# STATISTICAL METHODS FOR RECEIVER OPERATING CHARACTERISTIC (ROC) CURVES IN COMPLEX SAMPLING SURVEYS

Tamy Harumy Moraes Tsujimoto

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics.

Chapel Hill
2022

Approved by:

Jianwen Cai

Donglin Zeng

Bonnie Shook-Sa

Feng-Chang Lin

Colin Orr

# ABSTRACT

Tamy Harumy Moraes Tsujimoto: Statistical Methods for Receiver Operating
Characteristic (ROC) Curves in Complex Sampling Surveys
(Under the direction of Dr. Jianwen Cai)

Sample surveys are critical in providing information in a broad range of areas, serving
as a valuable resource for guiding actions and policies. Classical methods in inferential
statistics assume that the observations were selected according to a simple random
sampling from a population of interest. However, in large-scale surveys, the final sample
usually does not represent a simple random sample of independent, identically distributed
observations from an infinite population. Instead, these studies use complex survey designs,
including stratification, multistage cluster sampling, and unequal selection probabilities
to obtain a representative sample more efficiently in terms of time and cost. Failure
to account for the complex survey design may result in biased parameter estimators,
underestimated standard errors, and possibly misleading conclusions.

The receiver operating characteristic (ROC) curve is the most popular tool used to
evaluate the accuracy of diagnostic tests measured on a continuous scale. Currently,
analyses based on the ROC curve have been performed on data arising from complex survey
samples ignoring the sampling scheme. For our first topic, we propose a nonparametric
estimator for the ROC curve that accounts for complex survey sampling and establish its
uniform convergence. The properties of the estimator are evaluated through simulation
studies and illustrated using the National Health and Nutrition Examination Survey
(NHANES).

Nonresponse is a common issue in surveys and can induce bias if not adequately
accounted for. For our second topic, we propose an IPW estimator for the ROC curve to

accommodate the case where the diagnostic test is missing. The theoretical properties of the estimator are developed and evaluated using simulation studies. The proposed estimator is then applied to the NHANES data.

In many applications, it is desired to study the covariate effects on the accuracy of a diagnostic test. In our third topic, we adapt a popular model referred to as ROC-GLM to account for complex survey designs. Simulation studies show that our design-adjusted ROC-GLM model performs well compared to the original model, which was developed for simple random sampling. To illustrate our method, we study the effect of age on the accuracy of a diabetes assessment calculator.

*To my past and future self, with love and compassion.*

# ACKNOWLEDGEMENTS

First, I would like to thank my advisor, Dr. Jianwen Cai, for her support and guidance during the dissertation process. Her constructive suggestions and kindness in moments of doubt were crucial for the success of this work. I am also grateful to my committee members, Dr. Feng-Chang Lin, Dr. Colin Orr, Dr. Bonnie Shook-Sa, and Dr. Donglin Zeng, for their valuable feedback and suggestions for this research.

I would not have pursued this Ph.D. if it were not for great mentors that empowered me. Thank you, Dr. Antonio Carlos Pedroso de Lima and Dr. Julio Singer, for encouraging me to improve as a statistician. Thank you also to Dr. Pranab Sen for being my first contact at UNC and being vital in securing the funding from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) (88881.128210/2016-01) that supported this work.

Thank you to my dear friends Andre Assumpcao, Ana Carolina Rios, and their daughter Stella Assumpcao Rios for being my family here in Chapel Hill, sharing special moments that will forever be in my heart. Thank you to my friends Hunyong Cho and Jinyoung Park for all the fantastic moments we spent together. And thank you to my dear Brazilian friends Damaris Regina and Tuany Castro for staying close, even from far away, and for constantly reminding me to be kind to myself.

Thank you to my parents, Elaine Tsujimoto and Luiz Tsujimoto, and my siblings, Thais Tsujimoto and Leonardo Tsujimoto, for all the love and for believing in me even when I did not believe in myself. To my love and friend, Victor Ritter, for supporting me in every way possible during this process. Thank you to my dog, Mochi, who patiently stayed by my side, filling me with love and joy. Finally, thank you, Tamy, for not giving up. You are more capable than you think, and this work is proof that you can do anything you dream of.

# TABLE OF CONTENTS

# LIST OF TABLES

x

# LIST OF FIGURES

xiii

# CHAPTER 1: INTRODUCTION

Sample surveys are critical in providing information in a broad range of areas, serving as a valuable resource for guiding actions and policies. In the United States, the National Center for Health Statistics (NCHS) is the principal health statistics agency under the Center for Disease Control (CDC). It conducts several population surveys, such as the National Health and Nutrition Examination Survey (NHANES), the National Health Interview Survey (NHIS), and the National Survey of Family Growth (NSFG). These large-scale surveys use complex sampling designs, including stratification, multistage cluster sampling, and unequal selection probabilities to obtain a representative sample more efficiently in terms of time and cost (Pfeffermann and Rao, 2009). Failure to account for the complex survey design may result in biased parameter estimators, underestimated standard errors, and possibly misleading conclusions (Heeringa et al., 2017).

A common use of large-scale surveys is to evaluate the accuracy of diagnostic tests. For example, Olson et al. (2010) and Higgins et al. (2011) compared the accuracy of A1C and oral glucose tolerance tests (OGTT) as diagnostic tests for diabetes and pre-diabetes using NHANES data. Laurson et al. (2011) evaluated the accuracy of percent body fat in identifying metabolic syndrome in adolescents, using data from NHANES.

The receiver operating characteristic (ROC) curve is the most popular tool used to evaluate the accuracy of diagnostic tests measured on a continuous scale. This work is motivated by the recent applications of ROC curves in large-scale surveys without completely accounting for the complex survey design. Some applications use outputs from design-adjusted regression models as inputs when constructing the (unweighted) ROC curves (Gaziano et al. (2008), Zhang et al. (2014)), while others use ROC curves to

1

evaluate the discriminatory performance of biomarkers ignoring the complex sampling design (Ferraro et al. (2010), DeBoer and Gurka (2014)).

Currently, there is little literature dealing with ROC curves in the context of complex survey data. Bisoffi et al. (2000) used bootstrap and jackknife methods to approximate standard error for the area under the ROC curve (AUC) for a two-phase sampling design. More recently, Yao et al. (2015) proposed a nonparametric estimator for the AUC that accounts for complex sampling, and employed jackknife method and balanced repeated replication for the variance estimation. To the best of our knowledge, the theoretical properties of a nonparametric estimator of the ROC curve accounting for complex survey sampling have not been fully developed, and will be the main focus in the first topic of this dissertation. We propose a nonparametric estimator for the ROC curve that accounts for complex survey sampling, and establish the uniform convergence using results presented in Han and Wellner (2021), combined with empirical processes arguments. Simulation studies showed that our proposed estimator is approximately unbiased and its estimated asymptotic variance is close to the empirical variance, with improved performance for larger sample size and disease proportions. The proposed method was applied to the National Health and Nutrition Examination Survey (NHANES) to evaluate the discriminatory ability of a traditional risk calculator for undiagnosed diabetes.

A common issue in large-scale surveys is the presence of missing biomarker values due to various reasons such as study drop-out or loss of information caused by uncontrollable factors. Due to the limited availability of statistical methods, analyses using ROC curves on complex survey data is currently done by ignoring the sampling scheme, and include only participants without missing data on the variables of interest. Most methods for handling nonresponse fall within inverse probability weighting (IPW) techniques and imputation strategies. IPW methods rely on a model for the response probability given a set of predictors (response model). In contrast, imputation approaches based on statistical prediction rules rely on models for the partially observed variables given the fully observed

variables (imputation models). A third class of methods combines IPW and imputation methods, called augmented IPW (AIPW) estimators, and has the property of being doubly robust to model misspecifications. For our second topic, we propose an IPW estimator for the ROC curve to accommodate the case where the diagnostic test is missing. The theoretical aspects of the estimator are developed following Han and Wellner (2021), combined with empirical process arguments. Simulation studies showed that our proposed estimator performs well, with a better performance than the estimators that do not account for survey design or missingness, especially for higher missing proportions. The proposed estimator was applied to the same data from Chapter 3 to evaluate the discriminatory ability of the risk calculator for undiagnosed diabetes accounting for the missing observations.

In many applications, it is desired to study the effects of covariates on the discriminating capacity of a diagnostic test. For example, disease severity may impact the marker accuracy, with less severe cases being more dificult to distinguish from controls. Currently, there are three major existing approaches to evaluate the covariate effects on the ROC curve (Pepe, 1998). In the first class of approaches, called induced methods, the distributions of the diagnostic test in the diseased and non-diseased populations are modeled separately, from which the ROC curve is computed. The second class considers regression models for the area under the ROC curve (AUC). Lastly, in the third approach, called direct methods, the covariate effects on the ROC curve are modeled directly. The latter class is often referred to as parametric distribution-free (PDF) models, since a parametric model is assumed for the ROC curve, but the distributions of the diagnostic tests remaing unspecified. For our third topic, we adapt the PDF model referred to as ROC-GLM model proposed by Pepe (2000a) and Alonzo and Pepe (2002) to account for complex survey designs. Simulation studies show that our design-adjusted ROC-GLM model performs well compared to the original model, which was developed for simple

random sampling. To illustrate our method, we study the effect of age on the accuracy of a diabetes assessment calculator.

This dissertation aims to fill the gaps by proposing methods for the ROC curve in the context of complex survey designs, and it is organized as follows. Chapter 2 provides the literature review comprising methods in complex survey data and the ROC curve. Chapter 3 presents a design-adjusted nonparametric estimator for the ROC curve and uses empirical process arguments to develop the estimator's asymptotic properties. Chapter 4 proposes an IPW estimator for the ROC curve to handle the case where a portion of the diagnostic test is missing in survey sampling. Finally, in Chapter 5 we adapt the ROC-GLM model proposed by Pepe (2000a) and Alonzo and Pepe (2002) to account for complex survey designs.

## CHAPTER 2: LITERATURE REVIEW

## 2.1 Complex Surveys

Sample surveys play a critical role in providing information in a broad range of areas, serving as a valuable resource for guiding actions and policies. In the United States, the National Center for Health Statistics (NCHS) is the principal health statistics agency under the Center for Disease Control (CDC). It conducts several population surveys, such as the National Health and Nutrition Examination Survey (NHANES), the National Health Interview Survey (NHIS), and the National Survey of Family Growth (NSFG).

Classical methods in inferential statistics assume that the observations were selected according to a simple random sampling from a population of interest. However, in large-scale surveys the final sample usually does not represent a simple random sample of independent, identically distributed observations from a population. Instead, these studies generally use complex survey designs, including stratification, multistage cluster sampling, and unequal selection probabilities to obtain the ultimate study sample. Failure to account for the complex survey design may result in biased parameter estimators, underestimated standard errors, and possibly misleading conclusions (Heeringa et al., 2017).

For further elaborations, let $U_N$ be a finite population of size $N$ with elements indexed by $i \in \{1, \cdots, N\}$. Each index $i$ is associated with a unique vector $(y_i, x_i, z_i) \in \mathbb{R}^p \times \mathbb{R}^k \times \mathbb{R}^q_+$ representing the characteristics of interest, the auxiliary information, and the sampling design information available at the time of the design of the survey on all units, respectively. For a sample $s$ drawn according to a sampling design $p(.)$, where $p(s) = P(\text{sample } s \text{ is selected})$, define the sampling indicator $\xi_i = I(i \in s)$, to

specify whether the $i$-th unit is in the sample or not, and the inclusion probability $\pi_i = \sum_{s:i \in s} p(s) = P(\xi_i = 1)$ of the unit $i$ being in the sample, computed from all possible samples that contain the unit $i$.

### 2.1.1 Modes of Inference

When analyzing complex survey data, one may formulate the question of interest around finite-population quantities, such as population totals, means, and proportions, or around the statistical model generating the finite population. We refer to the first as descriptive studies and the latter as analytic studies.

When targeting finite-population quantities, the design-based (or randomization-based) inference remains the dominant approach. In this framework, the $y_i$'s are considered fixed but unknown quantities, and the only source of randomness comes from the joint distribution of the random variables $\{\xi_1, \cdots, \xi_N\}$ (Pfeffermann (2000), Lohr (2019)). This framework also requires that the selection probabilities $\pi_i$'s are known for all units in the population.

For studies where the questions of interest are around the parameters of the statistical model generating the finite population, the model-based approach is often employed. In this context, it is assumed that $\{(y_i, x_i, z_i)\}_{i=1}^N$ are realizations of random variables $(Y, X, Z)$ defined on the same probability space. The distribution of $(Y, X, Z)$ is called a super-population model, and supplies the link between units in the sample and units not in the sample (Rubin-Bleuer et al. (2005), Lohr (2019)).

An alternative approach of inference for survey sampling combines design-based and model-based inference, and it is known as joint inference. Under this framework, the finite population is viewed as a realization of a statistical model (super-population model), and a sample is drawn from this finite population according to a sampling design. Inference under this approach requires explicitly accounting for two sources of randomness. The first source is the model-based randomness arising from the difference between the

finite-population quantity and the superpopulation parameter, and the second source is the design-based randomness resulting from the difference between the sample-based estimator and the finite-population quantity (Pfeffermann, 2000).

### 2.1.2 Nonresponse in survey sampling

A common issue in large-scale surveys is the presence of nonresponse. If the respondents and nonrespondents have similar behavior, the nonresponse can be ignored in the statistical analysis. However, ignoring the nonresponse may lead to biased results if the characteristics of respondents and nonrespondents are considerably different.

Data can be missing for the entire unit observation (unit nonresponse), or partially missing when at least one questionnaire item is missing (item nonresponse). To elaborate on methods to handle nonresponse, define the random variable $R_i$ that assumes 1 when the $i$-th unit responds, and 0 otherwise. After the sampling, a value for $y_i$ is observed if the realization $r_i$ of the random variable $R_i$ is equal to 1. We will also consider the partition $Y_i = (Y_i^o, Y_i^m)$ of the outcomes of interest, with $Y_i^o$ and $Y_i^m$ corresponding to the responses that are observed and missing, respectively.

#### 2.1.2.1 Missing Data Mechanisms

As described in Molenberghs et al. (2014) and Lohr (2019), the nonresponse is said to be missing completely at random (MCAR) if $R$ is independent of the outcome of interest $Y = (Y^o, Y^m)$, the auxiliary information $X$, and the survey design information $Z$. This is the missing mechanism implicitly adopted when nonresponse is ignored.

When $R$ is conditionally independent of $Y^m$ given $Y^o$, $X$, and $Z$, we say the data are missing at random (MAR). In this case, the nonresponse depends only on observed variables, and it can be validly inferred if a model is correctly specified for $R$. This is also known as an ignorable mechanism since the nonresponse can be ignored once its mechanism is completely explained by the model (Lohr, 2019). However, completely

ignoring this type of nonresponse without using a model for the nonresponse mechanism would lead to biased results.

Finally, if the conditional distribution of $R$ given $Y^o$, $X$, and $Z$ still depends on at least some components of $Y^m$, the data is said to be missing not at random (MNAR). This mechanism is often referred to as non-ignorable since the nonresponse cannot be fully explained by the observed data. Although standard methods to handle this type of nonresponse lead to biased results, they are still used to reduce the bias since the nonresponse may also depend on known variables (Lohr, 2019).

The main approaches for handling survey nonresponse are weighting and imputation. Typically, a combination of weighting and imputation is used, where imputation is employed for item nonresponse, and weighting is then applied to compensate for unit response (Särndal and Lundström, 2005). Approaches strictly using weighting or strictly using imputation are also possible.

### 2.1.2.2   Weighting Methods

In design-based inference, the sampling weights $w_i = \pi_i^{-1}$ are used to estimate population quantities for various sampling schemes. The weighting, in this case, can be viewed as a means of compensating the nonsampled units in the population in the sampling phase (Brick and Montaquila, 2009). In design-based theory, the inclusion probabilities $\pi_i$ are assumed to be known for all units in the population. When nonresponse is present, however, the unknown nature of the response probability model makes further assumptions necessary. Most of the methods assume that data are MAR, with the response probabilities estimated from information known for all units in the sample (Lohr, 2019).

One method of nonresponse weighting adjustment is to model the response propensities $\phi_i = P(R_i = 1)$ directly for the sampled units and use the inverse of the estimated propensities as the weight adjustment (Brick and Montaquila, 2009). This approach is referred to as inverse probability weighting (IPW).

An alternative method of nonresponse weighting adjustment is calibration. In general, calibration techniques use an auxiliary set of variables to adjust the sampling weights to improve estimation for the variables of interest. The idea is to scale the sampling weights so that the population totals for the auxiliary variables match the known population totals, resulting in a more representative sample of the population.

### 2.1.2.3   Imputation Methods

Imputation is the procedure used to replace missing values with substitutes based on a specific rule. According to Särndal and Lundström (2005), these imputed values can be constructed in various ways, and the imputation by statistical rules can be classified into two major categories. The first category is based on a statistical prediction rule, where a parametric model $m(X; \gamma)$ for $E(Y|X)$ (imputation model) is specified, and the parameter $\gamma$ is estimated from the complete cases. The second category, also described as the donor-based method, is based on values observed for responding units, but not for the nonresponding units (donor-based method). In this category, the imputed value for a given nonrespondent unit is borrowed from an observed unit (donor) that is considered very similar in a statistical sense.

Examples of donor-based methods are nearest neighbor imputation and hot-deck imputation. The former is a deterministic procedure, where the donor for a nonresponding unit is identified using distance minimization. The latter is a random procedure, where the imputed value for a nonresponding unit is randomly selected from a group of all potential (respondent) donor units.

## 2.2   Receiver Operating Characteristic (ROC) Curve

### 2.2.1   Diagnostic medicine

Diagnostic medicine is the process of identifying the disease or condition that a patient has, and ruling out conditions that the patient does not have, through assessment of the

patient's signs, symptoms, and results of various diagnostic tests (Zhou et al., 2009). It has evolved over the years as the advances in technology allowed the development of new diagnostic tests for detecting diseases. Examples of diagnostic tests include biochemical serum markers, such as prostate specific antigen (PSA) for prostate cancer, CA-125 for ovarian cancer, creatinine for kidney dysfunction, and cholesterol and blood pressure for cardiovascular disease (Pepe, 2000b). Given the importance of diagnostic medicine to population's overall health, and understanding of disease mechanism, statistical methods that assess the accuracy of diagnostic tests in a reliable way are crucial.

### 2.2.2 Definition of ROC curve

The receiver operating characteristic (ROC) curve is the most popular method to assess the performance of a continuous diagnostic test. The curve is defined as the plot of the false positive rate (1-specificity) versus the true-positive rate (sensitivity) across all possible cutoffs of the diagnostic test. The false-positive rate (FPR) is the proportion of non-diseased individuals that test positive for the disease based on the diagnostic test, and the true-positive rate (TPR) is the proportion of diseased individuals that test positive for the disease. The curve is especially useful to compare the performance of different diagnostic tests and obtain optimal cutoffs for the diagnostic test to minimize the misclassification of diseased and non-diseased individuals. (Pepe (2000b), Inácio et al. (2021)).

