1(\boldsymbol{O}_k;\widehat{\boldsymbol{\gamma}})\\ &+\left\{\frac{1}{n(n-1)}\sum_{i\neq j}\left.\frac{\partial l(\boldsymbol{O}_i,\boldsymbol{O}_j;\widehat{\tau}_{\mathrm{DR}},\widehat{\boldsymbol{\gamma}},\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}}\right|_{\alpha=\widehat{\alpha}}\right\}\\ &\left\{\frac{1}{n}\sum_{i=1}^{n}\left.\frac{\partial h_2(\boldsymbol{O}_i;\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}}\right|_{\alpha=\widehat{\alpha}}\right\}^{-1} h_2(\boldsymbol{O}_k;\widehat{\boldsymbol{\alpha}})\\ &-\frac{1}{n}\left[\sum_{j=1}^{n}\left\{l(\boldsymbol{O}_k,\boldsymbol{O}_j;\widehat{\tau}_{\mathrm{DR}},\widehat{\boldsymbol{\gamma}},\widehat{\boldsymbol{\alpha}}) + l(\boldsymbol{O}_j,\boldsymbol{O}_k;\widehat{\tau}_{\mathrm{DR}},\widehat{\boldsymbol{\gamma}},\widehat{\boldsymbol{\alpha}})\right\}\right].\end{aligned} \tag{23}$$

Analogous theorems for the asymptotic distribution and consistent variance estimators for the IOW estimator and the outcome model estimator are presented in Supplementary Web Appendices D.2 and D.3, respectively.

For completeness, we also give a result about the asymptotic distribution of the solution to the estimating equation in (16), when condition A2 does not hold. Define τ^* as the solution to the population estimating equation $\mathrm{E}[U_n(\tau,\boldsymbol{\gamma}^*,\boldsymbol{\alpha}^*)] = 0$, which we do not assume is necessarily equal to τ_0 (as will be the case when condition A2 does not hold). Lemma 3 in Supplementary Web Appendix D.1 shows that, even when the solution to $\mathrm{E}[U_n(\tau,\boldsymbol{\gamma}^*,\boldsymbol{\alpha}^*)] = 0$ cannot be interpreted as an estimator of the AUC in the target population, the estimator $\widehat{\tau}_{\mathrm{DR}}$ is still asymptotically normal (with an asymptotic distribution now centered around τ^*) and its asymptotic variance can still be consistently estimated by $\sum_{k=1}^{n}\widehat{Q}_k^2/(n\widehat{\rho}^2)$.

6 | SIMULATIONS

We conducted simulations to evaluate the finite-sample performance of the proposed estimators under different scenarios with correct and incorrect specification of the outcome model and/or the model for the conditional expectation of the source population indicator.

6.1 | Simulation settings

We simulated 500 datasets each with 1000 observations. For each dataset, we generated the source and target super populations using the following procedure. We generated 10,000 covariate vectors $\boldsymbol{X} = \boldsymbol{X}^* = (\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{X}_3)$ from a multivariate Gaussian distribution with mean $\mu_{\boldsymbol{X}} = (0.2,-1,2)$ and a variance matrix $\Sigma_{\boldsymbol{X}} = \mathrm{diag}(0.1,0.1,0.1)$. Next, we generated the indicator of being from the superpopulation underlying the source population (S) from a Bernoulli distribution with mean $\pi(\boldsymbol{X})$, where $\mathrm{logit}(\pi(\boldsymbol{X})) = -4.8 - 3.5\boldsymbol{X}_1 + 5\boldsymbol{X}_1^2 + 2.5\boldsymbol{X}_2 + 2\boldsymbol{X}_3$. We generated the outcome Y from a Bernoulli distribution with mean $q(\boldsymbol{X})$, where $\mathrm{logit}(q(\boldsymbol{X})) = 0.2 + 3\boldsymbol{X}_1 - 5\boldsymbol{X}_1^2 + 2\boldsymbol{X}_2 + \boldsymbol{X}_3$. All observations with $S = 1$ were selected in the sample from the source population, resulting in an average sample size of the data from the source population of 697. The sample from the target population was selected as a random sample from the superpopulation underlying the target data with a sampling fraction of 0.0325. The average size of the sample from the target population was 303. The average implied proportion of positive outcomes in the sample from the target population was 53.4% and the average proportion of positive outcomes in the sample from the source population was 47.9%. We estimated the AUC in the target population of a logistic regression model $g(\boldsymbol{X}^*;\boldsymbol{\beta}) = \mathrm{logit}^{-1}(0.2 + 3\boldsymbol{X}_1 - 5\boldsymbol{X}_1^2 + 2\boldsymbol{X}_2 + \boldsymbol{X}_3)$, that is, the "true" outcome model.