To introduce the ROC curve more formally, consider a binary indicator of disease status $D$, with $D = 1$ indicating diseased subjects, and a continuous diagnostic test $Y$, which larger values are more indicative of disease. We denote the cumulative distribution function (cdf) of $Y$ conditioned on $D = 0$ as $G$, and similarly, the cdf of $Y$ conditioned

on $D = 1$ as $F$. The corresponding FPR and TPR at the cutoff $c \in \mathbb{R}$ are given by:

$$\text{FPR}(c) = P(Y \geq c | D = 0) = P(Y_{\bar{D}} \geq c) = 1 - G(c)$$

$$\text{TPR}(c) = P(Y \geq c | D = 1) = P(Y_D \geq c) = 1 - F(c)$$

The ROC curve is defined as the plot $\{(\text{FPR}(c), \text{TPR}(c)) : c \in \mathbb{R}\}$, or equivalently, as the plot $\{(t, R(t)) : t \in [0, 1]\}$, where $R(t) = 1 - F \circ G^{-1}(1 - t)$, with $G^{-1}(s) = \inf\{x \in \mathbb{R} : G(x) \geq s\}$, and $F \circ G^{-1}(.) \equiv F(G^{-1}(.))$.

### 2.2.3 ROC curve estimation

Let $\{Y_{D_i}\}_{i=1}^{n_D}$ and $\{Y_{\bar{D}_i}\}_{i=1}^{n_{\bar{D}}}$ be two independent random samples of sizes $n_D$ and $n_{\bar{D}}$, respectively, of continuous diagnostic tests for the diseased and nondiseased populations.

#### 2.2.3.1 Non-parametric estimation

The ROC curve can be non-parametrically estimated by plugging in the corresponding empirical distributions for diseased and nondiseased, as described in Hsieh et al. (1996):

$$\hat{G}(c) = \frac{1}{n_{\bar{D}}} \sum_{i=1}^{n_{\bar{D}}} I(Y_{\bar{D}_i} \leq c) \qquad \hat{F}(c) = \frac{1}{n_D} \sum_{i=1}^{n_D} I(Y_{D_i} \leq c).$$

Hsieh et al. (1996) also provides a formal derivation including uniform consistency and weak convergence for the empirical ROC estimator.

Note that the empirical ROC estimator is an increasing discrete function, whereas the true ROC curve is a continuous function. To overcome the lack of smoothness of the empirical estimator, kernel-based methods for estimating ROC curve have also been developed. Zou et al. (1997) proposed using kernel density estimation to estimate the density function for each population, and Zou et al. (1998) proposed estimating the distribution functions using normal kernel.

### 2.2.3.2 Parametric estimation

Fully parametric estimation for the ROC curve is also possible. In this approach, the distribution functions are parametrically estimated for the diseased and nondiseased populations, and then used to compute the ROC curve. Typically, it is assumed that $Y^{\bar{D}} \sim \mathrm{N}(\mu_{\bar{D}}, \sigma_{\bar{D}}^2)$ and $Y^D \sim \mathrm{N}(\mu_D, \sigma_D^2)$, leading to the binormal model for the ROC curve:

$$R(t) = \Phi(a + b\Phi^{-1}(t)), \qquad (2.1)$$

where $a = (\mu_D - \mu_{\bar{D}})/\sigma_D$ and $b = \sigma_{\bar{D}}/\sigma_D$. In this case, the estimation is perfomed using maximum likelihood. In general, if $Y_{\bar{D}}$ and $Y_D$ have probability distribution functions with survival functions $S(y - \mu_D/\sigma_D)$ and $S(y - \mu_{\bar{D}}/\sigma_{\bar{D}})$ respectively, then the associated ROC curve has the form $S(-a + bS^{-1}(t))$ where $a = (\mu_D - \mu_{\bar{D}})/\sigma_D$ and $b = \sigma_{\bar{D}}/\sigma_D$ (Pepe et al., 2000).

### 2.2.3.3 Semiparametric estimation

Semiparametric approaches to ROC curve estimation have been proposed recently. In this framework, the strategies can be classified into two major categories. The first category aims to estimate semiparametrically the diagnostic test results, and obtain the induced estimate for the ROC curve. In the second category, a parametric form is assumed for the ROC curve, but no assumptions are made about the distribution from the diagnostic tests.

The first class of semiparametric approaches is the location-scale models, where a semiparametric location-scale model is proposed for the diagnostic test results:

$$Y_{\bar{D}_i} = \mu_{\bar{D}} + \sigma_{\bar{D}}\epsilon_i$$

$$Y_{D_i} = \mu_D + \sigma_D\epsilon_i,$$

where $\epsilon$ are independent random variables with mean zero and variance one with survival function $S$. The location-scale model for the ROC curve is given by

$$R(t) = S((\mu_{\bar{D}} - \mu_D)/\sigma_D + (\sigma_{\bar{D}}/\sigma_D)S^{-1}(t)).$$

The estimation for the location-scale model is done by computing sample means and variances, and estimating $S$ nonparametrically from the standardized residuals $(Y_{\bar{D}_i} - \hat{\mu}_{\bar{D}})/\hat{\sigma}_{\bar{D}}$ and $(Y_{D_i} - \hat{\mu}_D)/\hat{\sigma}_D$. Inferences for this estimator is based on resampling methods (Pepe, 2000b).

The second class of semiparametric approaches is referred to as parametric distribution-free methods. The binormal model remains the most popular approach, which assumes the existence of some unspecified strictly increasing transformation $H$, such that $H(Y_D)$ and $H(Y_{\bar{D}})$ follow a normal distribution, and then the ROC model is the binormal model written as in (2.1). The estimation is then carried for the parameters $a$ and $b$ in the binormal model. Metz et al. (1998) proposed the LABROC procedure, which uses ordinal regression methods applied to continuous data with a separate ordinal category defined for each non-diseased observation. Pepe (2000a) and Alonzo and Pepe (2002) suggested methods based on estimating equations. Zou and Hall (2000) proposed using maximum likelihood estimation based on the ranks of the diagnostic tests. These procedures share the attractive feature of providing a smooth estimate of the ROC curve. However, the statistical properties of these procedures have not yet been well characterized (Pepe et al., 2000).

### 2.2.4 Covariate-specific ROC curve

In many applications, the discriminatory capacity of a diagnostic test may be affected by a variety of factors. Currently, there are two approaches to evaluate the covariate effects on the ROC curve (Pepe, 1998). In the first approach, called the induced method, the distributions of the diagnostic test in diseased and non-diseased populations are

modeled separately, and the induced ROC curve is calculated. In the second approach, called the direct method, the covariate effects on the ROC curve are modeled directly.

### 2.2.4.1 Induced Methods

Similarly as described in sections 2.2.3.2 and 2.2.3.3, the induced methods assume that the distribution for the diagnostic test $Y$ follows a location-scale model, conditional on the disease status $D$ and covariates $X$:

$$Y = \mu(D, X) + \sigma(D, X)\epsilon,$$

where $\mu(D, X) = \mathrm{E}(Y|D, X)$, $\sigma^2(D, X) = \mathrm{Var}(Y|D, X)$, and $\epsilon$ are independent random variables with mean zero and variance one with survival function $S$. The corresponding covariate-specific ROC curve is given by

$$R(t) = S(-a(X) + b(X)S^{-1}(t)),$$

with $a(X) = (\mu(1, X) - \mu(0, X))/\sigma(1, X)$ and $b(X) = \sigma(0, X)/\sigma(1, X)$. The estimation can be performed using parametric, semiparametric, or nonparametric approaches. Faraggi (2003) proposed a fully parametric approach, assuming a linear model with homoscedastic normal distribution for each population. In a semiparametric context, Pepe (1998) proposed estimating the distribution function of the errors in each population by the corresponding empirical distribution function of the standardized residuals. In the nonparametric context, kernel-based approaches using continuous covariates were proposed by Yao et al. (2010), González-Manteiga et al. (2011), and Rodríguez-Álvarez et al. (2011).

### 2.2.4.2 Direct Methods

In contrast with induced methodologies, the direct methods model the covariate effects directly on the ROC curve. Pepe (1997) proposed a general regression modelling

framework for ROC curves given by:

$$R(t|X) = g(\alpha_0(t), \beta X),\qquad(2.2)$$

where $g$ is a specified function and $\alpha_0(t)$ is a parametric function. Both $g$ and $\alpha_0$ need to be chosen so that $R(t|X)$ is a monotone increasing function. Pepe (1997) proposed to estimate (2.2) by combining quantile regression and estimating equations. A special case of (2.2) is the class of ROC regression models of the generalized linear models (ROC-GLM) form

$$g\{R(t|X)\} = h_0(t) + \mu(X),\qquad(2.3)$$

where $\mu(X)$ models the covariate effects on the ROC curve, $h_0(t)$ is an unknown monotonic increasing function of the FPR related to the shape of the ROC curve, and $g$ is the link function. The estimation for the model (2.3) relies on the fact that

$$R(t) = 1 - F \circ G^{-1}(1 - t) = P(Y_D \geq G^{-1}(1 - t)) = \mathrm{E}\{I(Y_D \geq G^{-1}(1 - t))\},$$

and, therefore, the model (2.3) can be seen as a model for the binary variable $I(Y_D \geq G^{-1}(1 - t))$. Different model proposals were made by varying $g$, $h_0$ and $\mu$. Pepe (2000a) and Alonzo and Pepe (2002) assumed $g(.) = \Phi^{-1}(.)$, $\mu(X) = X'\beta$ and a parametric form for $h_0(t)$. Cai and Pepe (2002) and Cai (2004) proposed a more flexible model with $g(.) = \Phi^{-1}(.)$, $\mu(X) = X'\beta$, and $h_0(t)$ completely unspecified.

### 2.2.5 ROC curve in the presence of incomplete data

The methods described so far rely on the assumption of complete data. However, in diagnostic medicine, the presence of incomplete data is quite frequent. In the context of the ROC curve, there are two problem setups concerning incomplete data. First, the case

where the disease status $D$ is missing, and second, the case where the diagnostic test $Y$ is missing.

### 2.2.5.1  Missing disease status

It is common to have subjects not undergoing the definitive assessment of disease in practice due to the verification procedure being expensive or invasive. This results in missing information on the disease status and can lead to biased estimates for the ROC curve. This bias is referred to as verification bias.

As described in Fluss et al. (2009), the methods developed for correcting for verification bias are usually based on the assumptions that the diagnostic test has an ordinal scale (Begg and Greenes (1983), Gray et al. (1984), Toledano and Gatsonis (1996), Zhou (1996), Zhou (1998), Rodenberg and Zhou (2000)). Alonzo and Pepe (2005) proposed to correct verification bias for continuous diagnostic tests assuming that the missing disease status is MAR. Rotnitzky et al. (2006) and Fluss et al. (2009) developed doubly-robust estimators for the ROC curve under non-ignorable verification bias for continuous diagnostic tests. Their approaches rely on specifying a value for the nonignorable parameter (the log odds ratio of verification for diseased versus nondiseased individuals) to account for the identifiability problem. Liu and Zhou (2010) proposed to estimate the nonignorable parameter using likelihood approach.

### 2.2.5.2  Missing diagnostic test

In many applications, especially in observational studies, some subjects might have a missing diagnostic test due to various reasons. The presence of missing the diagnostic test may lead to biased estimates of the ROC curve. Many methods have been developed to construct the ROC curve in the presence of missing diagnostic tests recently. The first group of methods rely on inverse probability weighting (IPW) methods (Long et al. (2011b), Li and Ning (2015), Li et al. (2021)), while the second group of methods rely on

imputation methods (Long et al. (2011a), Liu and Zhao (2012), Qin and Wang (2012), Yang and Zhao (2015), Karakaya et al. (2015)).

## 2.3 ROC curve for complex survey sampling

Many applications include analyses based on the ROC curve performed using complex survey data. Some applications use outputs from regression models that account for complex survey design as outcomes when constructing the (unweighted) ROC curves (Gaziano et al. (2008), Zhang et al. (2014)), while others use ROC curves to evaluate the discriminatory performance of biomarkers without accounting for the complex design (Ferraro et al. (2010), DeBoer and Gurka (2014)). Bisoffi et al. (2000) used bootstrap and jackknife methods to approximate standard deviations for the area under the ROC curve (AUC) for a two-phase sampling design. Recently, Yao et al. (2015) proposed a nonparametric estimator for the AUC that accounts for complex sampling. However, the theoretical aspects of the ROC curve in the context of complex survey data still need to be further developed.

# CHAPTER 3: RECEIVER OPERATING CHARACTERISTIC CURVE FOR COMPLEX SURVEY DATA

## 3.1  Introduction

Diagnostic medicine is the process of identifying the disease or condition that a patient has, and ruling out conditions that the patient does not have, through assessment of the patient's signs, symptoms, and results of various diagnostic tests (Zhou et al., 2009). It has evolved over the years as the advances in technology allowed the development of new diagnostic tests for detecting diseases. Examples of diagnostic tests include biochemical serum markers, such as prostate specific antigen (PSA) for prostate cancer, CA-125 for ovarian cancer, creatinine for kidney dysfunction, and cholesterol and blood pressure for cardiovascular disease. Given the importance of diagnostic medicine to population's overall health, and understanding of disease mechanism, statistical methods that assess the accuracy of diagnostic tests in a reliable way are crucial.

The receiver operating characteristic (ROC) curve is the most popular method to assess the performance of a continuous diagnostic test. The curve is defined as the plot of the false positive rate (1-specificity) versus the true-positive rate (sensitivity) across all possible cutoffs of the diagnostic test. The false-positive rate (FPR) is the proportion of non-diseased individuals that test positive for the disease based on the diagnostic test, and the true-positive rate (TPR) is the proportion of diseased individuals that test positive for the disease. The curve is especially useful to compare the performance of different diagnostic tests and obtain optimal cutoffs for the diagnostic test to minimize the misclassification of diseased and non-diseased individuals. Although we focus on medical diagnosis, the ROC curve is widely used in many binary classification problems.

A comprehensive discussion on ROC curves can be found in Pepe (2000b) and Inácio et al. (2021), for example.

The ROC curve has been widely used in the analysis of data arising from complex surveys. Sample surveys play a critical role in providing essential information in a broad range of areas, serving as an essential resource to guide actions and policies. In the United States, the National Center for Health Statistics (NCHS) is the principal health statistics agency under the Center of Disease Control (CDC), and conducts several population surveys, such as the National Health and Nutrition Examination Survey (NHANES), the National Health Interview Survey (NHIS), and the National Survey of Family Growth (NSFG).

In large scale surveys, the final sample usually does not represent a simple random sample of independent, identically distributed observations from an infinite population. Instead, these studies generally use complex survey designs, including stratification, multistage cluster sampling and unequal selection probabilities to obtain a representative sample in the most effective manner from a finite population. Failure to account for complex survey design may result in biased and inconsistent parameter estimators, underestimated standard errors, and possibly misleading conclusions.

Due to the limited availability of statistical methods, analyses using ROC curves on complex survey data is currently done by ignoring the sampling scheme, even in papers that correctly account for the survey design in other aspects of the analysis. For example, Pandya et al. (2011) assessed the discrimination of traditional cardiovascular disease risk scores in the Third National Health and Nutrition Examination Survey (NHANES III) using unweighted ROC curves. Similarly, DeBoer and Gurka (2014) used an ROC curve to assess the ability of metabolic syndrome Z-score to discriminate impaired glucose tolerance in adolescents, without accounting for the survey design.

Currently, there is little previous literature dealing with ROC curves in the context of complex survey data. Bisoffi et al. (2000) used bootstrap and jackknife methods to

approximate standard error for the area under the ROC curve (AUC) for a two-phase sampling design. More recently, Yao et al. (2015) proposed a nonparametric estimator for the AUC that accounts for complex sampling, and employed jackknife method and balanced repeated replication for the variance estimation. To the best of our knowledge, the asymptotic properties of a nonparametric estimator of the ROC curve accounting for complex survey data have not been fully developed.

In this paper, we propose a nonparametric estimator for the ROC curve for complex survey data. The proposed estimator's asymptotic properties are developed and evaluated through simulation studies, and the method is applied to the National Health and Nutrition Examination Survey (NHANES).

## 3.2   Methods

### 3.2.1   Setup

Classical sampling theory concerns the inference for finite population quantities (parameters). In this context, the design-based (also called randomization-based) inference is often employed, where the characteristics of interest are considered fixed quantities associated with the finite population. The source of randomness is resulting from the sampling scheme, with random variables indicating whether the population unit is contained in the sample. When the questions of interest are based on parameters of a statistical model, the model-based (also called prediction-based) inference is often preferred. In this framework, the characteristics of interest are considered to be random variables generated from a statistical model.

In this paper, we handle the model-based and design-based inference jointly, using the super-population framework described in Rubin-Bleuer et al. (2005), and followed by Boistard et al. (2017) and Han and Wellner (2021). Under this approach, the finite population is viewed as a realization from a statistical model (superpopulation model), and a sample is drawn from this finite population according to the sampling design.

Inference under this approach requires to explicitly account for two sources of randomness: the model-based randomness, accounting for the difference between the finite population parameter and the superpopulation model parameter, and the design-based randomness, accounting for the difference between the sample estimator and the finite population parameter (Pfeffermann, 2000).

Consider a sequence of finite populations $\mathcal{U}^N$ of size $N = 1, 2, \cdots$, with corresponding set of indices $U_N = \{1, \cdots, N\}$. Each index $i \in U_N$ is associated with a unique vector $(y_i, z_i) \in \mathbb{R}^p \times \mathbb{R}_+^q$ representing, respectively, the characteristics of interest, and the sampling design information available at the time of the design of the survey on all units. We assume that $\{(y_i, z_i)\}_{i=1}^N$ are realizations of random variables $(Y, Z)$, $Y : \Omega \mapsto \mathbb{R}^p$, $Z : \Omega \mapsto \mathbb{R}_+^q$, defined on a common probability space $(\Omega, \mathfrak{F}, \mathbb{P}_m)$, and denote $\boldsymbol{y}^N = (y_1, \cdots, y_N)$, $\boldsymbol{Y}^N = (Y_1, \cdots, Y_N)$, $\boldsymbol{z}^N = (z_1, \cdots, z_N)$, and $\boldsymbol{Z}^N = (Z_1, \cdots, Z_N)$.

Let $\mathfrak{S}_N = \{s : s \subset U_N\}$ be the collection of subsets of $U_N$ selected under a given sampling scheme and let $\sigma(\mathfrak{S}_N)$ be the $\sigma$-algebra generated by $\mathfrak{S}_N$. A sampling design associated with a sampling scheme is a function $p : \sigma(\mathfrak{S}_N) \times \mathbb{R}_+^{q \times N} \mapsto [0, 1]$ such that

(i) for all $s$ in $\mathfrak{S}_N$, $\boldsymbol{z}^N \mapsto p(s, \boldsymbol{z}^N)$ is Borel-measurable on $\mathbb{R}_+^{q \times N}$;

(ii) for $\boldsymbol{z}^N \in \mathbb{R}_+^{q \times N}$, $A \mapsto p(A, \boldsymbol{z}^N)$ is a probability measure on $\sigma(\mathfrak{S}_N)$.

Note that since $p$ does not depend on $\boldsymbol{y}^N$, only non-informative sampling designs are considered. Similarly to Boistard et al. (2017), for each $\omega \in \Omega$ we define a probability measure $A \mapsto \mathbb{P}_d(A, \omega) = \sum_{s \in A} p(s, \boldsymbol{Z}^N(\omega))$, and we say that $(\mathfrak{S}_N, \sigma(\mathfrak{S}_N), \mathbb{P}_d)$ is the design probability space. We will work on a product probability space $(\mathfrak{S}_N \times \Omega, \sigma(\mathfrak{S}_N) \times \mathfrak{F}, \mathbb{P}_{d,m})$ that includes both the super-population and the design space with probability measure $\mathbb{P}_{d,m}$ defined as $\mathbb{P}_{d,m}(s \times E) = \int_E \mathbb{P}_d(s, \omega) \, \mathrm{d}P_m(\omega)$, with $(s, E) \in \sigma(\mathfrak{S}_N) \times \mathfrak{F}$. We adopt $\mathbb{E}_d$, $\mathbb{E}_m$ and $\mathbb{E}_{d,m}$ to denote the expectation with respect to the probability space $(\mathfrak{S}_N, \sigma(\mathfrak{S}_N), \mathbb{P}_d)$, $(\Omega, \mathfrak{F}, \mathbb{P}_m)$ and $(\mathfrak{S}_N \times \Omega, \sigma(\mathfrak{S}_N) \times \mathfrak{F}, \mathbb{P}_{d,m})$, respectively. For a sample $s$ drawn according to a sampling design $p$, the sampling indicators $\xi_i = I(i \in s)$ are random variables defined on $(\mathfrak{S}_N \times \Omega, \sigma(\mathfrak{S}_N) \times \mathfrak{F}, \mathbb{P}_{d,m})$, with first-order inclusion

probabilities defined as $\pi_i(\omega) = \mathbb{E}_{d,m}[\xi_i | \boldsymbol{Z}^N(\omega)]$, and second-order inclusion probabilities defined as $\pi_{ij}(\omega) = \mathbb{E}_{d,m}[\xi_i \xi_j | \boldsymbol{Z}^N(\omega)]$.

### 3.2.2 Assumptions and Notations

Following Han and Wellner (2021), we impose the following assumptions for the asymptotic convergence:

(A1.1) $\min_{1 \leq i \leq N} \pi_i \geq \pi_0 > 0$

(A1.2) $\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left( \frac{\xi_i}{\pi_i} - 1 \right) = O_{\mathbb{P}_{d,m}}(1)$

(A1.3) There exist constant $K > 0$ such that

$$\sup_{N \in \mathbb{N}} \sup_{1 \leq i \neq j \leq N} N|\pi_{ij} - \pi_i \pi_j| \leq K$$

(A1.4) The sample size $n$ increases as the population size $N$ increases, with

$$\lim_{N \to \infty} \frac{n}{N} = \lambda, \quad 0 < \lambda < 1$$

Let $\{V_i\}$ be a sequence of bounded i.i.d random variables defined on $(\Omega, \mathfrak{F}, \mathbb{P}_m)$. Let $S_N^2$ be the design-based variance of the Horvitz-Thompson estimator of the population mean, that is,

$$S_N^2 = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} V_i V_j$$

(A2.1) Suppose that for $N$ sufficiently large

$$\frac{1}{S_N} \left( \frac{1}{N} \sum_{i=1}^{N} \frac{\xi_i}{\pi_i} V_i - \frac{1}{N} \sum_{i=1}^{N} V_i \right) \to N(0, 1), \quad \omega\text{-a.s}$$

in distribution under $\mathbb{P}_d$

(A2.2) There exist constants $\mu_{\pi 1}, \mu_{\pi 2} \in \mathbb{R}$ such that

$$\frac{1}{N} \sum_{i=1}^{N} \frac{\pi_{ii} - \pi_i^2}{\pi_i^2} \to \mu_{\pi 1} \quad \text{in } \mathbb{P}_m$$

$$\frac{1}{N} \sum_{i \neq j}^{N} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} \to \mu_{\pi 2} \quad \text{in } \mathbb{P}_m$$

Following Han and Wellner (2021), for $\{\pi_i\}_{i=1}^{N}$, $\{\xi_i\}_{i=1}^{N}$, $\{Y_i\}_{i=1}^{N}$, and a class $\mathcal{F}$ of real functions $f$ we define the Horvitz-Thompson empirical measure as

$$\mathbb{P}_N^{\pi}(f) = \frac{1}{N} \sum_{i=1}^{N} \frac{\xi_i}{\pi_i} f(Y_i), \quad f \in \mathcal{F}, \tag{3.4}$$

and the associated Horvitz-Thompson empirical process as

$$\mathbb{G}_N^{\pi}(f) = \sqrt{N}(\mathbb{P}_N^{\pi} - P)(f), \quad f \in \mathcal{F}. \tag{3.5}$$

The usual empirical measure and empirical process ($\xi_i/\pi_i = 1$ for all $i = 1, \cdots, N$) will be denoted as $\mathbb{P}_N$ and $\mathbb{G}_N$.

### 3.2.3 ROC curve for complex survey sampling

Let $\{(Y_i, Z_i) = (X_i, D_i, Z_i) \in \mathbb{R} \times \{0,1\} \times \mathbb{R}_+^q\}_{i=1}^{N}$ be i.i.d realizations of the diagnostic test measure $X$, the disease indicator $D$, and the sampling design information $Z$. We denote the cumulative distribution function (cdf) of $X$ conditioned on $D = 0$ as $G$, and similarly, the cdf of $X$ conditioned on $D = 1$ as $F$. We assume that $F$ and $G$ have continuous probability density functions (pdf) $f$ and $g$, respectively. The ROC curve is defined as the plot of $\{(1 - G(c), 1 - F(c)) : c \in \mathbb{R}\}$, or equivalently, as the plot of $\{(s, R(s)) : s \in [0,1]\}$, where $R(s) = 1 - F \circ G^{-1}(1 - s)$, with $G^{-1}(s) = \inf\{x \in \mathbb{R} : G(x) \geq s\}$, and $F \circ G^{-1}(.) \equiv F(G^{-1}(.))$. The area under the ROC curve (AUC-ROC) is $A = \int_0^1 R(s) \, ds$. The corresponding finite-population quantities are $R_N(s) =$

$1 - F_N \circ G_N^{-1}(1-s)$ and $A_N = \int_0^1 R_N(s)ds$, where $G_N(x) = N_0^{-1} \sum_{i=1}^N I(X_i \leq x, D_i = 0)$, $F_N(x) = N_1^{-1} \sum_{i=1}^N I(X_i \leq x, D_i = 1)$, and $N_d = \sum_{i=1}^N I(D_i = d)$, $d = 0, 1$.

Consider a sample $\mathfrak{s}$, consisting of $0 \leq n \leq N$ units drawn from the finite population using a sampling design $p$. A survey-weighted estimator for the ROC curve can be obtained by substituting $F$ and $G$ by their Hájek type estimators:

$$R_n(s) = 1 - F_n \circ G_n^{-1}(1-s), \tag{3.6}$$

where

$$G_n(x) = \frac{1}{\hat{N}_0} \sum_{i=1}^N \frac{\xi_i}{\pi_i} I(X_i \leq x, D_i = 0) \quad \text{and} \quad F_n(x) = \frac{1}{\hat{N}_1} \sum_{i=1}^N \frac{\xi_i}{\pi_i} I(X_i \leq x, D_i = 1),$$
$$\tag{3.7}$$

with $\hat{N}_d = \sum_{i=1}^N \xi_i \pi_i^{-1} I(D_i = d)$, $d = 0, 1$.

The correspondent estimator for the area $A$ under $R(s)$ is

$$A_n = \int_0^1 R_n(s) \; ds. \tag{3.8}$$

The estimators $F_n(.)$ and $G_n(.)$ can be seen as ratios of Horvitz-Thompson empirical measures (4.11) with respect to the class of function $\mathcal{F} = \{f_{s,l}(y) \equiv f_{s,l}(x,d) = I(x \leq s, d = l) : s \in \mathbb{R}, l \in \{0,1\}\}$. Note that this class of functions is P-Donsker (Kosorok, 2008). The finite-dimensional convergence of $\mathbb{G}_N^\pi(f_{s,l})$ can be shown similarly as done in Boistard et al. (2017) using Crámer-Wold device. By Corollary 3.13 from Han and Wellner (2021), $\sqrt{n}(\mathbb{P}_N^\pi - \mathbb{P}_N) \leadsto \mathbb{G}^\pi$ in $\ell^\infty(\mathcal{F})$, where $\mathbb{G}^\pi$ is a tight Gaussian process with covariance function

$$\text{Cov}(\mathbb{G}^\pi(f_{s,d}), \mathbb{G}^\pi(f_{u,d'})) = \lambda\left(\mu_{\pi_1} P(f_{s,d} f_{u,d'}) + \mu_{\pi_2}(P f_{s,d})(P f_{u,d'})\right) \quad f_{s,d}, f_{u,d'} \in \mathcal{F},$$

with $P(f_{s,l}) = \int_{\mathcal{Y}} f_{s,l}(y) P(dy) = P(X \leq s, D = l)$ and $P(f_{s,d} f_{u,d'}) = \int_{\mathcal{Y}} f_{s,d}(y) f_{u,d'}(y) P(dy) = P(X \leq s \wedge u, D = d)$ for $d' = d$, and zero if $d' \neq d$.

The proposed estimator $R_n$ for the ROC curve depends on the pair $(G_n, F_n)$ through the map $\psi(A, B) = B(A^{-1})$, where $A^{-1}$ is the inverse map of $A$. Combining the results from Han and Wellner (2021) and Functional Delta Method (Vaart and Wellner, 1996) arguments presented in the Appendix, the following result will follow:

**Theorem 3.2.1** (FINITE POPULATION INFERENCE). *Consider the estimators $F_n$, $G_n$, $R_n$ and $A_n$ as defined in (3.6), (3.7) and (3.8). Suppose that conditions (A1.1)-(A2.2) hold.*

(a) *(Survey-weighted empirical distributions).*

$$\sqrt{N} \begin{bmatrix} G_n - G_N \\ F_n - F_N \end{bmatrix} \rightsquigarrow \begin{bmatrix} \mathbb{G}_0^\pi \\ \mathbb{G}_1^\pi \end{bmatrix} = \begin{bmatrix} \{(1-p)^{-1} \mu_{\pi_1}\}^{1/2} B_1(G) \\ \{p^{-1} \mu_{\pi_1}\}^{1/2} B_2(F) \end{bmatrix},$$

*where $B_1(.)$ and $B_2(.)$ denote two independent Brownian bridges and $p = P(D = 1)$.*

(b) *(Survey-weighted ROC curve). Suppose that $F$ and $G$ have continuous positive densities $f$ and $g$, respectively, on $[G^{-1}(a) - \epsilon, G^{-1}(b) + \epsilon]$ and that $f(G^{-1})/g(G^{-1})$ is bounded on any subinterval $(a, b)$, $0 < a < b < 1$. Then, for $0 < s < 1$*

$$\sqrt{n} \left( F_n \circ G_n^{-1}(s) - F_N \circ G_N^{-1}(s) \right) \rightsquigarrow \sqrt{\lambda \mu_{\pi_1}} \left\{ p^{-1/2} B_2(F \circ G^{-1}(s)) + \right.$$
$$\left. + (1-p)^{-1/2} \frac{f(G^{-1}(s))}{g(G^{-1}(s))} B_1(s) \right\}$$

*where $B_1(.)$ and $B_2(.)$ denote two independent Brownian bridges. This result implies that $\sqrt{n}(R_n(s) - R_N(s)) \rightsquigarrow \mathbb{W}(G^{-1}(1 - s))$, where $\mathbb{W}(u)$ is a Gaussian process with*

25

*mean zero and covariance function $\mathbb{E}_{d,m}\{\mathbb{W}(u)\mathbb{W}(t)\} = \sigma^2(u,t)$ given by*

$$\sigma^2(u,t) = \lambda\mu_{\pi_1}\left\{p^{-1}(F(u\wedge t) - F(u)F(t)) + \right.$$
$$\left. + (1-p)^{-1}\frac{f(u)f(t)}{g(u)g(t)}(G(u\wedge t) - G(u)G(t))\right\} \qquad (3.9)$$

*(c) (Survey-weighted AUC)*

$$\sqrt{n}\left(A_n - A_N\right) \to N(0, \delta^2)$$

*in distribution, where*

$$\delta^2 = \int_0^1\int_0^1 \sigma^2\{G^{-1}(1-s), G^{-1}(1-t)\}\ ds\ dt = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \sigma^2(s,t)\ dG(s)\ dG(t)$$

The proof for Theorem 3.2.1 are presented in Appendix A of the Supplementary Material.

The results for the super-population inference can be obtained from the decomposition

$$\sqrt{n}\left(F_n\circ G_n^{-1} - F\circ G^{-1}\right) = \sqrt{n}\left(F_n\circ G_n^{-1} - F_N\circ G_N^{-1}\right) + \sqrt{n}\left(F_N\circ G_N^{-1} - F\circ G^{-1}\right).$$

From the results presented in Theorem 3.2.1, we have that the first component converges to a zero mean Gaussian process under $\mathbb{P}_{d,m}$ (and $\mathbb{P}_d$). Using similar arguments, combined with classical empirical processes results, we have that the second component also converges to a zero mean Gaussian process under $\mathbb{P}_m$. Theorem 5.1(iii) from Rubin-Bleuer et al. (2005) imply that the two components are asymptotically independent, leading in the following result:

**Theorem 3.2.2** (SUPER POPULATION INFERENCE). *Consider the estimators $F_n$, $G_n$, $R_n$ and $A_n$ as defined in (3.6), (3.7) and (3.8). Suppose that conditions (A1.1)-(A2.2) hold.*

(a) *(Survey-weighted ROC curve). Suppose that $F$ and $G$ have continuous positive densities $f$ and $g$, respectively, on $[G^{-1}(a) - \epsilon, G^{-1}(b) + \epsilon]$ and that $f(G^{-1})/g(G^{-1})$ is bounded on any subinterval $(a, b)$, $0 < a < b < 1$. Then, for $0 < s < 1$*

$$\sqrt{n}\left(F_n \circ G_n^{-1}(s) - F \circ G^{-1}(s)\right) \rightsquigarrow \sqrt{\lambda(1 + \mu_{\pi_1})}\left\{p^{-1/2}B_2(F \circ G^{-1}(s)) + \right.$$

$$\left. + (1-p)^{-1/2}\frac{f(G^{-1}(s))}{g(G^{-1}(s))}B_1(s)\right\}$$

*where $B_1(.)$ and $B_2(.)$ denote two independent Brownian bridges. This result implies that $\sqrt{n}(R_n(s) - R(s)) \rightsquigarrow \mathbb{W}(G^{-1}(1-s))$, where $\mathbb{W}(u)$ is a Gaussian process with mean zero and covariance function $\mathbb{E}_{d,m}\{\mathbb{W}(u)\mathbb{W}(t)\} = \sigma^2(u, t)$ given by*

$$\sigma^2(u, t) = \lambda(1 + \mu_{\pi_1})\left\{p^{-1}(F(u \wedge t) - F(u)F(t)) + \right.$$

$$\left. + (1-p)^{-1}\frac{f(u)f(t)}{g(u)g(t)}(G(u \wedge t) - G(u)G(t))\right\}$$

(b) *(Survey-weighted AUC).*

$$\sqrt{n}\left(A_n - A\right) \to N(0, \delta^2)$$

*in distribution, where*

$$\delta^2 = \int_0^1 \int_0^1 \sigma^2\{G^{-1}(1-s), G^{-1}(1-t)\} \, ds \, dt = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sigma^2(s, t) \, dG(s) \, dG(t)$$

*Theorem 3.2.2 implies that $\sqrt{n}(R_n(s) - R(s))$ converges in distribution to $N(0, \sigma^2(s))$, with $\sigma^2(s)$ given by*

$$\sigma^2(s) = \lambda(1 + \mu_{\pi_1})\left\{p^{-1}R(s)(1 - R(s)) + (1-p)^{-1}\frac{f(G^{-1}(1-s))^2}{g(G^{-1}(1-s))^2}s(1-s)\right\} \quad (3.10)$$

Let $\hat{\sigma}^2(s)$ be the survey-weighted empirical version of $\sigma^2(s)$ with $(R, p, F, G, f, g)$ replaced by their survey-weighted estimates. An approximate level $1 - \alpha$ pointwise confidence interval for $R(s)$ is given by $R_n(s) \pm z_{1-\alpha/2}[\hat{\sigma}(s)^2/n]^{1/2}$, where $z_\alpha$ is such that $P(Z \leq z_\alpha) = \alpha$ with $Z \sim N(0, 1)$.

## 3.3 Simulation studies

In the simulation studies, we investigate the performance of the proposed estimator for the ROC curve under stratified simple random sampling (SSRS) and stratified two-stage cluster sampling (STSCS). For each sampling scheme, a total of 8 scenarios were considered according to different finite population sizes $N = 50,000$ and $100,000$, disease proportions $p = 5\%, 25\%$ and sampling fractions $\lambda = 5\%, 10\%$.

We generated populations subdivided in five strata containing 5%, 10%, 25%, 30% and 30% of the observations. We set the AUC for the strata to 0.95, 0.9, 0.8, 0.7, 0.6 respectively, and for each stratum $h = 1, \cdots, 5$, we generated $X_h = \alpha_h D + \epsilon$, where $D \sim \text{Ber}(p)$ and $\epsilon \sim \text{N}(0, 1)$. The ROC curve in each stratum is given by the binormal model $R(s) = \Phi(\alpha_h + \Phi^{-1}(s))$, where $\Phi(.)$ is the standard normal cdf and $\alpha_h$ is determined from the corresponding AUC specified for the $h$-th stratum. For STSCS, $M = 5,000$ and $10,000$ clusters of sizes 5, 10 and 15 were generated using quantiles of $X_h + \tau, \tau \sim N(0, 1)$, in addition to the steps already described.

Samples with size determined by the sampling fraction were drawn assuming uniform allocation. For the first stage of STSCS, $m = 310$ and 625 clusters were selected from the population of size $N = 50,000$, and $m = 625$ and 1250 clusters were sampled from the population of size $N = 100,000$. At the second stage, 80% of the observations were sampled from each cluster.

We evaluated the performance at $s \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ in terms of Relative Bias (RB), Empirical Standard Error (ESE), Asymptotic Standard Error (ASE), and Coverage Probability (CP) for 95% confidence intervals. We compare our method

(SVY) to the unweighted ROC curve (UN), where the sampling weights are ignored, and the asymptotic variance is computed following Hsieh et al. (1996). We also include the case (WT) where the sampling weights are used to compute weighted estimates for the ROC curve and the asymptotic variance from Hsieh et al. (1996). Results are obtained by generating 2,000 finite populations and selecting one sample from each of the finite populations.