With each simulated dataset, we estimated the parameter $\boldsymbol{\alpha}$ in the outcome model $q(\boldsymbol{X};\boldsymbol{\alpha})$ and the parameter $\boldsymbol{\gamma}$ in the model $\pi(\boldsymbol{X};\boldsymbol{\gamma})$ for the conditional expectation of the source population indicator using both correctly specified and misspecified models. The misspecified models for both $q(\boldsymbol{X};\boldsymbol{\alpha})$ and $\pi(\boldsymbol{X};\boldsymbol{\gamma})$ were main effects logistic regression models that omitted the quadratic term for $\boldsymbol{X}_1$.

In Supplementary Web Appendix E, we present additional simulations using a sample size of 2500 (Table E2), and where we evaluated the performance of a fixed but misspecified model in the target population (Table E3). For both modifications, the results show similar patterns as those discussed in the following section.

TABLE 1 Simulation results comparing the performance of the source population AUC estimator $\hat{\tau}_S$, the inverse-odds weighting estimator $\hat{\tau}_{\text{IOW}}$, the outcome model estimator $\hat{\tau}_{\text{OM}}$, and the doubly robust estimator $\hat{\tau}_{\text{DR}}$. We compared the estimators in terms of their average parameter estimates (Mean), average bias (Bias), relative average bias (RB), Monte Carlo standard deviation of the parameter estimates (SD), the ratio of standard error estimates and Monte Carlo standard deviation of the parameter estimates (SE/SD), square root of mean squared errors (SMSE), and coverage rate of 95% Wald confidence intervals (CR). The true value of the AUC in the target population τ_0 was 0.72 (obtained numerically, by averaging target data AUC across 500 simulations)

	Mean	Bias	RB(%)	SD	SE/SD	SMSE	CR(%)
$\hat{\tau}_S$	0.82	0.10	13.87	0.02	1.04	0.10	0.0
Both $\pi(\boldsymbol{X};\boldsymbol{\gamma})$ and $q(\boldsymbol{X};\boldsymbol{\alpha})$ correctly specified							
$\hat{\tau}_{\text{IOW}}$	0.72	0.0	−0.08	0.04	0.93	0.04	92.2
$\hat{\tau}_{\text{OM}}$	0.72	0.0	0.25	0.02	0.98	0.02	95.4
$\hat{\tau}_{\text{DR}}$	0.72	0.0	0.04	0.04	0.93	0.04	93.0
$\pi(\boldsymbol{X};\boldsymbol{\gamma})$ misspecified; $q(\boldsymbol{X};\boldsymbol{\alpha})$ correctly specified							
$\hat{\tau}_{\text{IOW}}$	0.81	0.09	12.51	0.03	0.97	0.09	14.4
$\hat{\tau}_{\text{DR}}$	0.72	0.00	0.20	0.03	1.00	0.03	94.0
$q(\boldsymbol{X};\boldsymbol{\alpha})$ misspecified; $\pi(\boldsymbol{X};\boldsymbol{\gamma})$ correctly specified							
$\hat{\tau}_{\text{OM}}$	0.66	−0.06	−7.97	0.02	0.98	0.06	13.4
$\hat{\tau}_{\text{DR}}$	0.72	0.00	0.08	0.04	0.93	0.04	92.6
Both $\pi(\boldsymbol{X};\boldsymbol{\gamma})$ and $q(\boldsymbol{X};\boldsymbol{\alpha})$ misspecified							
$\hat{\tau}_{\text{DR}}$	0.94	0.22	29.97	0.06	0.92	0.23	1.6

6.2 | Simulation results

Under each simulation scenario, we compared the performance of the three estimators $\hat{\tau}_{\text{IOW}}$, $\hat{\tau}_{\text{OM}}$, and $\hat{\tau}_{\text{DR}}$. For comparison, we also obtained results for the source population AUC estimator ($\hat{\tau}_S$) that only uses data from the source population. We compared all estimators in terms of their bias, standard deviation, square root of mean squared errors, and coverage rate of 95% Wald confidence intervals. Table 1 presents results from the simulations.

The AUC estimator $\hat{\tau}_S$ that only uses source population data was biased and had low coverage rate for a (nominal) 95% Wald confidence interval. In contrast, the estimators we propose behaved better, as would be expected from the theoretical results presented above. Specifically, when both models for $\pi(\boldsymbol{X};\boldsymbol{\gamma})$ and $q(\boldsymbol{X};\boldsymbol{\alpha})$ were correctly specified, all three estimators—$\hat{\tau}_{\text{IOW}}$, $\hat{\tau}_{\text{OM}}$, and $\hat{\tau}_{\text{DR}}$—were essentially unbiased, with standard deviation 0.04, 0.02, and 0.04, respectively. When the model for $\pi(\boldsymbol{X};\boldsymbol{\gamma})$ was misspecified, $\hat{\tau}_{\text{IOW}}$ was biased, whereas $\hat{\tau}_{\text{DR}}$ was unbiased, provided that $q(\boldsymbol{X};\boldsymbol{\alpha})$ was correctly specified. When the model for $q(\boldsymbol{X};\boldsymbol{\alpha})$ was misspecified, $\hat{\tau}_{\text{OM}}$ was biased, whereas $\hat{\tau}_{\text{DR}}$ was unbiased, provided that the model for $\pi(\boldsymbol{X};\boldsymbol{\gamma})$ was correctly specified. When the models for both $\pi(\boldsymbol{X};\boldsymbol{\gamma})$ and $q(\boldsymbol{X};\boldsymbol{\alpha})$ were misspecified, $\hat{\tau}_{\text{DR}}$ was also biased. Overall, the outcome model estimator had the lowest standard deviation. This is not unexpected because in simpler settings, the asymptotic variance of the outcome model estimator with correctly specified parametric models is no greater than the asymptotic variance of DR or IOW estimators (Tan, 2007).