The results for the relative biases under SSRS and STSCS are reported in Figure 3.5.1 and Tables S1 and S2 in the Supplementary Materials. As expected, the relative bias for the UN estimator is quite large, especially at the beginning of the ROC curve, with relative biases close to 30%. In contrast, the values for the SVY and WT estimators never exceed 0.5%.

The estimates for the empirical and asymptotic standard errors under SSRS and STSCS are presented in Figure 3.5.2 and Tables S3 and S4 Supplementary Materials. For the SWY estimator, the values were obtained by plugging survey-weighted estimates of $p$, $F$, $G$, $f$, and $g$ into the expression (3.10). For the UN and WT estimators, unweighted and survey-weighted estimates of $p$, $F$, $G$, $f$, and $g$ were plugged into the asymptotic variance expression presented in Hsieh et al. (1996). In general, our method estimates are close to the ESE, with better performance for larger sample sizes and disease proportions. The variance estimator that ignores the complex-survey design leads to underestimated standard errors, even when the sampling weights are used.

Figure 3.5.3 and Tables S5 and S6 in the Supplementary Materials give the coverage probabilities of the 95% confidence interval for the ROC curve. In general, the coverage probabilities based on our method are closer to 95% at the beginning of the ROC curve and decrease as we increase the FPR, except for FPR $= 0.9$ in the case of the smallest finite population, sample size, and disease proportion. The WT estimator presents coverage probabilities close to 92% at most, and the UN estimator performs poorly due to the significant bias and underestimated variances.

## 3.4 Application

Diabetes and its complications are major causes of morbidity and mortality worldwide. Currently, clinical practice guidelines recommend screening for pre-diabetes and type 2 diabetes with an informal assessment of risk factors or validated risk calculator in asymptomatic adults to guide providers on whether performing a definitive diagnostic test is necessary (Draznin et al., 2022). The current risk assessment tool used by the American Diabetes Association (ADA) to screen for pre-diabetes and type 2 diabetes is adapted from the algorithm developed in Bang et al. (2009) to estimate the risk of undiagnosed diabetes.

In this application, we wish to evaluate the discrimination of the algorithm developed by Bang et al. (2009) using the National Health and Nutrition Examination Survey (NHANES) between 1999-2006. NHANES is an annual survey conducted by the Centers for Disease Control and Prevention's (CDC) National Center for Health Statistics (NCHS) that utilizes a complex, multistage probability sampling design to select a representative sample of the non-institutionalized resident population of the United States.

Similarly as presented in Bang et al. (2009), we consider participants aged 20 years or more, excluding pregnant women, that had fasting plasma glucose (FPG) results. The participants are classified into four groups of diabetes status: known diabetes (if answered "yes" to the question "Other than during pregnancy, have you ever been told by a doctor or health professional that you have diabetes or sugar diabetes?"), normal glucose metabolism (FPG < 100 mg/dL), pre-diabetes (FPG 100-125 mg/dL), and undiagnosed diabetes (FPG > 125 mg/dL). The participants classified as "known diabetes" are not included in the analysis, and the undiagnosed diabetes was used as the binary outcome. The risk score was computed using age (< 40, 40-49, 50-59, > 59), sex (female, male), family history of diabetes (yes, no), history of hypertension (yes, no), obesity (not overweight, overweight, obese, extremely obese), physically active (yes, no).

In the 1999-2006 NHANES, 20,159 non-pregnant adults aged 20 years or more were enrolled. Out of this sample, 17,696 observations were classified as either normal glucose metabolism, pre-diabetes, and undiagnosed diabetes, and 7,348 observations had information for all variables needed to compute the risk score. In this final analytic sample, the proportion of undiagnosed diabetes is 3.1% (95% CI: 2.6, 3.5).

Figure 3.5.4 shows both survey-weighted and unweighted estimates of the ROC curve, as well as its corresponding AUC. The most considerable discrepancies between unweighted and survey-weighted estimates are observed between FPR 0.1-0.5, with the unweighted ROC curve being lower than the survey-weighted ROC curve. As a result, the AUC (survey-weighted = 0.83, unweighted = 0.80) is smaller when the survey weights are not considered.

## 3.5    Discussion

In this paper, we studied a nonparametric estimator for the ROC curve in the context of complex survey data. We examined the asymptotic properties of the proposed estimator and evaluated its performance in finite samples through simulation studies. The asymptotic properties of the proposed estimator were developed using empirical process arguments in the super-population framework described in Rubin-Bleuer et al. (2005), where the sources of randomness from both model-based and design-based inference are jointly taken into account.

The uniform convergence for the ROC curve in the finite population and super-population levels were established using key results presented in Han and Wellner (2021), combined with empirical processes arguments. The asymptotic distribution for the finite population and the super-population level AUC was also presented. Simulation studies showed that our proposed estimator performed well in the practical situations considered. The estimator was then applied to a national-level health survey to evaluate the discriminatory ability of a traditional risk calculator of undiagnosed diabetes.

The methods presented in this paper serve as a basis for nonparametric estimation of the ROC curve in the context of complex survey data. The weakly convergence results make it possible to further compute confidence bands for the ROC curve in both super-population and finite population levels. The proposed estimator may serve as an option when using data arising from complex survey data, preventing from biased results and possibly misleading conclusions by ignoring the sampling design.

The proposed estimator is a discrete function, whereas the true ROC curve for continuous data is a continuous function. To have a smooth estimate, the study of semiparametric and parametric models for ROC curve estimation in the context of complex survey data deserves attention. In addition to smoothness, if the models are correctly specified, these alternative approaches might be more efficient in estimating the ROC curve in the context of complex survey data. Our method also assumes that the sampling is noninformative, and further investigation for informative sampling will be worthwhile. There is also little literature exploring the accuracy of a diagnostic test that varies according to a set of characteristics in the context of complex survey data. To address this issue, the estimation of covariate-specific ROC curve for complex survey data is currently under investigation.
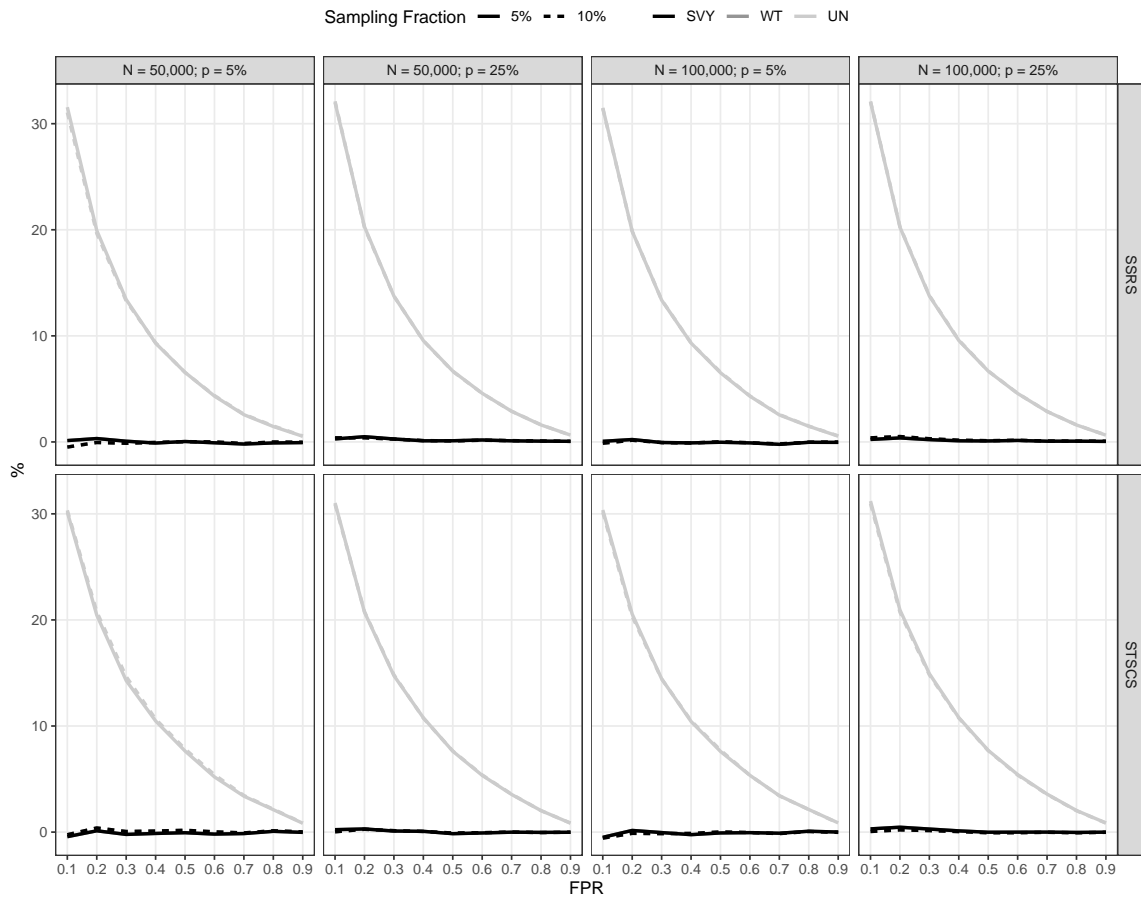
Figure 3.5.1: Relative Bias (in %) of the UN, WT, and SVY estimators for the super-population ROC curve with finite population size $N$, disease proportion $p$, and sampling fraction $\lambda$.
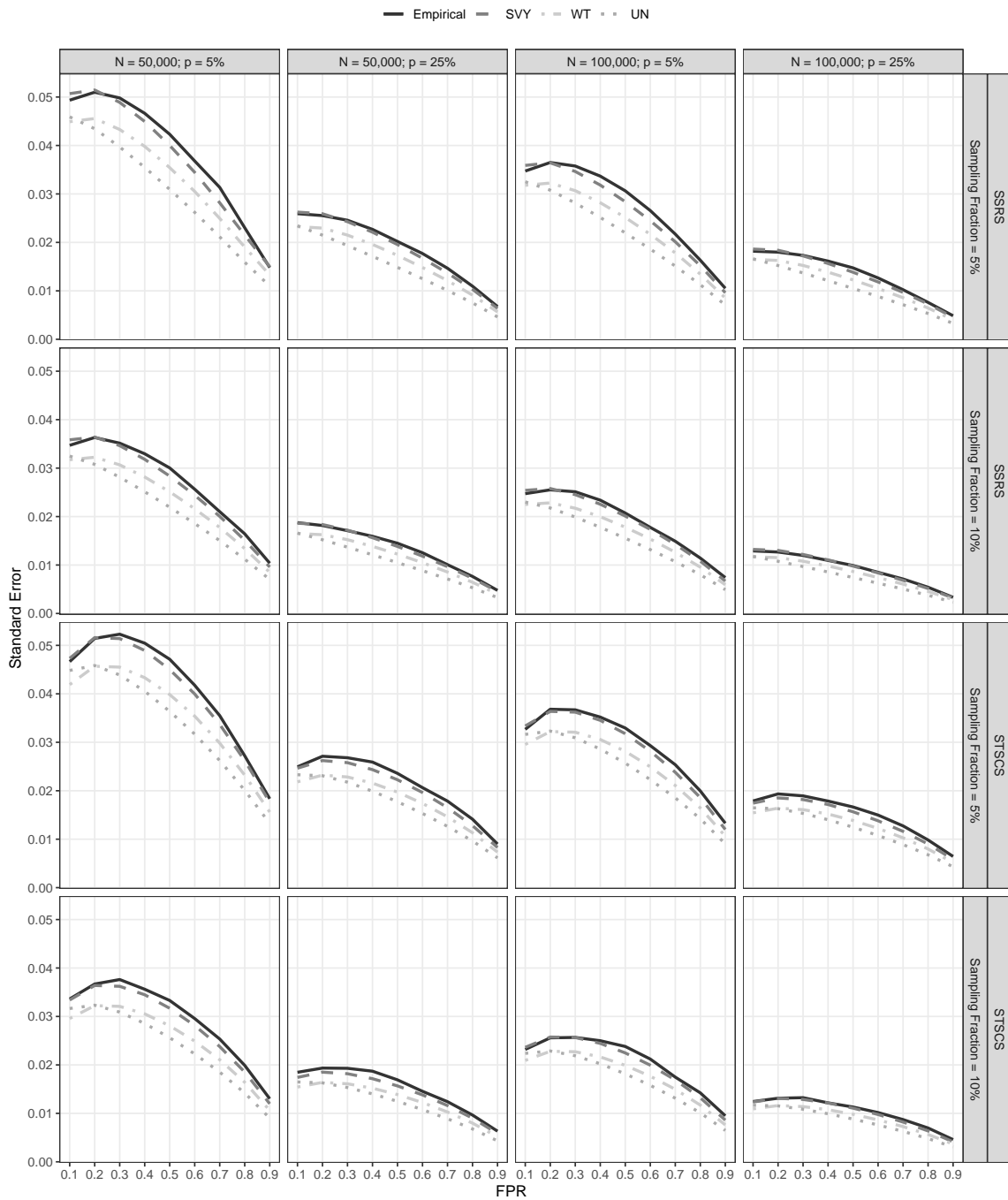
Figure 3.5.2: Empirical and Asymptotic Standard Error (in %) of the UN, WT, and SVY estimators for the super-population ROC curve with finite population size $N$, disease proportion $p$, and sampling fraction $\lambda$.
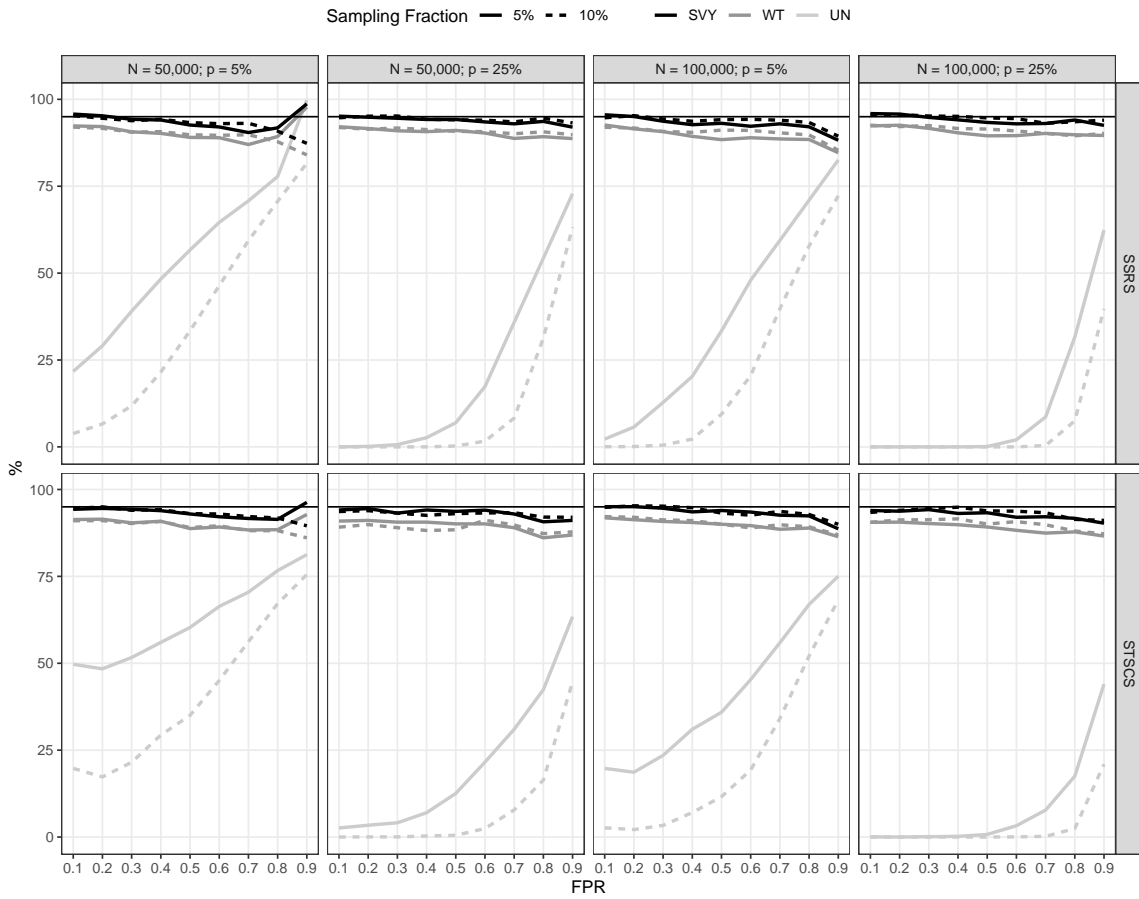
Figure 3.5.3: Coverage Probabilities (in %) of the UN, WT, and SVY estimators for the super-population ROC curve with finite population size $N$, disease proportion $p$, and sampling fraction $\lambda$.
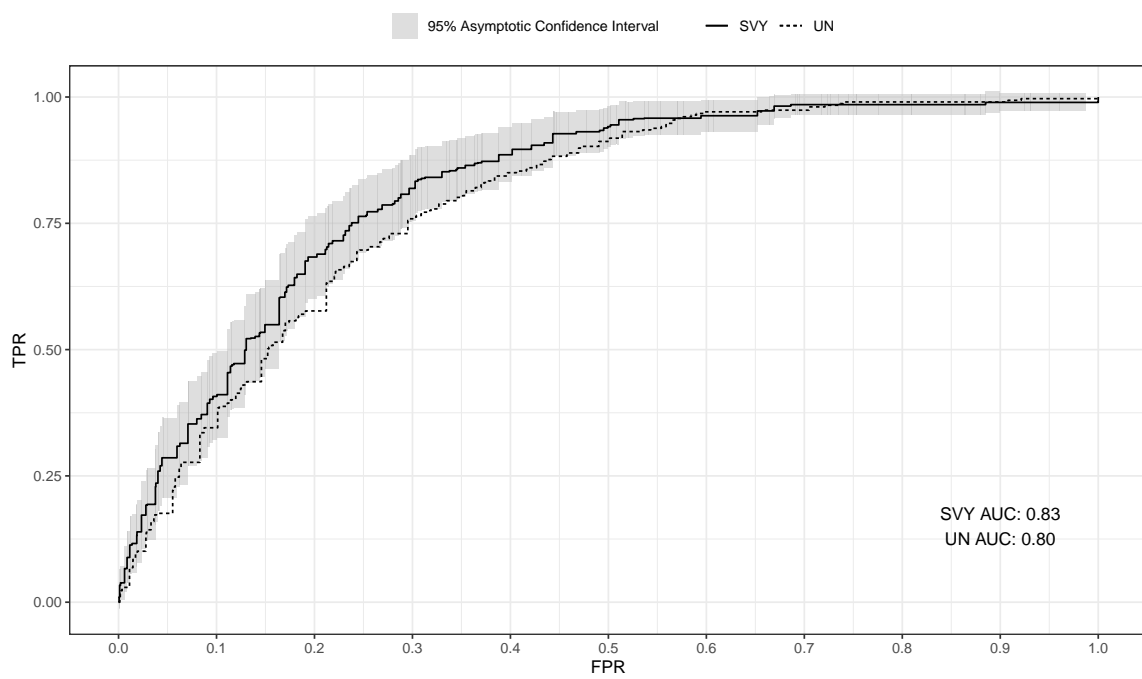
Figure 3.5.4: Unweighted (UN) and survey weighted (SVY) estimates ROC curves and survey weighted 95% confidence interval for NHANES data.

# CHAPTER 4: RECEIVER OPERATING CHARACTERISTIC CURVE FOR COMPLEX SURVEY DATA IN THE PRESENCE OF MISSING BIOMARKER

## 4.1 Introduction

The receiver operating characteristic (ROC) curve is the most popular method to assess the performance of a continuous diagnostic test. The curve is defined as the plot of the false positive rate (1-specificity) versus the true-positive rate (sensitivity) across all possible cutpoints of the diagnostic test. The false-positive rate (FPR) is the proportion of non-diseased individuals that test positive for the disease based on the diagnostic test, and the true-positive rate (TPR) is the proportion of diseased individuals that test positive for the disease. The curve is especially useful to compare the performance of different diagnostic tests and obtain optimal cutpoints for the diagnostic test to minimize the misclassification of diseased and non-diseased individuals. Although we focus on medical diagnosis, the ROC curve is widely used in many binary classification problems. A comprehensive discussion on ROC curves can be found in Pepe (2000b) and Inácio et al. (2021), for example.

The ROC curve has been widely used in the analysis of data arising from complex survey designs. Sample surveys play a critical role in providing essential information in a broad range of areas, serving as an essential resource to guide actions and policies. In the United States, the National Center for Health Statistics (NCHS) is the principal health statistics agency under the Center of Disease Control (CDC), and conducts several population surveys, such as the National Health and Nutrition Examination Survey (NHANES), the National Health Interview Survey (NHIS), and the National Survey of Family Growth (NSFG).

A common issue in large-scale surveys is the presence of missing biomarker values due to various reasons such as study drop-out or loss of information caused by uncontrollable factors. Due to the limited availability of statistical methods, analyses using ROC curves on complex survey data is currently done by ignoring the sampling scheme, and include only participants without missing data on the variables of interest.

Most methods for handling nonresponse fall within inverse probability weighting (IPW) techniques and imputation strategies. IPW methods rely on a model for the response probability given a set of predictors (response model). In contrast, imputation approaches based on statistical prediction rules rely on models for the partially observed variables given the fully observed variables (imputation models). A third class of methods combines IPW and imputation methods, called augmented IPW (AIPW) estimators, and has the property of being doubly robust to model misspecifications.

In this paper, we propose a IPW estimator for the ROC curve for complex survey data with the biomarker being subject to missing at random (MAR). The proposed estimator's asymptotic properties are developed and evaluated through simulation studies, and the method is applied to the National Health and Nutrition Examination Survey (NHANES).

## 4.2 Methods

### 4.2.1 Setup

Classical sampling theory concerns the inference for finite population quantities (parameters). In this context, the design-based (also called randomization-based) inference is often employed, where the characteristics of interest are considered fixed quantities associated with the finite population. The source of randomness is resulting from the sampling scheme, with random variables indicating whether the population unit is contained in the sample. When the questions of interest are based on parameters of a statistical model, the model-based (also called prediction-based) inference is often preferred. In this

framework, the characteristics of interest are considered to be random variables generated from a statistical model.

In this paper, we handle the model-based and design-based inference jointly, using the super-population framework described in Rubin-Bleuer et al. (2005), and followed by Boistard et al. (2017) and Han and Wellner (2021). Under this approach, the finite population is viewed as a realization from a statistical model (superpopulation model), and a sample is drawn from this finite population according to the sampling design. Inference under this approach requires to explicitly account for two sources of randomness: the model-based randomness, accounting for the difference between the finite population parameter and the superpopulation model parameter, and the design-based randomness, accounting for the difference between the sample estimator and the finite population parameter (Pfeffermann, 2000).

Consider a finite population $\mathcal{U}^N$ of size $N$, with corresponding set of indices $U = \{1, \cdots, N\}$. Each index $i \in U$ is associated with a unique vector $(y_i, x_i, z_i) \in \mathbb{R}^p \times \mathbb{R}^k \times \mathbb{R}_+^q$ representing, respectively, the characteristics of interest, the complete auxiliary information, and the sampling design information available at the time of the design of the survey on all units. We assume that $\{(y_i, x_i, z_i)\}_{i=1}^N$ are realizations of random variables $(Y, X, Z)$ defined according to a superpopulation model. We denote $\boldsymbol{y}^N = (y_1, \cdots, y_N)$, $\boldsymbol{Y}^N = (Y_1, \cdots, Y_N)$, $\boldsymbol{x}^N = (x_1, \cdots, x_N)$, and $\boldsymbol{X}^N = (X_1, \cdots, X_N)$, $\boldsymbol{z}^N = (z_1, \cdots, z_N)$, and $\boldsymbol{Z}^N = (Z_1, \cdots, Z_N)$.

We further assume that $Y$ is subject to missing, and let $R$ be the response indicator variable that assumes the value one if $Y$ is observed and zero otherwise. We assume that $\{r_i\}_{i=1}^N$ are realizations of the random variable $R$ defined in the same probability space as $(Y, X, Z)$. We assume that $P(R = 1|Y, X, Z) = P(R = 1|X) = p(X; \phi)$, that is, the response mechanism does not depend on $Y$, and thus, the data are missing at random (MAR). Note that the MAR condition is assumed in the population level as done in Kim

and Riddles (2012), because an individual's decision on wether or not to respond to a survey is at his or her own discretion.

Let $\mathfrak{S}_N = \{s : s \subset U_N\}$ be the collection of subsets of $U_N$ selected under a given sampling scheme and let $\sigma(\mathfrak{S}_N)$ be the $\sigma$-algebra generated by $\mathfrak{S}_N$. A sampling design associated with a sampling scheme is a function $p : \sigma(\mathfrak{S}_N) \times \mathbb{R}_+^{q \times N} \mapsto [0, 1]$ such that

(i) for all $s$ in $\mathfrak{S}_N$, $\boldsymbol{z}^N \mapsto p(s, \boldsymbol{z}^N)$ is Borel-measurable on $\mathbb{R}_+^{q \times N}$;

(ii) for $\boldsymbol{z}^N \in \mathbb{R}_+^{q \times N}$, $A \mapsto p(A, \boldsymbol{z}^N)$ is a probability measure on $\sigma(\mathfrak{S}_N)$.

Note that since $p$ does not depend on $\boldsymbol{y}^N$, only non-informative sampling designs are considered. Similarly to Boistard et al. (2017), for each $\omega \in \Omega$ we define a probability measure $A \mapsto \mathbb{P}_d(A, \omega) = \sum_{s \in A} p(s, \boldsymbol{Z}^N(\omega))$, and we say that $(\mathfrak{S}_N, \sigma(\mathfrak{S}_N), \mathbb{P}_d)$ is the design probability space. We will work on a product probability space $(\mathfrak{S}_N \times \Omega, \sigma(\mathfrak{S}_N) \times \mathfrak{F}, \mathbb{P}_{d,m})$ that includes both the super-population and the design space with probability measure $\mathbb{P}_{d,m}$ defined as $\mathbb{P}_{d,m}(s \times E) = \int_E \mathbb{P}_d(s, \omega) \, \mathrm{d}P_m(\omega)$, with $(s, E) \in \sigma(\mathfrak{S}_N) \times \mathfrak{F}$. We adopt $\mathbb{E}_d$, $\mathbb{E}_m$ and $\mathbb{E}_{d,m}$ to denote the expectation with respect to the probability space $(\mathfrak{S}_N, \sigma(\mathfrak{S}_N), \mathbb{P}_d)$, $(\Omega, \mathfrak{F}, \mathbb{P}_m)$ and $(\mathfrak{S}_N \times \Omega, \sigma(\mathfrak{S}_N) \times \mathfrak{F}, \mathbb{P}_{d,m})$, respectively. For a sample $s$ drawn according to a sampling design $p$, the sampling indicators $\xi_i = I(i \in s)$ are random variables defined on $(\mathfrak{S}_N \times \Omega, \sigma(\mathfrak{S}_N) \times \mathfrak{F}, \mathbb{P}_{d,m})$, with first-order inclusion probabilities defined as $\pi_i(\omega) = \mathbb{E}_{d,m}[\xi_i | \boldsymbol{Z}^N(\omega)]$, and second-order inclusion probabilities defined as $\pi_{ij}(\omega) = \mathbb{E}_{d,m}[\xi_i \xi_j | \boldsymbol{Z}^N(\omega)]$.

### 4.2.2 Assumptions and Notations

Following Han and Wellner (2021), we impose the following assumptions for the asymptotic convergence:

(A1.1) $\min_{1 \leq i \leq N} \pi_i \geq \pi_0 > 0$

(A1.2) $\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left( \frac{\xi_i}{\pi_i} - 1 \right) = O_{\mathbb{P}_{d,m}}(1)$

(A1.3) There exist constant $K > 0$ such that

$$\sup_{N \in \mathbb{N}} \sup_{1 \leq i \neq j \leq N} N|\pi_{ij} - \pi_i \pi_j| \leq K$$

(A1.4) The sample size $n$ increases as the population size $N$ increases, with

$$\lim_{N \to \infty} \frac{n}{N} = \lambda, \quad 0 < \lambda < 1$$

Let $\{V_i\}$ be a sequence of bounded i.i.d random variables defined on $(\Omega, \mathfrak{F}, \mathbb{P}_m)$. Let $S_N^2$ be the design-based variance of the Horvitz-Thompson estimator of the population mean, that is,

$$S_N^2 = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} V_i V_j$$

(A2.1) Suppose that for $N$ sufficiently large

$$\frac{1}{S_N} \left( \frac{1}{N} \sum_{i=1}^{N} \frac{\xi_i}{\pi_i} V_i - \frac{1}{N} \sum_{i=1}^{N} V_i \right) \to N(0, 1), \quad \omega\text{-a.s}$$

in distribution under $\mathbb{P}_d$

(A2.2) The first and second-order inclusion probabilities satisfy

$$\frac{1}{N} \sum_{i=1}^{N} \frac{\pi_{ii} - \pi_i^2}{\pi_i^2} \to \mu_{\pi 1} \quad \text{in } \mathbb{P}_m$$

$$\frac{1}{N} \sum_{i \neq j}^{N} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} \to \mu_{\pi 2} \quad \text{in } \mathbb{P}_m$$

where $\mu_{\pi 1}, \mu_{\pi 2} \in \mathbb{R}$ are nonrandom quantities.

Following Han and Wellner (2021), for $\{\pi_i\}_{i=1}^{N}$, $\{\xi_i\}_{i=1}^{N}$, $\{Y_i\}_{i=1}^{N}$, and a class $\mathcal{F}$ of real functions $f$ we define the Horvitz-Thompson empirical measure as

$$\mathbb{P}_N^{\pi}(f) = \frac{1}{N} \sum_{i=1}^{N} \frac{\xi_i}{\pi_i} f(Y_i), \quad f \in \mathcal{F}, \tag{4.11}$$

41

Similarly, we define the IPW Horvitz-Thompson empirical measure:

$$\mathbb{P}^{\pi}_{N,\phi}(f) = \frac{1}{N} \sum_{i=1}^{N} \frac{\xi_i}{\pi_i} \frac{R_i}{p(X_i; \phi)} f(Y_i), \tag{4.12}$$

where $p(X_i; \phi) = P(R_i = 1 | X_i)$. The corresponding finite-population empirical measures are defined as

$$\mathbb{P}_N(f) = \frac{1}{N} \sum_{i=1}^{N} f(Y_i) \tag{4.13}$$

$$\mathbb{P}_{N,\phi}(f) = \frac{1}{N} \sum_{i=1}^{N} \frac{R_i}{p(X_i; \phi)} f(Y_i). \tag{4.14}$$

### 4.2.3 ROC curve for complex survey sampling in the presence of missing biomarker

Let $\{(Y_i, X_i, Z_i) = (Y_i, D_i, W_i, Z_i) \in \mathbb{R} \times \{0,1\} \times \mathbb{R}^k \times \mathbb{R}^q_+\}_{i=1}^{N}$ be i.i.d realizations of the diagnostic test measure $Y$, the disease indicator $D$, the auxiliary variables $W$, and the sampling design information $Z$. In this paper, we focus on the case where the disease status $D$ is always confirmed, but the diagnostic test values $Y$ might be missing for some units. The auxiliary variables $W$ are also assumed to be complete.

We denote the cumulative distribution function (cdf) of $Y$ conditioned on $D = 0$ as $G$, and similarly, the cdf of $Y$ conditioned on $D = 1$ as $F$. We assume that $F$ and $G$ have continuous probability density functions (pdf) $f$ and $g$, respectively. The ROC curve is defined as the plot of $\{(1 - G(c), 1 - F(c)) : c \in \mathbb{R}\}$, or equivalently, as the plot of $\{(s, R(s)) : s \in [0,1]\}$, where $R(s) = 1 - F \circ G^{-1}(1-s)$, with $G^{-1}(s) = \inf\{x \in \mathbb{R} : G(x) \geq s\}$, and $F \circ G^{-1}(.) \equiv F(G^{-1}(.))$. The area under the ROC curve (AUC-ROC) is $A = \int_0^1 R(s) \, ds$.

The finite-population quantities are defined as:

$$\text{ROC}_{\text{IPW}}(s) = 1 - F_{\text{IPW}} \circ (G_{\text{IPW}})^{-1}(1 - s),$$

$$G_{\text{IPW}}(s) = \left[ \sum_{i=1}^{N} \frac{R_i}{p(X_i; \phi_N)} I(D_i = 0) \right]^{-1} \sum_{i=1}^{N} \frac{R_i}{p(X_i; \phi_N)} I(Y_i \leq s, D_i = 0)$$

$$F_{\text{IPW}}(s) = \left[ \sum_{i=1}^{N} \frac{R_i}{p(X_i; \phi_N)} I(D_i = 1) \right]^{-1} \sum_{i=1}^{N} \frac{R_i}{p(X_i; \phi_N)} I(Y_i \leq s, D_i = 1),$$

with $\phi_N$ being the solution of

$$U_N(\phi) = \sum_{i=1}^{N} \{R_i - p(X_i; \phi)\} h_i(\phi) = \sum_{i=1}^{N} u_i(\phi) = 0,$$

where $h_i(\phi) = \partial \text{logit}(p(X_i; \phi))/\partial \phi = \{p(X_i; \phi)(1 - p(X_i; \phi))\}^{-1}(\partial p(X_i; \phi)/\partial \phi)$, which are unbiased estimating equations for the true parameter $\phi^*$.

Consider a sample $\mathfrak{s}$, consisting of $0 \leq n \leq N$ units drawn from the finite population using a sampling design $p$. The proposed survey-weighted IPW estimator for the ROC curve is:

$$\text{ROC}_{\text{IPW}}^{\pi}(s) = 1 - F_{\text{IPW}}^{\pi} \circ (G_{\text{IPW}}^{\pi})^{-1}(1 - s), \tag{4.15}$$

where

$$G_{\text{IPW}}^{\pi}(s) = \left[ \sum_{i=1}^{N} \frac{\xi_i}{\pi_i} \frac{R_i}{p(X_i; \phi_N^{\pi})} I(D_i = 0) \right]^{-1} \sum_{i=1}^{N} \frac{\xi_i}{\pi_i} \frac{R_i}{p(X_i; \phi_N^{\pi})} I(Y_i \leq s, D_i = 0) \tag{4.16}$$

$$F_{\text{IPW}}^{\pi}(s) = \left[ \sum_{i=1}^{N} \frac{\xi_i}{\pi_i} \frac{R_i}{p(X_i; \phi_N^{\pi})} I(D_i = 1) \right]^{-1} \sum_{i=1}^{N} \frac{\xi_i}{\pi_i} \frac{R_i}{p(X_i; \phi_N^{\pi})} I(Y_i \leq s, D_i = 1), \tag{4.17}$$

with $\phi_N^{\pi}$ being the solution of

$$U_N^{\pi}(\phi) = \sum_{i=1}^{N} \frac{\xi_i}{\pi_i} \{R_i - p(X_i; \phi)\} h_i(\phi) = \sum_{i=1}^{N} \frac{\xi_i}{\pi_i} u_i(\phi) = 0.$$

The estimators $F_{\text{IPW}}^\pi$ and $G_{\text{IPW}}^\pi$ can be seen as ratios of the IPW Horvitz-Thompson empirical measures (4.12) with respect to the class of function $\mathcal{F} = \{f_{s,l}(y,d) = I(y \leq s, d = l) : s \in \mathbb{R}, l \in \{0,1\}\}$. The proposed IPW estimator for the ROC curve depends on the pair $(G_{\text{IPW}}^\pi, F_{\text{IPW}}^\pi)$ through the map $\psi(A, B) = B(A^{-1})$, where $A^{-1}$ is the inverse map of $A$. Combining the results from Han and Wellner (2021) and Functional Delta Method (Vaart and Wellner, 1996) arguments presented in the Appendix, the following result will follow:

**Theorem 4.2.1** (FINITE POPULATION INFERENCE). *Consider the estimators* $F_{\text{IPW}}^\pi$, $G_{\text{IPW}}^\pi$, *and* $\text{ROC}_{\text{IPW}}^\pi$ *as defined in (4.15), (4.16), and (4.17), and let* $\phi^*$ *be the superpopulation parameter. Suppose that conditions (A1.1)-(A2.2) hold.*

(a) *(Survey-weighted propensity score parameter).*

$$\sqrt{N}(\phi_N^\pi - \phi_N) \to I^{-1}(\phi^*)\text{N}(0, V(\phi^*))$$

*where* $I$ *and* $V$ *are such that*

$$I(\phi_0) = \lim_{N \to \infty} N^{-1} \frac{\partial U_N(\phi)}{\partial \phi}\bigg|_{\phi=\phi_0}$$

$$V(\phi_0) = \lim_{N \to \infty} N^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} u_i(\phi_0) u_j(\phi_0)$$

(a) *(Survey-weighted IPW empirical distributions).*

$$\sqrt{N} \begin{bmatrix} G_{\text{IPW}}^\pi - G_{\text{IPW}} \\ F_{\text{IPW}}^\pi - F_{\text{IPW}} \end{bmatrix} \rightsquigarrow \begin{bmatrix} \mathbb{G}_0^\pi \\ \mathbb{G}_1^\pi \end{bmatrix} = \begin{bmatrix} \left\{\mu_{\pi_1}(1-\delta)^{-1}p(\phi^*)^{-1}\right\}^{1/2} B_1(G) \\ \left\{\mu_{\pi_1}\delta^{-1}p(\phi^*)^{-1}\right\}^{1/2} B_2(F) \end{bmatrix},$$

*where* $B_1(.)$ *and* $B_2(.)$ *denote two independent Brownian bridges and* $\delta = P(D = 1)$.