For all three estimators, the average standard error estimate over the simulation runs was close to the Monte Carlo standard deviation, even when the estimators were biased. Wald confidence intervals had near-nominal coverage rate when the models needed for the consistency of the estimators were correctly specified. This suggests that when either the outcome model or the model for the conditional expectation of the source population indicator is correctly specified, the DR estimator can support valid inferences.

7 | EVALUATING A PREDICTION MODEL FOR LUNG CANCER DIAGNOSIS

We applied the proposed methods to estimate the AUC of a lung cancer risk prediction model in a nationally representative target population. Specifically, we estimated a risk prediction model for lung cancer diagnosis using source population data from the NLST (Team, 2011), a large clinical trial that evaluated the effect of screening with low-dose helical computed tomography versus chest radiography on lung-cancer-specific and overall mortality. The results of the trial helped inform national policy for lung cancer screening (Krist et al., 2021; Moyer, 2014), but studies have noted that participants in the NLST and other randomized controlled trials for lung cancer screening tend to differ from the general population recommended for screening (e.g., in terms of age, education, and comorbidities) (Moyer, 2014). We examined the performance of the estimated model in the target population represented in the NHANES.

We randomly split the NLST data into a training set (75%) and a test set (25%). We used the training set to estimate a logistic regression model ($g(\boldsymbol{X}^*;\boldsymbol{\beta})$) based on the PLCOm2012 model (Tammemägi et al., 2013). The outcome was a binary indicator for whether an individual had a diagnosis of lung cancer within 6-years from study enrollment. The predictors in the model ($\boldsymbol{X}^*$) were age, BMI, race, education, personal history of cancer, smoking status, smoking intensity, duration of smoking, and smoking quit time.

In the United States, lung cancer screening is implemented nationwide, so a natural target population for our prediction model would be the population of all individuals for whom screening is recommended, that is, the population of screening candidates. The NHANES is a cross-sectional study designed to assess the health

and nutritional status of noninstitutionalized adults and children in the United States; it is designed to be representative of this population. Hence, the subset of the NHANES participants who are eligible for lung cancer screening provides a reasonable sample from a well-defined target population.

For our analyses, we used the subset of NHANES 2009-2010 participants who also completed a smoking questionnaire. To avoid structural violations of the positivity assumption A1 (i), we selected NHANES participants who satisfied the NLST eligibility criteria (that we could implement in the data). That is, we used the subset of NHANES participants who were aged 55–74 years, had smoking history of at least 30 pack-years, were either current smokers or former smokers who had quit within the past 15 years, and had not previously been diagnosed with lung cancer as the NHANES target population. Because the NLST has been influential in informing guidelines for lung cancer screening, the criteria used to recommend lung cancer screening in the United States are similar to the eligibility criteria of NLST. For example, the 2014 U.S. Preventive Services Task Force (USPSTF) Recommendation Statement guidelines (Moyer, 2014) recommended screening using the same criteria as the NLST eligibility criteria and the 2021 USPSTF criteria slightly lowered the age and pack year limit (Krist et al., 2021). Because each cycle of the NHANES is cross-sectional, no follow-up is available for participants, and thus, we do not have information on the outcome of interest (lung cancer diagnosis within 6 years)—outcome information is unavailable in the sample of the target population. NHANES used a formal survey sampling design resulting in each observation being assigned a sampling weight. The people enrolled in the NLST are part of the target population of people eligible for lung cancer screening in the United States. Hence, there is some potential for nonidentifiable overlap between the NLST and NHANES samples, but the overlap is likely minimal given the large number of individuals eligible for lung cancer screening (roughly nine millions per year; Dyer, 2021).

For simplicity, and because missing data were limited (86.0% of the NLST and 95.9% of the NHANES observations have no missing data, and the censoring probability for the outcome of lung cancer diagnosis is 10.5%), we restricted our analyses to NLST and NHANES participants who had complete information on the outcome and on the covariates listed in Table 2. Overall, our analyses used data from 45,976 NLST participants and 185 NHANES participants. Table 2 shows the distribution of covariates in the NLST and NHANES data (incorporating the sampling weights for the NHANES data). Table 2 shows important differences between the two covariate distributions (e.g., in terms of race, education level, and smoking status). The higher proportion of nonwhite and less educated individuals in the NHANES compared to the NLST is consistent with prior work, showing that these groups are less likely to participate in clinical trials (Murthy et al., 2004; Unger et al., 2013).