(b) *(Survey-weighted IPW ROC curve). Suppose that* $F$ *and* $G$ *have continuous positive densities* $f$ *and* $g$, *respectively, on* $[G^{-1}(a) - \epsilon, G^{-1}(b) + \epsilon]$ *and that* $f(G^{-1})/g(G^{-1})$

*is bounded on any subinterval $(a, b)$, $0 < a < b < 1$. Then, for $0 < s < 1$*

$$\sqrt{n}\left(F_{\mathrm{IPW}}^{\pi} \circ (G_{\mathrm{IPW}}^{\pi})^{-1}(s) - F_{\mathrm{IPW}} \circ (G_{\mathrm{IPW}})^{-1}(s)\right) \rightsquigarrow$$

$$\rightsquigarrow \sqrt{\frac{\lambda \mu_{\pi_1}}{p(\phi^*)}}\left\{\delta^{-1/2} B_2(F \circ G^{-1}(s)) + (1 - \delta)^{-1/2}\frac{f(G^{-1}(s))}{g(G^{-1}(s))}B_1(s)\right\}$$

*where $B_1(.)$ and $B_2(.)$ denote two independent Brownian bridges. This result implies that $\sqrt{n}(R_{\mathrm{IPW}}^{\pi}(s) - R_{\mathrm{IPW}}(s)) \rightsquigarrow \mathbb{W}(G^{-1}(1 - s))$, where $\mathbb{W}(u)$ is a Gaussian process with mean zero and covariance function $\mathbb{E}_{d,m}\{\mathbb{W}(u)\mathbb{W}(t)\} = \sigma^2(u, t)$ given by*

$$\sigma^2(u, t) = \lambda \mu_{\pi_1} p(\phi^*)^{-1}\left\{\delta^{-1}(F(u \wedge t) - F(u)F(t)) + \right.$$

$$\left. + (1 - \delta)^{-1}\frac{f(u)f(t)}{g(u)g(t)}(G(u \wedge t) - G(u)G(t))\right\} \quad (4.18)$$

The super-population inference of $\phi_N^{\pi}$ in the product space follows from Theorem 6.1 from Rubin-Bleuer et al. (2005). For the ROC curve, the super-population inference can be obtained from the decomposition

$$\sqrt{n}\left(F_{\mathrm{IPW}}^{\pi} \circ (G_{\mathrm{IPW}}^{\pi})^{-1}(s) - F \circ G^{-1}(s)\right) =$$

$$\sqrt{n}\left(F_{\mathrm{IPW}}^{\pi} \circ (G_{\mathrm{IPW}}^{\pi})^{-1}(s) - F_{\mathrm{IPW}} \circ (G_{\mathrm{IPW}})^{-1}(s)\right) + \sqrt{n}\left(F_{\mathrm{IPW}} \circ (G_{\mathrm{IPW}})^{-1}(s) - F \circ G^{-1}(s)\right)$$

From the results presented in Theorem 4.2.1, we have that the first component converges to a zero mean Gaussian process under $\mathbb{P}_{d,m}$ (and $\mathbb{P}_d$). Using similar arguments, combined with classical empirical processes results, we have that the second component also converges to a zero mean Gaussian process under $\mathbb{P}_m$. Theorem 5.1(iii) from Rubin-Bleuer et al. (2005) imply that the two components are asymptotically independent, leading to the following result:

**Theorem 4.2.2** (SUPER POPULATION INFERENCE). *Consider the estimators* $F_{\text{IPW}}^{\pi}$, $G_{\text{IPW}}^{\pi}$, *and* $\text{ROC}_{\text{IPW}}^{\pi}$ *as defined in (4.15), (4.16), and (4.17), and let $\phi^*$ be the superpopulation parameter. Suppose that conditions (A1.1)-(A2.2) hold.*

(a) *(Survey-weighted propensity score parameter).*

$$\sqrt{n}(\phi_N^{\pi} - \phi^*) \to \text{N}(0, \Gamma)$$

*where*

$$\Gamma = I^{-1}(\phi^*) \left[ V(\phi_N) + \lambda I(\phi^*) \right] I^{-1}(\phi^*)$$

$$I(\phi^*) = \lim_{N \to \infty} N^{-1} \frac{\partial U_N(\phi)}{\partial \phi} \Big|_{\phi = \phi_0}$$

$$V(\phi_N) = \lim_{N \to \infty} N^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} u_i(\phi_N) u_j(\phi_N)$$

(a) *(Survey-weighted IPW empirical distributions).*

$$\sqrt{N} \begin{bmatrix} G_{\text{IPW}}^{\pi} - G \\ F_{\text{IPW}}^{\pi} - F \end{bmatrix} \rightsquigarrow \begin{bmatrix} \mathbb{G}_0^{\pi} \\ \mathbb{G}_1^{\pi} \end{bmatrix} = \begin{bmatrix} \left\{ (1 + \mu_{\pi_1})(1 - \delta)^{-1} p(\phi^*)^{-1} \right\}^{1/2} B_1(G) \\ \left\{ (1 + \mu_{\pi_1}) \delta^{-1} p(\phi^*)^{-1} \right\}^{1/2} B_2(F) \end{bmatrix},$$

*where $B_1(.)$ and $B_2(.)$ denote two independent Brownian bridges and $\delta = P(D = 1)$.*

(b) *(Survey-weighted IPW ROC curve). Suppose that $F$ and $G$ have continuous positive densities $f$ and $g$, respectively, on $[G^{-1}(a) - \epsilon, G^{-1}(b) + \epsilon]$ and that $f(G^{-1})/g(G^{-1})$ is bounded on any subinterval $(a, b)$, $0 < a < b < 1$. Then, for $0 < s < 1$*

$$\sqrt{n} \left( F_{\text{IPW}}^{\pi} \circ (G_{\text{IPW}}^{\pi})^{-1}(s) - F \circ G^{-1}(s) \right) \rightsquigarrow$$

$$\rightsquigarrow \sqrt{\frac{\lambda(1 + \mu_{\pi_1})}{p(\phi^*)}} \left\{ \delta^{-1/2} B_2(F \circ G^{-1}(s)) + (1 - \delta)^{-1/2} \frac{f(G^{-1}(s))}{g(G^{-1}(s))} B_1(s) \right\}$$

where $B_1(.)$ and $B_2(.)$ denote two independent Brownian bridges. This result implies that $\sqrt{n}(R_{\text{IPW}}^{\pi}(s) - R(s)) \rightsquigarrow \mathbb{W}(G^{-1}(1-s))$, where $\mathbb{W}(u)$ is a Gaussian process with mean zero and covariance function $\mathbb{E}_{d,m}\{\mathbb{W}(u)\mathbb{W}(t)\} = \sigma^2(u,t)$ given by

$$\sigma^2(u,t) = \lambda(1+\mu_{\pi_1})p(\phi^*)^{-1}\bigg\{\delta^{-1}(F(u \wedge t) - F(u)F(t))+$$
$$+ (1-\delta)^{-1}\frac{f(u)f(t)}{g(u)g(t)}(G(u \wedge t) - G(u)G(t))\bigg\}$$

(4.19)

## 4.3   Simulation Studies

In this simulation study, we investigate the performance of the proposed IPW estimator for the ROC curve under stratified simple random sampling (SSRS) and stratified two-stage cluster sampling (STSCS). For each sampling scheme, a total of 18 scenarios were considered according to different missing proportions ($10\%, 25\%, 50\%$), propensity score model specification (M1: correctly specified, M2: overspecified, M3: underspecified), and sampling fractions ($5\%, 10\%$).

For each scenario, we generated populations of size $N = 100,000$ subdivided in five strata containing 5%, 10%, 25%, 30% and 30% of the observations. For each stratum $h = 1, \cdots, 5$, we generated the auxiliary variables $\boldsymbol{W} = (W_1, W_2, W_3, W_4)^{\top}$ from a multivariate normal distribution with mean $\boldsymbol{0}_4$ and covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{I}_4$. Then, we generated the diagnostic test $X_h = \alpha_0^h D + \boldsymbol{\alpha}_1^{\top}\boldsymbol{W}_0 + \epsilon$, where $D \sim \text{Ber}(0.25)$, $\boldsymbol{W}_0 = (W_1, W_2)^{\top}$, and $\epsilon \sim \text{N}(0,1)$. We set $\alpha_0^1 = 2.33, \alpha_0^2 = 1.81, \alpha_0^3 = 1.19, \alpha_0^4 = 0.74$, $\alpha_0^5 = 0.54, \boldsymbol{\alpha}_1 = \boldsymbol{1}_2$, such that the corresponding AUC for each strata were 0.84, 0.78, 0.69, 0.63, 0.59. For STSCS, $M = 10,000$ clusters of sizes 5, 10 and 15 were generated using quantiles of $X_h + \tau, \tau \sim N(0,1)$. Finally, the missing indicators $R$ were generated from the Bernoulli distribution with probability $p$ such that $\text{logit}(p) = \beta_0 + \beta_1 D + \boldsymbol{\beta}_2^{\top}\boldsymbol{W}_0$,

with $\beta_0 = (2.31, 1.10, -0.12)$ for the missing proportions of (10%, 25%, 50%), $\beta_1 = 0.5$, and $\boldsymbol{\beta}_2 = -0.5 \times \mathbf{1}_2$.

Samples with size determined by the sampling fraction were drawn assuming uniform allocation. For the first stage of STSCS, $m = 125$ and $250$ clusters were sampled from each stratum, and 80% of the observations were sampled from each cluster at the second stage.

We evaluated the performance at $s \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ in terms of Relative Bias (RB), Empirical Standard Error (ESE), Asymptotic Standard Error (ASE), and Coverage Probability for 95% confidence intervals (CP). We compare our method (IPW) to the survey-weighted complete-case ROC curve (SCC), using estimator proposed in Chapter 2.3. We also include unweighted ROC curve (UN), where the sampling weights are ignored, and the asymptotic variance is computed following Hsieh et al. (1996). For the IPW estimator, we use a design-adjusted logistic regression model with covariates $D$ and $\boldsymbol{Z}$ to estimate the propensity scores. The performance of the proposed IPW estimator is evaluated under the following model specifications: (M1) correctly specified model using the correct set of covariates, i.e. $\boldsymbol{Z} = \boldsymbol{W}_0$; (M2) overspecified model using all the auxiliary variables, i.e. $\boldsymbol{Z} = \boldsymbol{W}$; (M3) underspecified model using only one of the covariates, i.e. $\boldsymbol{Z} = W_1$. The simulation results are compared with a population of size $N = 1,000,000$ with no missing observations as a approximation of the super-population quantities.

The results for the RB under SSRS and STSCS are reported in Figure 4.6.1, and Tables 4.6.1 and 4.6.2. Overall, the IPW estimator performs well when using the correctly specified (IPW-M1) and over-specified (IPW-M2) model for the propensity score, with RB never exceeding 1.2%. When the propensity score is under-specified (IPW-M3), we have slightly biased results that worsen as the missing proportion increases. However, the IPW-M3 estimator has less bias than the SCC estimator in all scenarios. Finally, the UN estimator that does not account for the sampling weights and missingness performs poorly, with RB as large as 41% when the missing proportion is 50% in SSRS.

The ESE and ASE under SSRS and STSCS are presented in Figure 4.6.2, and Tables 4.6.3 and 4.6.4. The ESE was obtained by computing the empirical standard deviation of all repetitions of the IPW estimator in each simulation scenario. The ASE was obtained for the IPW and SCC estimators by plugging survey-weighted estimates of $p$, $F$, $G$, $f$, and $g$ into the expression developed in (3.10) and Chapter 2.3. For the UN estimator, unweighted estimates of $p$, $F$, $G$, $f$, and $g$ were plugged into the asymptotic variance expression presented in Hsieh et al. (1996). In general, the ASE for the IPW estimator is closer to the ESE, with higher departures for higher missing proportions. The variance estimators for the SCC and UN estimators underestimate the standard errors as the missing proportion increases.

Figure 4.6.3 and Tables 4.6.5 and 4.6.6 give the CP for the ROC curve. The coverage probabilities based on the IPW estimator are closer to 95%, with higher departures with higher missing proportions and an under-specified model (IPW-M3). The SCC estimator presents coverage probabilities close to 92% at most, with a decrease as the missing proportion increases. Finally, the UN estimator performs poorly due to the large biases and underestimated variances.

## 4.4 Application

Diabetes and its complications are major causes of morbidity and mortality worldwide. Currently, clinical practice guidelines recommend screening for pre-diabetes and type 2 diabetes with an informal assessment of risk factors or validated risk calculator in asymptomatic adults to guide providers on whether performing a definitive diagnostic test is necessary (Draznin et al., 2022). The current risk assessment tool used by the American Diabetes Association (ADA) to screen for pre-diabetes and type 2 diabetes is adapted from the algorithm developed in Bang et al. (2009) to estimate the risk of undiagnosed diabetes.

In this application, we wish to evaluate the discrimination of the algorithm developed by Bang et al. (2009) using the National Health and Nutrition Examination Survey (NHANES) between 1999-2006. NHANES is an annual survey conducted by the Centers for Disease Control and Prevention's (CDC) National Center for Health Statistics (NCHS) that utilizes a complex, multistage probability sampling design to select a representative sample of the non-institutionalized resident population of the United States.

Similarly as presented in Bang et al. (2009), we consider participants aged 20 years or more, excluding pregnant women, that had fasting plasma glucose (FPG) results. The participants are classified into four groups of diabetes status: known diabetes (if answered "yes" to the question "Other than during pregnancy, have you ever been told by a doctor or health professional that you have diabetes or sugar diabetes?"), normal glucose metabolism (FPG < 100 mg/dL), pre-diabetes (FPG 100-125 mg/dL), and undiagnosed diabetes (FPG > 125 mg/dL). The participants classified as "known diabetes" are not considered in the analysis, and the undiagnosed diabetes was used as the endpoint. The risk score was computed using age (< 40, 40-49, 50-59, > 59), sex (female, male), family history of diabetes (yes, no), history of hypertension (yes, no), obesity (not overweight, overweight, obese, extremely obese), physically active (yes, no) according to the model presented in Bang et al. (2009).

In the 1999-2006 NHANES, 20,159 non-pregnant adults aged 20 years or more were enrolled. Out of this sample, 17,696 observations had FPG results and were classified as either normal glucose metabolism, pre-diabetes, and undiagnosed diabetes. The proportion of undiagnosed diabetes was 2.3% (95% CI: 1.9, 2.7), and 10,511 (59%) observations had missing value for the risk of undiagnosed diabetes (Table 4.6.7).

Figure 4.6.4 shows the estimated ROC curves and their corresponding AUC according to three methods. The first method is the unweighted estimate, where the design weights are ignored. The second method is the survey-weighted ROC curve proposed in Chapter 2.3 that accounts for the design weights, but uses only the complete-cases

for the estimation. Finally, the third method is the IPW estimator proposed in this Chapter, with the propensity score modelled using survey-weighted logistic regression model including gender, race, age, obesity, weekly minutes of physical activity and level of adherence to the 2008 Physical Activity Guidelines for Americans (U.S. Department of Health and Human Services, 2008). From the plot, we see that the unweighted ROC curve provides the smallest estimates, followed by the complete-case estimate and IPW estimate, with the most considerable discrepancies between FPR 0.1-0.5.

## 4.5 Discussion

In this paper, we proposed a design-adjusted IPW estimator for the ROC curve to accommodate the case where the diagnostic test $Y$ is missing. We developed the asymptotic properties of the estimator using the super-population framework described in Rubin-Bleuer et al. (2005), combined with the general results for Horwitz-Thompson empirical measure from Boistard et al. (2017) and Han and Wellner (2021). Our proposed estimator presented a better performance in the simulation studies compared to methods that do not account for the complex survey design and missingness. Our approach was then applied to NHANES to evaluate the discriminatory ability of a traditional risk calculator for undiagnosed diabetes.

The proposed estimator serves as a first step toward accounting for the common issue of nonresponse in large-scale surveys when computing ROC curves. A natural extension of our proposed estimator is the Augmented IPW (AIPW) estimator, where a model is assumed for the diagnostic test, in addition to the model for the missing mechanism. In this context, one can obtain an estimator that is doubly robust to model specification. Alternative methods using imputation for complex survey data might also be a possibility.

In the context of ROC curves, there is also the issue concerning the missingness of the disease status $D$, also referred to as verification bias. The development of estimators

to accommodate this second possibility of incomplete data for ROC curves may deserve attention.