We used the NHANES data and the NLST test set data to estimate the AUC in the target population represented by the NHANES using the methods from Section 4. In addition to the predictors used in the prediction model $g(\boldsymbol{X}^*;\boldsymbol{\beta})$, the covariate vector used to satisfy the conditional exchangeability assumption (i.e., $\boldsymbol{X}$) also included gender, age at smoking initiation, average number of cigarettes smoked per day, marital status, pack years (total years smoked × cigarettes per day / 20), and whether the participant has asthma, diabetes, emphysema, heart disease, hypertension, and stroke. Both the outcome model and the model for the conditional expectation of the source population indicator were estimated using elastic net with mixing parameter 0.5, using 10-fold cross-validation to select the elastic net penalization parameter incorporating the sampling weights as described below.

As previously mentioned, the NHANES uses a formal survey sampling design and each observation is assigned a sampling weight that accounts for oversampling, survey nonresponse, and poststratification adjustments. For the NLST data, all observations were assigned a sampling weight of one. To account for the sampling weights in the IOW estimator, we estimated the parameter $\boldsymbol{\gamma}$ using a weighted elastic net estimator that weights each observation by its sampling weight and denote the estimator by $\widetilde{\boldsymbol{\gamma}}$. Define

$$d^{\text{IOW-w}}(\boldsymbol{O}_i,\boldsymbol{O}_j;\boldsymbol{\gamma},k(\boldsymbol{O}_i,\boldsymbol{O}_j)) = w(\boldsymbol{X}_i,\boldsymbol{X}_j;\boldsymbol{\gamma})I(Y_i=1,Y_j = 0,S_i=1,S_j=1)k(\boldsymbol{O}_i,\boldsymbol{O}_j), \quad (24)$$

and the weighted IOW estimator is given by

$$\widehat{\tau}_{\text{IOW-w}} = \frac{\sum_{i\neq j} d^{\text{IOW-w}}\left(\boldsymbol{O}_i,\boldsymbol{O}_j;\widetilde{\boldsymbol{\gamma}},I\left(g(\boldsymbol{X}_i^*;\boldsymbol{\beta})>g(\boldsymbol{X}_j^*;\boldsymbol{\beta})\right)\right)}{\sum_{i\neq j} d^{\text{IOW-w}}(\boldsymbol{O}_i,\boldsymbol{O}_j;\widetilde{\boldsymbol{\gamma}},1)}. \quad (25)$$

The weighted outcome model estimator $\widehat{\tau}_{\text{OM-w}}$ is given by

$$\widehat{\tau}_{\text{OM-w}} = \frac{\sum_{i\neq j} d^{\text{OM-w}}\left(\boldsymbol{O}_i,\boldsymbol{O}_j;\widehat{\boldsymbol{\alpha}},I\left(g(\boldsymbol{X}_i^*;\boldsymbol{\beta})>g(\boldsymbol{X}_j^*;\boldsymbol{\beta})\right)\right)}{\sum_{i\neq j} d^{\text{OM-w}}(\boldsymbol{O}_i,\boldsymbol{O}_j;\widehat{\boldsymbol{\alpha}},1)}, \quad (26)$$

where

$$d^{\text{OM-w}}(\boldsymbol{O}_i,\boldsymbol{O}_j;\boldsymbol{\alpha},k(\boldsymbol{O}_i,\boldsymbol{O}_j)) = u_iu_jm(\boldsymbol{X}_i,\boldsymbol{X}_j;\boldsymbol{\alpha})I(S_i=0,S_j=0)k(\boldsymbol{O}_i,\boldsymbol{O}_j), \quad (27)$$

TABLE 2 Summary of the variables used in the transportability analysis of lung cancer prediction models. Continuous variables are summarized by their weighted mean (weighted standard deviation) and categorical variables are summarized by weighted percentage in each category. The "$g(\boldsymbol{X}^*;\boldsymbol{\beta})$" column indicates whether the variable is included as a predictor in the prediction model $g(\boldsymbol{X}^*;\boldsymbol{\beta})$ whose AUC we are evaluating in the target population. The "$\pi(\boldsymbol{X};\boldsymbol{\gamma})$" column indicates whether the elastic net procedure selected that variable for the model for the conditional expectation of the source population indicator. The "$q(\boldsymbol{X};\boldsymbol{\alpha})$" column indicates whether the elastic net procedure selected that variable for the outcome model.