## 4.6 Figures and Tables
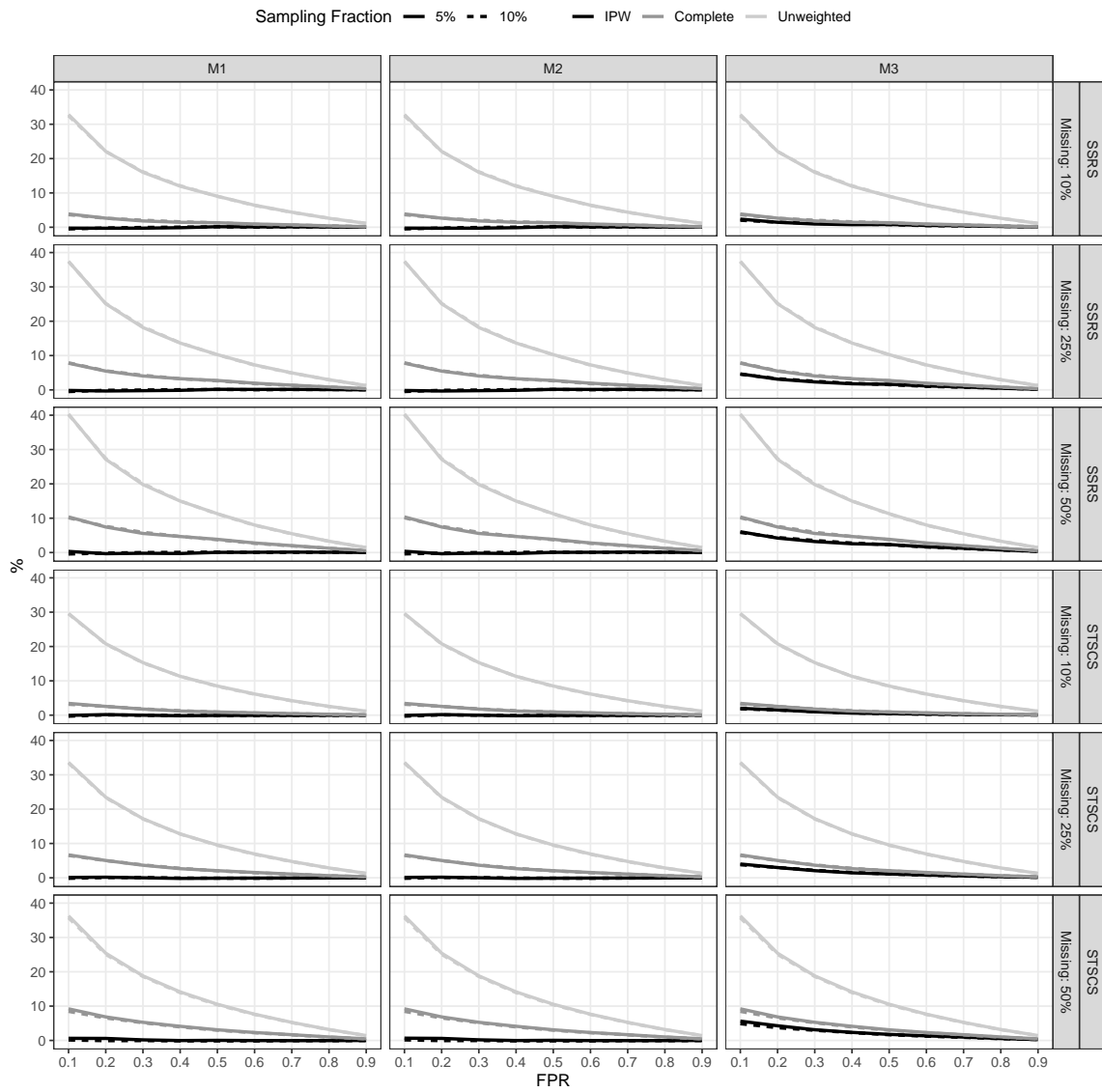


Figure 4.6.1: Relative Bias (in %) of the IPW, CC, and UN estimators for the super population ROC curve.
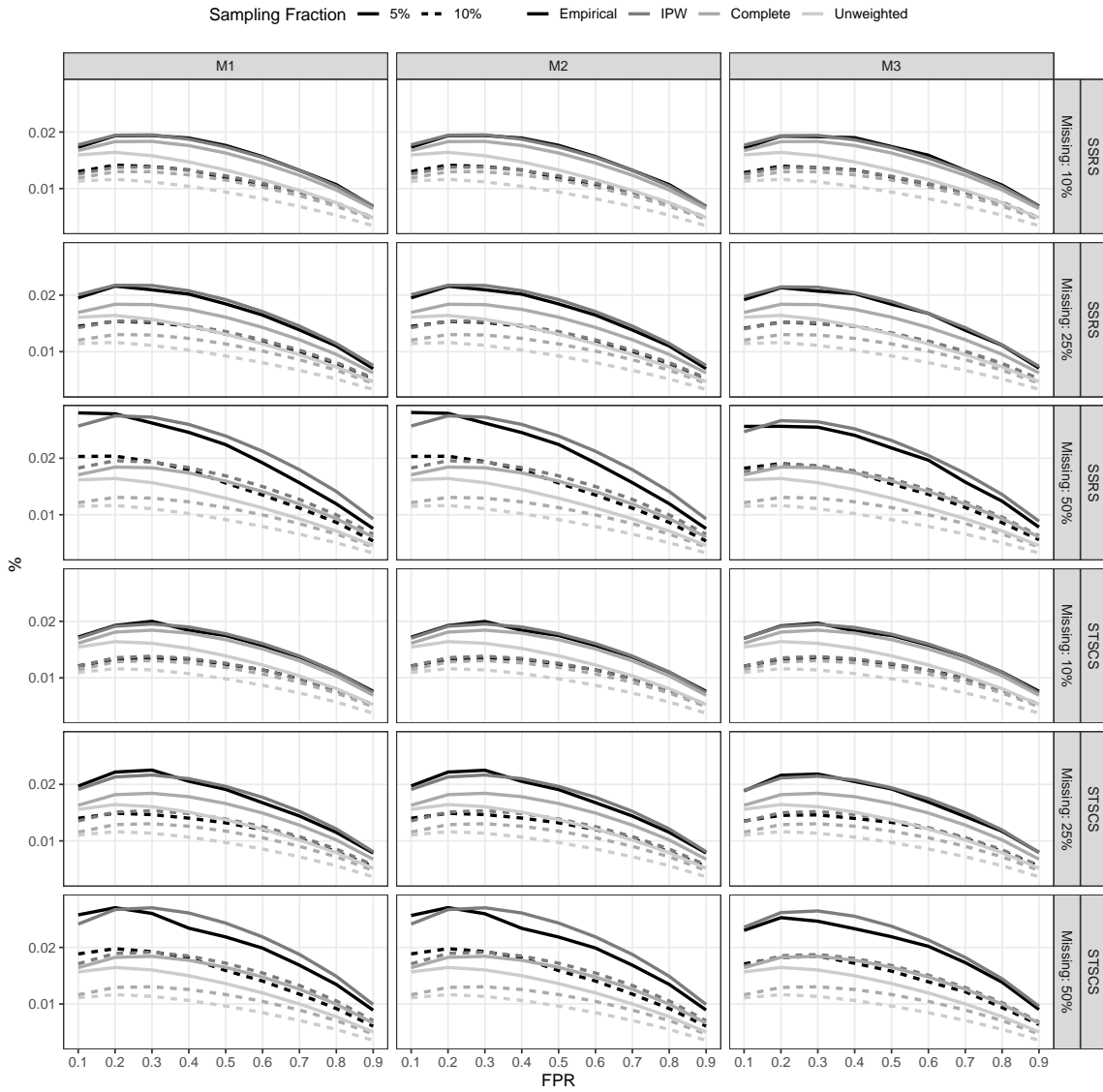
Figure 4.6.2: Empirical and Asymptotic Standard Errors of the IPW, CC, and UN estimators for the super population ROC curve.
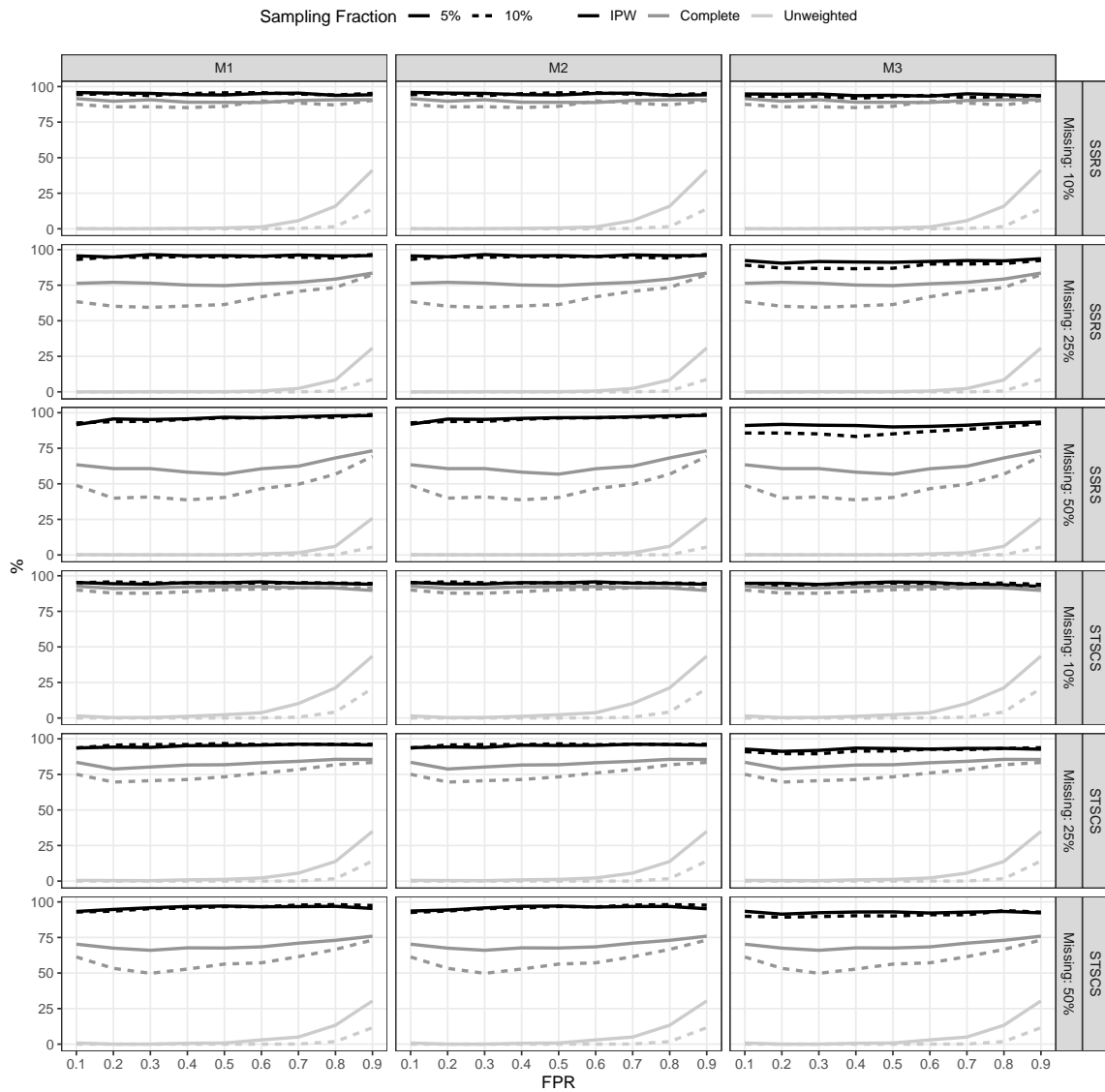
Figure 4.6.3: Coverage Probabilities (in %) of the IPW, CC, and UN estimators for the super-population ROC curve.

Table 4.6.1: Relative Bias (in %) of the IPW, CC, and UN estimators for the super-population ROC curve under SSRS.

| Missing | FPR | Sampling fraction: 5% | | | | | Sampling fraction: 10% | | | | |
|---------|-----|--------|--------|--------|-----|------|--------|--------|--------|------|------|
| | | IPW-M1 | IPW-M2 | IPW-M3 | CC | UN | IPW-M1 | IPW-M2 | IPW-M3 | CC | UN |
| 10% | 0.1 | 0.0 | 0.0 | 2.6 | 4.1 | 33.2 | -0.2 | -0.2 | 2.3 | 3.9 | 32.8 |
| | 0.2 | -0.4 | -0.4 | 1.3 | 2.5 | 22.0 | -0.3 | -0.3 | 1.4 | 2.5 | 21.9 |
| | 0.3 | -0.4 | -0.4 | 0.9 | 1.7 | 15.9 | -0.1 | -0.1 | 1.1 | 2.0 | 16.1 |
| | 0.4 | -0.3 | -0.3 | 0.5 | 1.2 | 11.7 | -0.2 | -0.2 | 0.7 | 1.4 | 11.9 |
| | 0.5 | -0.1 | -0.1 | 0.6 | 1.1 | 8.8 | -0.1 | -0.1 | 0.6 | 1.1 | 8.8 |
| | 0.6 | 0.2 | 0.2 | 0.6 | 1.0 | 6.5 | 0.1 | 0.1 | 0.6 | 0.9 | 6.5 |
| | 0.7 | 0.0 | 0.0 | 0.3 | 0.6 | 4.3 | 0.0 | 0.0 | 0.3 | 0.5 | 4.2 |
| | 0.8 | 0.1 | 0.1 | 0.2 | 0.4 | 2.6 | 0.1 | 0.1 | 0.3 | 0.4 | 2.6 |
| | 0.9 | 0.0 | 0.0 | 0.1 | 0.2 | 1.1 | 0.0 | 0.0 | 0.1 | 0.2 | 1.1 |
| 25% | 0.1 | 0.1 | 0.1 | 4.9 | 8.1 | 37.8 | -0.2 | -0.2 | 4.8 | 8.2 | 37.7 |
| | 0.2 | -0.4 | -0.4 | 3.0 | 5.4 | 25.0 | -0.2 | -0.2 | 3.2 | 5.4 | 25.1 |
| | 0.3 | -0.3 | -0.3 | 2.2 | 4.0 | 18.1 | -0.1 | -0.1 | 2.4 | 4.2 | 18.3 |
| | 0.4 | -0.3 | -0.3 | 1.6 | 3.0 | 13.4 | -0.1 | -0.1 | 1.7 | 3.1 | 13.5 |
| | 0.5 | -0.1 | 0.0 | 1.4 | 2.5 | 10.0 | -0.1 | -0.1 | 1.3 | 2.4 | 10.0 |
| | 0.6 | 0.2 | 0.2 | 1.2 | 2.0 | 7.4 | 0.1 | 0.1 | 1.1 | 1.9 | 7.3 |
| | 0.7 | 0.1 | 0.1 | 0.8 | 1.3 | 4.9 | 0.0 | 0.0 | 0.7 | 1.2 | 4.8 |
| | 0.8 | 0.1 | 0.1 | 0.5 | 0.8 | 3.0 | 0.1 | 0.1 | 0.5 | 0.8 | 2.9 |
| | 0.9 | 0.0 | 0.0 | 0.2 | 0.4 | 1.3 | 0.0 | 0.0 | 0.2 | 0.4 | 1.3 |
| 50% | 0.1 | 0.6 | 0.6 | 6.3 | 10.7 | 40.7 | -0.2 | -0.1 | 6.1 | 10.4 | 40.6 |
| | 0.2 | -0.4 | -0.5 | 4.1 | 7.3 | 27.0 | -0.3 | -0.3 | 4.3 | 7.5 | 27.3 |
| | 0.3 | -0.3 | -0.3 | 3.1 | 5.5 | 19.7 | -0.1 | -0.1 | 3.4 | 5.8 | 20.0 |
| | 0.4 | -0.5 | -0.5 | 2.4 | 4.5 | 14.8 | -0.1 | -0.1 | 2.6 | 4.5 | 14.8 |
| | 0.5 | -0.2 | -0.2 | 2.1 | 3.6 | 11.1 | -0.1 | -0.1 | 2.1 | 3.5 | 11.0 |
| | 0.6 | 0.2 | 0.2 | 1.8 | 2.9 | 8.1 | 0.1 | 0.1 | 1.7 | 2.8 | 8.0 |
| | 0.7 | 0.0 | 0.0 | 1.1 | 1.9 | 5.4 | 0.0 | 0.0 | 1.1 | 1.8 | 5.3 |
| | 0.8 | 0.1 | 0.1 | 0.7 | 1.2 | 3.3 | 0.1 | 0.1 | 0.7 | 1.2 | 3.3 |
| | 0.9 | 0.0 | 0.0 | 0.4 | 0.6 | 1.4 | 0.0 | 0.0 | 0.3 | 0.5 | 1.4 |

Table 4.6.2: Relative Bias (in %) of the IPW, CC, and UN estimators for the super-population ROC curve under STSCS.

| Missing | FPR | Sampling fraction: 5% | | | | | Sampling fraction: 10% | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | IPW-M1 | IPW-M2 | IPW-M3 | CC | UN | IPW-M1 | IPW-M2 | IPW-M3 | CC | UN |
| 10% | 0.1 | 0.5 | 0.5 | 2.5 | 4.0 | 30.3 | 0.2 | 0.2 | 2.4 | 3.7 | 30.2 |
| | 0.2 | 0.1 | 0.1 | 1.5 | 2.5 | 20.7 | 0.0 | 0.0 | 1.4 | 2.5 | 20.6 |
| | 0.3 | -0.2 | -0.2 | 0.8 | 1.5 | 15.0 | -0.2 | -0.2 | 0.8 | 1.6 | 15.0 |
| | 0.4 | -0.2 | -0.2 | 0.6 | 1.2 | 11.3 | -0.1 | -0.1 | 0.7 | 1.3 | 11.3 |
| | 0.5 | 0.1 | 0.1 | 0.6 | 1.1 | 8.7 | 0.0 | 0.0 | 0.6 | 1.1 | 8.6 |
| | 0.6 | 0.1 | 0.1 | 0.5 | 0.8 | 6.3 | 0.0 | 0.0 | 0.4 | 0.8 | 6.3 |
| | 0.7 | 0.0 | 0.0 | 0.3 | 0.5 | 4.3 | 0.0 | 0.0 | 0.3 | 0.5 | 4.2 |
| | 0.8 | 0.0 | 0.0 | 0.1 | 0.3 | 2.6 | -0.1 | -0.1 | 0.1 | 0.2 | 2.5 |
| | 0.9 | 0.0 | 0.0 | 0.0 | 0.1 | 1.1 | 0.0 | 0.0 | 0.0 | 0.1 | 1.1 |
| 25% | 0.1 | 0.6 | 0.7 | 4.6 | 7.2 | 34.3 | 0.4 | 0.4 | 4.3 | 7.1 | 34.1 |
| | 0.2 | 0.1 | 0.1 | 2.9 | 5.0 | 23.4 | 0.0 | 0.0 | 3.0 | 5.0 | 23.3 |
| | 0.3 | -0.2 | -0.2 | 1.8 | 3.4 | 17.0 | -0.2 | -0.2 | 2.0 | 3.5 | 16.9 |
| | 0.4 | -0.2 | -0.2 | 1.4 | 2.6 | 12.8 | -0.1 | -0.1 | 1.6 | 2.8 | 12.7 |
| | 0.5 | 0.0 | 0.0 | 1.3 | 2.2 | 9.7 | 0.1 | 0.1 | 1.3 | 2.2 | 9.6 |
| | 0.6 | 0.1 | 0.1 | 1.0 | 1.7 | 7.1 | 0.0 | 0.0 | 0.9 | 1.6 | 7.0 |
| | 0.7 | 0.0 | 0.0 | 0.6 | 1.2 | 4.8 | 0.0 | 0.0 | 0.6 | 1.1 | 4.8 |
| | 0.8 | -0.1 | -0.1 | 0.3 | 0.6 | 2.9 | -0.1 | -0.1 | 0.3 | 0.6 | 2.8 |
| | 0.9 | 0.0 | 0.0 | 0.1 | 0.3 | 1.3 | 0.0 | 0.0 | 0.1 | 0.3 | 1.3 |
| 50% | 0.1 | 1.2 | 1.2 | 6.2 | 9.8 | 37.0 | 0.6 | 0.7 | 5.4 | 9.0 | 36.4 |
| | 0.2 | 0.5 | 0.5 | 4.2 | 6.8 | 25.3 | -0.1 | -0.1 | 3.7 | 6.5 | 25.0 |
| | 0.3 | -0.1 | 0.0 | 2.9 | 5.0 | 18.5 | -0.3 | -0.3 | 2.7 | 4.9 | 18.4 |
| | 0.4 | 0.0 | 0.0 | 2.3 | 4.1 | 14.1 | -0.1 | -0.1 | 2.3 | 3.9 | 13.9 |
| | 0.5 | 0.2 | 0.2 | 1.9 | 3.3 | 10.7 | 0.0 | 0.0 | 1.8 | 3.1 | 10.5 |
| | 0.6 | 0.1 | 0.1 | 1.5 | 2.5 | 7.8 | 0.0 | 0.0 | 1.4 | 2.4 | 7.7 |
| | 0.7 | 0.0 | 0.0 | 1.1 | 1.7 | 5.3 | 0.0 | 0.0 | 1.0 | 1.7 | 5.3 |
| | 0.8 | 0.0 | 0.0 | 0.6 | 1.0 | 3.2 | -0.1 | -0.1 | 0.5 | 1.0 | 3.1 |
| | 0.9 | 0.0 | 0.0 | 0.2 | 0.5 | 1.4 | -0.1 | -0.1 | 0.2 | 0.4 | 1.4 |