Variable	NLST	NHANES	$g(\boldsymbol{X}^*;\boldsymbol{\beta})$	$\pi(\boldsymbol{X};\boldsymbol{\gamma})$	$q(\boldsymbol{X};\boldsymbol{\alpha})$
Age	61.3 (5.0)	62.7(5.4)	Yes	Yes	Yes
BMI	27.9 (5.0)	29.1 (6.5)	Yes	Yes	Yes
Race or ethnic group			Yes	Yes	Yes
Black	4.0%	7.7%			
White	90.6%	82.1%			
Hispanic	1.5%	6.2%			
Other	3.8%	4.0%			
Education level			Yes	Yes	Yes
less than high school	5.8%	35.5%			
high school graduate	38.5%	30.6%			
AA degree/some college	23.6%	21.9%			
college graduate	32.2%	12.0%			
Personal history of cancer	4.3%	22.5%	Yes	Yes	Yes
Smoking status			Yes	Yes	Yes
Current	47.5%	62.2%			
Former	52.5%	37.8%			
Smoking intensity[1]	28.4 (11.4)	27.3(12.1)	Yes	Yes	Yes
Duration of smoking (year)	39.6 (7.3)	42.1 (6.7)	Yes	Yes	Yes
Smoking quit time (year)[2]	7.3 (4.8)	6.5 (3.6)	Yes	Yes	Yes
Age started smoking	16.7 (3.7)	17.9 (3.8)	No	Yes	Yes
Gender			No	Yes	No
Male	58.7%	59.3%			
Female	41.3%	40.7%			
Marital status			No	Yes	Yes
Married or living as married	68.2%	61.6%			
Other	31.8%	38.4%			
Pack-years[3]	55.6 (23.5)	56.6 (25.8)	No	Yes	Yes
Asthma	6.1%	16.6%	No	Yes	Yes
Diabetes	9.2%	20.8%	No	Yes	Yes
Emphysema	7.3%	12.7%	No	Yes	Yes
Heart disease	12.3%	21.5%	No	Yes	Yes
Hypertension	34.8%	60.1%	No	Yes	Yes
Stroke	2.6%	6.1%	No	Yes	Yes

[1]Smoking intensity (the average number of cigarettes smoked per day) has a nonlinear association with lung cancer and was transformed by dividing by 10, exponentiating by the power −1 in models $g(\boldsymbol{X}^*;\boldsymbol{\beta})$, $\pi(\boldsymbol{X};\boldsymbol{\gamma})$ and $q(\boldsymbol{X};\boldsymbol{\alpha})$.

[2]Smoking quit time in former smokers (years).

[3]Pack-years are calculated by multiplying the number of packs of cigarettes smoked per day by the number of years the person has smoked divided by 20.

and u_i and u_j are the sampling weights for observations i and j, respectively. No adjustment for the sampling weights is needed when estimating $\boldsymbol{\alpha}$ because the parameter $\boldsymbol{\alpha}$ in the outcome model $q(\boldsymbol{X};\boldsymbol{\alpha})$ is estimated using only source population data.

Using the above notation, the weighted DR estimator is defined as

$$\widehat{\tau}_{\text{DR-w}} = \frac{\sum_{i\neq j} d^{\text{DR-w}}\left(\boldsymbol{O}_i, \boldsymbol{O}_j; \widetilde{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\alpha}}, k(\boldsymbol{O}_i, \boldsymbol{O}_j) = I(g(\boldsymbol{X}_i^*;\boldsymbol{\beta}) > g(\boldsymbol{X}_j^*;\boldsymbol{\beta}))\right)}{\sum_{i\neq j} d^{\text{DR-w}}\left(\boldsymbol{O}_i, \boldsymbol{O}_j; \widetilde{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\alpha}}, k(\boldsymbol{O}_i, \boldsymbol{O}_j) = 1\right)}, \tag{28}$$

TABLE 3 Estimated AUC in the target population using the source population data estimator $\widehat{\tau}_S$, the weighted inverse-odds weighting estimator $\widehat{\tau}_{\text{IOW-w}}$, the weighted outcome model estimator, $\widehat{\tau}_{\text{OM-w}}$, the weighted doubly robust estimator $\widehat{\tau}_{\text{DR-w}}$, and the associated standard error estimates

	$\widehat{\tau}_S$	$\widehat{\tau}_{\text{IOW-w}}$	$\widehat{\tau}_{\text{OM-w}}$	$\widehat{\tau}_{\text{DR-w}}$
AUC estimate	0.700	0.674	0.675	0.668
Standard error estimate	0.014	0.034	0.024	0.039

where

$$\begin{aligned} d^{\text{DR-w}}\left(\boldsymbol{O}_i, \boldsymbol{O}_j; \boldsymbol{\gamma}, \boldsymbol{\alpha}, k(\boldsymbol{O}_i, \boldsymbol{O}_j)\right) &= d^{\text{IOW-w}}(\boldsymbol{O}_i, \boldsymbol{O}_j; \boldsymbol{\gamma}, k(\boldsymbol{O}_i, \boldsymbol{O}_j)) \\ &+ d^{\text{OM-w}}(\boldsymbol{O}_i, \boldsymbol{O}_j; \boldsymbol{\alpha}, k(\boldsymbol{O}_i, \boldsymbol{O}_j)) \\ &- w(\boldsymbol{X}_i, \boldsymbol{X}_j; \boldsymbol{\gamma}) m(\boldsymbol{X}_i, \boldsymbol{X}_j; \boldsymbol{\alpha}) I \\ &(S_i = 1, S_j = 1) k(\boldsymbol{O}_i, \boldsymbol{O}_j). \end{aligned} \tag{29}$$

Table 3 shows the point estimates from the transportability estimators $\widehat{\tau}_{\text{IOW-w}}$, $\widehat{\tau}_{\text{OM-w}}$, $\widehat{\tau}_{\text{DR-w}}$, along with the AUC estimator that only uses source population data ($\widehat{\tau}_S$). Standard errors were from 1000 bootstraps that account for the sampling design used in NHANES (Rao & Wu, 1988; Shao, 2003) (see Supplementary Web Appendix E4 for additional details on the bootstrapping procedure). All four estimators produced similar results, suggesting that the classification performance of the logistic regression model is similar in the target and source populations.

In Supplementary Web appendix E4, we present results when data from the NHANES cycles 2001–2002, 2003–2004, 2005–2006, 2007–2008, or 2009–2010 were used as the sample from the target population. The results show that the performance is similar across NHANES cycles (25th quantile 0.720, median 0.725, and 75th quantile 0.734).

8 | DISCUSSION

We proposed methods for estimating the target population AUC for a prediction model developed using data from a source population that has a different covariate distribution compared to the target population. We considered the setting where covariate data are available from a sample of the source population and a separately obtained sample of the target population, but outcome data are only available from the sample of the source population. We provided conditions under which the AUC in the target population is identifiable and proposed three estimators for the AUC in the target population appropriate for use when no outcome information is available from the sample of the target population. We showed that the estimators are consistent and asymptotically normal and have good finite-sample performance. Last, we estimated the AUC of a lung cancer risk prediction model using source population data from the NLST and target population data from the NHANES. Because the NHANES uses a multistage clustering design with variable probability sampling and dependence between observations within each cluster, we modified the estimators to account for the complex survey design.

In our setting, estimating the model's AUC in the target population relies on two identifiability conditions: conditional independence of the outcome and data source and positivity of being in the source data. Conditional independence of the outcome and data source is an untestable assumption in our setting because outcome information is unavailable from the target population sample. Instead, the plausibility of the assumption should be evaluated using subject matter knowledge; future work could develop sensitivity analysis methods for exploring how violations of the assumption affect conclusions (Dahabreh et al., 2022; Robins et al., 2000).

The positivity condition is in principle testable, but formal assessments are challenging when the covariates needed to satisfy the conditional exchangeability assumption are high-dimensional (Petersen et al., 2012). When conducting transportability analyses, it is useful to examine the distribution of the estimated probabilities of being from the source population $\Pr[S = 1|\boldsymbol{X}]$ because estimated probabilities that are very close to zero may indicate near-violations of the positivity assumption (Dahabreh et al., 2020; Petersen et al., 2012). Even though the outcome model estimator does not involve an estimator of $\Pr[S = 1|\boldsymbol{X}]$, it is still useful to examine the distribution of the estimated probabilities because near-zero estimated values indicate that the outcome model estimator is "extrapolating" over parts of the covariate space that have no or very limited data in the source population (Dahabreh et al., 2020).

In practical applications of the methods we proposed, major challenges can arise in obtaining samples that are representative of the target populations of substantive interest (Barker et al., 2021). Nevertheless, we expect these challenges to be alleviated by increases in the amount and availability of routinely collected data (e.g., from electronic health records or medical claims) and by improved access to survey data that can be used to characterize target populations.

Another set of practical challenges involves tailoring the prediction model to the target population. If $X = X^*$, by condition A2, a consistent model in the source population is also consistent in the target population. And when condition A2 holds, the approach to tailoring a prediction model to the target population depends on whether the

model $g(\boldsymbol{X};\boldsymbol{\beta})$ is correctly specified or not. For correctly specified models, the maximum likelihood estimator calculated using only source population data is optimal (in the sense of minimizing asymptotic variance) (Shimodaira, 2000). For misspecified prediction models, however, the Kullback–Leibler divergence between the estimated and true conditional density of the outcome given covariates is minimized using a weighted maximum likelihood estimator with weights equal to $\Pr[S=0|\boldsymbol{X}]/\Pr[S=1|\boldsymbol{X}]$ (Shimodaira, 2000; Steingrimsson et al., 2021).

Ideally, the prediction model would be built and evaluated using a prospectively collected random sample from the target population. When that is infeasible, the methods we develop provide a useful alternative. Nevertheless, the potential availability of outcome data from the target population has interesting implications. First, the conditional independence of the outcome and data source becomes testable using tests for equivalence between conditional expectations (e.g., Luedtke et al., 2019). Nevertheless, conducting such tests with high-dimensional $\boldsymbol{X}$ can be challenging. Second, it may be possible to combine covariate and outcome data from both the target population and the source population both for model fitting and evaluation in the target population.

ACKNOWLEDGMENTS

This work was supported in part by grants U10CA180820 and U10CA180794 from the National Cancer Institute (NCI), National Library of Medicine (NLM) Award R01LM013616, awards ME-2019C3-17875, ME-2021C2-22365, and ME-1502-27794 from the Patient-Centered Outcomes Research Institute (PCORI), and Institutional Development Award U54GM115677 from the National Institute of General Medical Sciences of the National Institutes of Health (NIH), which funds Advance Clinical and Translational Research (Advance-CTR). We thank the NCI for access to the NLST data. The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of the NCI, PCORI, PCORI's Board of Governors, the PCORI Methodology Committee, the NIH, or the NLST.

DATA AVAILABILITY STATEMENT

The National Health and Nutrition Examination Survey data are publicly available and can be downloaded from https://cdc.gov/nchs/nhanes/default.aspx. The National Lung Screening Trial data can be requested from https://cdas.cancer.gov/learn/nlst/instructions/.

ORCID

Bing Li https://orcid.org/0000-0002-7360-6180
Issa J. Dahabreh https://orcid.org/0000-0002-2215-9931
Jon A. Steingrimsson https://orcid.org/0000-0003-2116-9377

REFERENCES

Agarwal, S., Graepel, T., Herbrich, R., Har-Peled, S. & Roth, D. (2005) Generalization bounds for the area under the roc curve. *Journal of Machine Learning Research*, 6, 393–425.

Bamber, D. (1975) The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of Mathematical Psychology*, 12(4), 387–415.

Barker, D.H., Dahabreh, I.J., Steingrimsson, J.A., Houck, C., Donenberg, G., DiClemente, R. et al. (2021) Causally interpretable meta-analysis: application in adolescent HIV prevention. *Prevention Science*, 23, 1–12.

Ben-David, S., Blitzer, J., Crammer, K., Kulesza, A., Pereira, F. & Vaughan, J.W. (2010) A theory of learning from different domains. *Machine Learning*, 79(1–2), 151–175.

Ben-David, S., Blitzer, J., Crammer, K. & Pereira, F. (2007) Analysis of representations for domain adaptation. In: *Advances in Neural Information Processing Systems* (pp. 137–144).

Chen, Y., Li, P. & Wu, C. (2020) Doubly robust inference with nonprobability survey samples. *Journal of the American Statistical Association*, 115(532), 2011–2021.

Cole, S.R. & Hernán, M.A. (2008) Constructing inverse probability weights for marginal structural models. *American Journal of Epidemiology*, 168(6), 656–664.

Cole, S.R. & Stuart, E.A. (2010) Generalizing evidence from randomized clinical trials to target populations: The ACTG 320 trial. *American Journal of Epidemiology*, 172(1), 107–115.

Dahabreh, I.J., Haneuse, S.J., Robins, J.M., Robertson, S.E., Buchanan, A.L., Stuart, E.A. et al. (2021) Study designs for extending causal inferences from a randomized trial to a target population. *American Journal of Epidemiology*, 190(8), 1632–1642.

Dahabreh, I.J. & Hernán, M.A. (2019) Extending inferences from a randomized trial to a target population. *European Journal of Epidemiology*, 34(8), 719–722.

Dahabreh, I.J., Robertson, S.E., Petito, L.C., Hernán, M.A. & Steingrimsson, J.A. (2019) Efficient and robust methods for causally interpretable meta-analysis: transporting inferences from multiple randomized trials to a target population. *arXiv preprint arXiv:1908.09230*.

Dahabreh, I.J., Robertson, S.E., Steingrimsson, J.A., Stuart, E.A. & Hernán, M.A. (2020) Extending inferences from a randomized trial to a new target population. *Statistics in Medicine*, 39(14), 1999–2014.

Dahabreh, I.J., Robertson, S.E., Tchetgen, E.J., Stuart, E.A. & Hernán, M.A. (2019) Generalizing causal inferences from individuals in randomized trials to all trial-eligible individuals. *Biometrics*, 75(2), 685–694.

Dahabreh, I.J., Robins, J.M., Haneuse, S.J., Robertson, S.E., Steingrimsson, J.A. & Hernán, M.A. (2022) Global sensitivity analysis for studies extending inferences from a randomized trial to a target population. *arXiv preprint arXiv:2207.09982*.

DeLong, E.R., DeLong, D.M. & Clarke-Pearson, D.L. (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, 44(3), 837–845.

Dyer, O. (2021) US task force recommends extending lung cancer screenings to over 50s. *BMJ: British Medical Journal (Online)*, 372.

Elliott, M.R. & Valliant, R. (2017) Inference for nonprobability samples. *Statistical Science*, 32(2), 249–264.

Hanley, J.A. & McNeil, B.J. (1982) The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, 143(1), 29–36.

Hoeffding, W. (1961) *The strong law of large numbers for u-statistics*. Technical report, Department of Statistics, North Carolina State University.

Krist, A.H., Davidson, K.W., Mangione, C.M., Barry, M.J., Cabana, M., Caughey, A.B. et al. (2021) Screening for lung cancer: US preventive services task force recommendation statement. *JAMA*, 325(10), 962–970.

Long, M., Cao, Y., Wang, J. & Jordan, M.I. (2015) Learning transferable features with deep adaptation networks. *arXiv preprint arXiv:1502.02791*.

Lu, Y., Scharfstein, D.O., Brooks, M.M., Quach, K. & Kennedy, E.H. (2019) Causal inference for comprehensive cohort studies. *arXiv preprint arXiv:1910.03531*.

Luedtke, A., Carone, M. & van der Laan, M.J. (2019) An omnibus non-parametric test of equality in distribution for unknown functions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 81(1), 75–99.

Mao, L. (2018) On causal estimation using u-statistics. *Biometrika*, 105(1), 215–220.

McNeil, B.J. & Hanley, J.A. (1984) Statistical approaches to the analysis of receiver operating characteristic (roc) curves. *Medical Decision Making*, 4(2), 137–150.

Moyer, V.A. (2014) Screening for lung cancer: US preventive services task force recommendation statement. *Annals of Internal Medicine*, 160(5), 330–338.

Murthy, V.H., Krumholz, H.M. & Gross, C.P. (2004) Participation in cancer clinical trials: race-, sex-, and age-based disparities. *Jama*, 291(22), 2720–2726.

Petersen, M.L., Porter, K.E., Gruber, S., Wang, Y. & Van Der Laan, M.J. (2012) Diagnosing and responding to violations in the positivity assumption. *Statistical Methods in Medical Research*, 21(1), 31–54.

Rao, J.N. & Wu, C. (1988) Resampling inference with complex survey data. *Journal of the American Statistical Association*, 83(401), 231–241.

Robins, J.M. (1988) Confidence intervals for causal parameters. *Statistics in Medicine*, 7(7), 773–785.

Robins, J.M., Rotnitzky, A. & Scharfstein, D.O. (2000) Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In: Halloran, M.E., & Berry, D.A. (Eds.) *Statistical models in epidemiology, the environment, and clinical trials* (pp. 1–94). New York: Springer.

Shao, J. (2003) Impact of the bootstrap on sample surveys. *Statistical Science*, 18(2), 191–198.

Shimodaira, H. (2000) Improving predictive inference under covariate shift by weighting the log-likelihood function. *Journal of Statistical Planning and Inference*, 90(2), 227–244.

Steingrimsson, J.A., Gatsonis, C. & Dahabreh, I.J. (2021) Transporting a prediction model for use in a new target population. *arXiv preprint arXiv:2101.11182*.

Sugiyama, M., Krauledat, M. & MÃžller, K.R. (2007) Covariate shift adaptation by importance weighted cross validation. *Journal of Machine Learning Research*, 8, 985–1005.

Sugiyama, M., Yamada, M. & du Plessis, M.C. (2013) Learning under nonstationarity: covariate shift and class-balance change. *Wiley Interdisciplinary Reviews: Computational Statistics*, 5(6), 465–477.

Tammemägi, M.C., Katki, H.A., Hocking, W.G., Church, T.R., Caporaso, N., Kvale, P.A. et al. (2013) Selection criteria for lung-cancer screening. *New England Journal of Medicine*, 368(8), 728–736.

Tan, Z. (2007) Comment: understanding OR, PS and DR. *Statistical Science*, 22(4), 560–568.

Team, N.L.S.T.R. (2011) The National Lung Screening Trial: overview and study design. *Radiology*, 258(1), 243–253.

Unger, J.M., Hershman, D.L., Albain, K.S., Moinpour, C.M., Petersen, J.A., Burg, K. et al. (2013) Patient income level and cancer clinical trial participation. *Journal of Clinical Oncology*, 31(5), 536.

Usunier, N., Amini, M.R. & Gallinari, P. (2005) A data-dependent generalisation error bound for the auc. In: *Proceedings of the ICML 2005 Workshop on ROC Analysis in Machine Learning*.

Westreich, D., Edwards, J.K., Lesko, C.R., Stuart, E. & Cole, S.R. (2017) Transportability of trial results using inverse odds of sampling weights. *American Journal of Epidemiology*, 186(8), 1010–1014.

Wieand, S., Gail, M.H., James, B.R. & James, K.L. (1989) A family of nonparametric statistics for comparing diagnostic markers with paired or unpaired data. *Biometrika*, 76(3), 585–592.

Zhou, X.H., McClish, D.K. & Obuchowski, N.A. (2009) *Statistical methods in diagnostic medicine*, volume 569. Hoboken, NJ: John Wiley & Sons.

SUPPORTING INFORMATION

Web Appendices and Tables referenced in Sections 2, 3, 5, 6, and 7 are available with this paper at the Biometrics website on Wiley Online Library. R code for the simulations along with implementation details is also provided.

Data S1

How to cite this article: Li, B., Gatsonis, C., Dahabreh, I.J. & Steingrimsson, J.A. (2023) Estimating the area under the ROC curve when transporting a prediction model to a target population. *Biometrics*, 79, 2382–2393. https://doi.org/10.1111/biom.13796