Table 4.6.3: Estimates of empirical (EMP) and asymptotic standard error of the IPW, CC, and UN estimators for the super-population ROC curve under SSRS.

| FPR | Method | Sampling proportion: 5% | | | | | | | | | Sampling proportion: 10% | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Missing: 10% | | | Missing: 25% | | | Missing: 50% | | | Missing: 10% | | | Missing: 25% | | | Missing: 50% | | |
| | | M1 | M2 | M3 | M1 | M2 | M3 | M1 | M2 | M3 | M1 | M2 | M3 | M1 | M2 | M3 | M1 | M2 | M3 |
| 0.1 | EMP | 0.017 | 0.017 | 0.017 | 0.020 | 0.020 | 0.019 | 0.028 | 0.028 | 0.026 | 0.013 | 0.013 | 0.013 | 0.014 | 0.014 | 0.014 | 0.020 | 0.020 | 0.018 |
| | IPW | 0.018 | 0.018 | 0.018 | 0.020 | 0.020 | 0.020 | 0.026 | 0.026 | 0.025 | 0.013 | 0.013 | 0.013 | 0.014 | 0.014 | 0.014 | 0.018 | 0.018 | 0.018 |
| | CC | 0.017 | 0.017 | 0.017 | 0.017 | 0.017 | 0.017 | 0.017 | 0.017 | 0.017 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 |
| | UN | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 | 0.012 | 0.012 | 0.012 |
| 0.2 | EMP | 0.019 | 0.019 | 0.019 | 0.022 | 0.022 | 0.021 | 0.028 | 0.028 | 0.026 | 0.014 | 0.014 | 0.014 | 0.015 | 0.015 | 0.015 | 0.020 | 0.020 | 0.019 |
| | IPW | 0.019 | 0.019 | 0.019 | 0.022 | 0.022 | 0.021 | 0.027 | 0.027 | 0.027 | 0.014 | 0.014 | 0.014 | 0.015 | 0.015 | 0.015 | 0.020 | 0.020 | 0.019 |
| | CC | 0.018 | 0.018 | 0.018 | 0.018 | 0.018 | 0.018 | 0.018 | 0.018 | 0.018 | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 |
| | UN | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 |
| 0.3 | EMP | 0.019 | 0.019 | 0.019 | 0.021 | 0.021 | 0.021 | 0.026 | 0.026 | 0.025 | 0.014 | 0.014 | 0.014 | 0.015 | 0.015 | 0.015 | 0.019 | 0.019 | 0.019 |
| | IPW | 0.020 | 0.020 | 0.019 | 0.022 | 0.022 | 0.021 | 0.027 | 0.027 | 0.026 | 0.014 | 0.014 | 0.014 | 0.015 | 0.015 | 0.015 | 0.019 | 0.019 | 0.019 |
| | CC | 0.018 | 0.018 | 0.018 | 0.018 | 0.018 | 0.018 | 0.018 | 0.018 | 0.018 | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 |
| | UN | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 |
| 0.4 | EMP | 0.019 | 0.019 | 0.019 | 0.020 | 0.020 | 0.020 | 0.025 | 0.025 | 0.024 | 0.013 | 0.013 | 0.013 | 0.015 | 0.015 | 0.015 | 0.018 | 0.018 | 0.018 |
| | IPW | 0.019 | 0.019 | 0.019 | 0.021 | 0.021 | 0.020 | 0.026 | 0.026 | 0.025 | 0.013 | 0.013 | 0.013 | 0.015 | 0.015 | 0.014 | 0.018 | 0.018 | 0.018 |
| | CC | 0.018 | 0.018 | 0.018 | 0.017 | 0.017 | 0.017 | 0.017 | 0.017 | 0.017 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 |
| | UN | 0.015 | 0.015 | 0.015 | 0.015 | 0.015 | 0.015 | 0.014 | 0.014 | 0.014 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 |
| 0.5 | EMP | 0.018 | 0.018 | 0.017 | 0.018 | 0.018 | 0.018 | 0.022 | 0.022 | 0.022 | 0.012 | 0.012 | 0.012 | 0.013 | 0.013 | 0.013 | 0.016 | 0.016 | 0.015 |
| | IPW | 0.017 | 0.017 | 0.017 | 0.019 | 0.019 | 0.019 | 0.024 | 0.024 | 0.023 | 0.012 | 0.012 | 0.012 | 0.014 | 0.014 | 0.013 | 0.017 | 0.017 | 0.016 |
| | CC | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.012 | 0.012 | 0.012 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 |
| | UN | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 |
| 0.6 | EMP | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.017 | 0.019 | 0.019 | 0.020 | 0.011 | 0.011 | 0.011 | 0.012 | 0.012 | 0.012 | 0.013 | 0.013 | 0.014 |
| | IPW | 0.016 | 0.016 | 0.015 | 0.017 | 0.017 | 0.017 | 0.021 | 0.021 | 0.020 | 0.011 | 0.011 | 0.011 | 0.012 | 0.012 | 0.012 | 0.015 | 0.015 | 0.014 |
| | CC | 0.015 | 0.015 | 0.015 | 0.014 | 0.014 | 0.014 | 0.014 | 0.014 | 0.014 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 |
| | UN | 0.012 | 0.012 | 0.012 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 |
| 0.7 | EMP | 0.013 | 0.013 | 0.013 | 0.014 | 0.014 | 0.014 | 0.016 | 0.016 | 0.016 | 0.009 | 0.009 | 0.009 | 0.010 | 0.010 | 0.010 | 0.011 | 0.011 | 0.011 |
| | IPW | 0.013 | 0.013 | 0.013 | 0.015 | 0.015 | 0.014 | 0.018 | 0.018 | 0.017 | 0.009 | 0.009 | 0.009 | 0.010 | 0.010 | 0.010 | 0.013 | 0.013 | 0.012 |
| | CC | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 | 0.008 | 0.008 | 0.008 |
| | UN | 0.010 | 0.010 | 0.010 | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 | 0.007 | 0.007 | 0.007 | 0.007 | 0.007 | 0.007 | 0.007 | 0.007 | 0.007 |
| 0.8 | EMP | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 | 0.012 | 0.012 | 0.012 | 0.007 | 0.007 | 0.008 | 0.008 | 0.008 | 0.008 | 0.009 | 0.009 | 0.009 |
| | IPW | 0.010 | 0.010 | 0.010 | 0.011 | 0.011 | 0.011 | 0.014 | 0.014 | 0.014 | 0.007 | 0.007 | 0.007 | 0.008 | 0.008 | 0.008 | 0.010 | 0.010 | 0.010 |
| | CC | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.009 | 0.009 | 0.009 | 0.007 | 0.007 | 0.007 | 0.007 | 0.007 | 0.007 | 0.007 | 0.007 | 0.007 |
| | UN | 0.008 | 0.008 | 0.008 | 0.007 | 0.007 | 0.007 | 0.007 | 0.007 | 0.007 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 |
| 0.9 | EMP | 0.007 | 0.007 | 0.007 | 0.007 | 0.007 | 0.007 | 0.008 | 0.008 | 0.008 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.006 |
| | IPW | 0.007 | 0.007 | 0.007 | 0.008 | 0.008 | 0.007 | 0.009 | 0.009 | 0.009 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.007 | 0.007 | 0.006 |
| | CC | 0.006 | 0.006 | 0.006 | 0.006 | 0.006 | 0.006 | 0.006 | 0.006 | 0.006 | 0.005 | 0.005 | 0.005 | 0.004 | 0.004 | 0.004 | 0.004 | 0.004 | 0.004 |
| | UN | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 |

Table 4.6.4: Estimates of empirical (EMP) and asymptotic standard error of the IPW, CC, and UN estimators for the super-population ROC curve under STSCS.

| FPR | Method | Sampling proportion: 5% | | | | | | | | | Sampling proportion: 10% | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Missing: 10% | | | Missing: 25% | | | Missing: 50% | | | Missing: 10% | | | Missing: 25% | | | Missing: 50% | | |
| | | M1 | M2 | M3 | M1 | M2 | M3 | M1 | M2 | M3 | M1 | M2 | M3 | M1 | M2 | M3 | M1 | M2 | M3 |
| 0.1 | EMP | 0.017 | 0.017 | 0.017 | 0.020 | 0.020 | 0.019 | 0.026 | 0.026 | 0.023 | 0.012 | 0.012 | 0.012 | 0.014 | 0.014 | 0.014 | 0.019 | 0.019 | 0.017 |
| | IPW | 0.017 | 0.017 | 0.017 | 0.019 | 0.019 | 0.019 | 0.024 | 0.024 | 0.024 | 0.012 | 0.012 | 0.012 | 0.014 | 0.014 | 0.013 | 0.017 | 0.017 | 0.017 |
| | CC | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.011 | 0.011 | 0.011 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 |
| | UN | 0.015 | 0.015 | 0.015 | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 |
| 0.2 | EMP | 0.019 | 0.019 | 0.019 | 0.022 | 0.022 | 0.022 | 0.027 | 0.027 | 0.025 | 0.013 | 0.013 | 0.013 | 0.015 | 0.015 | 0.015 | 0.020 | 0.020 | 0.018 |
| | IPW | 0.019 | 0.019 | 0.019 | 0.021 | 0.021 | 0.021 | 0.027 | 0.027 | 0.026 | 0.014 | 0.014 | 0.014 | 0.015 | 0.015 | 0.015 | 0.019 | 0.019 | 0.019 |
| | CC | 0.018 | 0.018 | 0.018 | 0.018 | 0.018 | 0.018 | 0.018 | 0.018 | 0.018 | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 |
| | UN | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 |
| 0.3 | EMP | 0.020 | 0.020 | 0.020 | 0.023 | 0.022 | 0.022 | 0.026 | 0.026 | 0.025 | 0.013 | 0.013 | 0.014 | 0.015 | 0.015 | 0.015 | 0.019 | 0.019 | 0.018 |
| | IPW | 0.020 | 0.020 | 0.019 | 0.022 | 0.022 | 0.021 | 0.027 | 0.027 | 0.026 | 0.014 | 0.014 | 0.014 | 0.015 | 0.015 | 0.015 | 0.019 | 0.019 | 0.019 |
| | CC | 0.018 | 0.018 | 0.018 | 0.018 | 0.018 | 0.018 | 0.018 | 0.018 | 0.018 | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 |
| | UN | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 |
| 0.4 | EMP | 0.018 | 0.018 | 0.018 | 0.021 | 0.021 | 0.021 | 0.023 | 0.023 | 0.023 | 0.013 | 0.013 | 0.013 | 0.014 | 0.014 | 0.014 | 0.018 | 0.018 | 0.017 |
| | IPW | 0.019 | 0.019 | 0.019 | 0.021 | 0.021 | 0.021 | 0.026 | 0.026 | 0.026 | 0.013 | 0.013 | 0.013 | 0.015 | 0.015 | 0.015 | 0.018 | 0.018 | 0.018 |
| | CC | 0.018 | 0.018 | 0.018 | 0.018 | 0.018 | 0.018 | 0.018 | 0.018 | 0.018 | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 |
| | UN | 0.015 | 0.015 | 0.015 | 0.015 | 0.015 | 0.015 | 0.015 | 0.015 | 0.015 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 |
| 0.5 | EMP | 0.017 | 0.017 | 0.017 | 0.019 | 0.019 | 0.019 | 0.022 | 0.022 | 0.022 | 0.012 | 0.012 | 0.012 | 0.013 | 0.013 | 0.013 | 0.016 | 0.016 | 0.016 |
| | IPW | 0.018 | 0.018 | 0.018 | 0.020 | 0.020 | 0.019 | 0.024 | 0.024 | 0.024 | 0.013 | 0.013 | 0.013 | 0.014 | 0.014 | 0.014 | 0.017 | 0.017 | 0.017 |
| | CC | 0.017 | 0.017 | 0.017 | 0.017 | 0.017 | 0.017 | 0.016 | 0.016 | 0.016 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 |
| | UN | 0.014 | 0.014 | 0.014 | 0.014 | 0.014 | 0.014 | 0.014 | 0.014 | 0.014 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 |
| 0.6 | EMP | 0.016 | 0.016 | 0.016 | 0.017 | 0.017 | 0.017 | 0.020 | 0.020 | 0.020 | 0.011 | 0.011 | 0.011 | 0.012 | 0.012 | 0.012 | 0.014 | 0.014 | 0.014 |
| | IPW | 0.016 | 0.016 | 0.016 | 0.018 | 0.018 | 0.017 | 0.022 | 0.022 | 0.021 | 0.011 | 0.011 | 0.011 | 0.013 | 0.013 | 0.012 | 0.015 | 0.015 | 0.015 |
| | CC | 0.015 | 0.015 | 0.015 | 0.015 | 0.015 | 0.015 | 0.015 | 0.015 | 0.015 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 | 0.010 | 0.010 | 0.010 |
| | UN | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 | 0.008 | 0.008 | 0.008 |
| 0.7 | EMP | 0.014 | 0.014 | 0.014 | 0.014 | 0.014 | 0.014 | 0.017 | 0.017 | 0.017 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.012 | 0.012 | 0.012 |
| | IPW | 0.014 | 0.014 | 0.014 | 0.015 | 0.015 | 0.015 | 0.019 | 0.019 | 0.018 | 0.010 | 0.010 | 0.010 | 0.011 | 0.011 | 0.011 | 0.013 | 0.013 | 0.013 |
| | CC | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 |
| | UN | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.007 | 0.007 | 0.007 | 0.007 | 0.007 | 0.007 | 0.007 | 0.007 | 0.007 |
| 0.8 | EMP | 0.011 | 0.011 | 0.011 | 0.012 | 0.012 | 0.012 | 0.013 | 0.013 | 0.014 | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 | 0.009 | 0.009 | 0.009 |
| | IPW | 0.011 | 0.011 | 0.011 | 0.012 | 0.012 | 0.012 | 0.015 | 0.015 | 0.014 | 0.008 | 0.008 | 0.008 | 0.009 | 0.009 | 0.008 | 0.011 | 0.011 | 0.010 |
| | CC | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.007 | 0.007 | 0.007 | 0.007 | 0.007 | 0.007 | 0.007 | 0.007 | 0.007 |
| | UN | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 | 0.006 | 0.006 | 0.006 | 0.006 | 0.006 | 0.006 | 0.005 | 0.005 | 0.005 |
| 0.9 | EMP | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 | 0.009 | 0.009 | 0.009 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.006 | 0.006 | 0.006 | 0.006 |
| | IPW | 0.007 | 0.007 | 0.007 | 0.008 | 0.008 | 0.008 | 0.010 | 0.010 | 0.010 | 0.005 | 0.005 | 0.005 | 0.006 | 0.006 | 0.006 | 0.007 | 0.007 | 0.007 |
| | CC | 0.007 | 0.007 | 0.007 | 0.007 | 0.007 | 0.007 | 0.007 | 0.007 | 0.007 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 |
| | UN | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.004 | 0.004 | 0.004 | 0.004 | 0.004 | 0.004 | 0.004 | 0.004 | 0.004 |

Table 4.6.5: Coverage Probability (in %) of the IPW, CC, and UN estimators for the super-population ROC curve under SSRS.

| Missing | FPR | Sampling fraction: 5% | | | | | Sampling fraction: 10% | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | IPW-M1 | IPW-M2 | IPW-M3 | CC | UN | IPW-M1 | IPW-M2 | IPW-M3 | CC | UN |
| 10% | 0.1 | 95.8 | 95.9 | 94.6 | 90.7 | 0.1 | 94.4 | 94.5 | 92.6 | 85.9 | 0.0 |
| | 0.2 | 95.2 | 95.2 | 94.6 | 90.0 | 0.0 | 94.8 | 94.8 | 93.5 | 86.5 | 0.0 |
| | 0.3 | 95.1 | 95.1 | 94.8 | 91.3 | 0.1 | 93.9 | 93.9 | 93.3 | 86.1 | 0.0 |
| | 0.4 | 94.3 | 94.2 | 93.9 | 89.8 | 0.4 | 94.8 | 94.8 | 92.5 | 86.9 | 0.0 |
| | 0.5 | 95.3 | 95.2 | 94.3 | 90.9 | 0.6 | 95.9 | 95.9 | 94.4 | 88.3 | 0.0 |
| | 0.6 | 94.9 | 94.9 | 93.0 | 87.7 | 0.9 | 95.6 | 95.7 | 93.5 | 88.4 | 0.0 |
| | 0.7 | 95.6 | 95.5 | 94.9 | 91.2 | 5.9 | 95.2 | 95.2 | 93.5 | 89.1 | 0.3 |
| | 0.8 | 93.7 | 93.7 | 94.3 | 90.7 | 16.0 | 94.1 | 94.2 | 92.4 | 87.2 | 1.5 |
| | 0.9 | 93.9 | 93.9 | 93.4 | 90.6 | 40.8 | 95.1 | 95.1 | 93.7 | 90.1 | 13.4 |
| 25% | 0.1 | 95.8 | 95.7 | 92.1 | 75.4 | 0.0 | 93.1 | 92.9 | 88.3 | 60.7 | 0.0 |
| | 0.2 | 94.8 | 94.8 | 91.1 | 77.4 | 0.0 | 95.0 | 95.1 | 87.6 | 61.3 | 0.0 |
| | 0.3 | 96.3 | 96.4 | 91.8 | 76.7 | 0.1 | 94.9 | 95.0 | 87.2 | 60.2 | 0.0 |
| | 0.4 | 95.6 | 95.5 | 91.7 | 76.2 | 0.1 | 95.3 | 95.4 | 88.3 | 64.7 | 0.0 |
| | 0.5 | 95.6 | 95.6 | 92.3 | 77.0 | 0.1 | 95.2 | 95.2 | 89.0 | 65.9 | 0.0 |
| | 0.6 | 95.3 | 95.4 | 90.8 | 74.9 | 0.5 | 95.2 | 95.2 | 89.1 | 65.3 | 0.0 |
| | 0.7 | 96.2 | 96.2 | 93.0 | 77.9 | 2.6 | 95.4 | 95.5 | 91.5 | 73.4 | 0.1 |
| | 0.8 | 95.7 | 95.9 | 92.1 | 79.4 | 8.4 | 94.1 | 94.1 | 90.2 | 73.5 | 0.6 |
| | 0.9 | 96.0 | 95.9 | 93.5 | 83.2 | 30.2 | 96.4 | 96.4 | 92.3 | 81.8 | 8.4 |
| 50% | 0.1 | 91.6 | 91.5 | 90.7 | 62.1 | 0.2 | 93.1 | 93.2 | 85.2 | 46.7 | 0.0 |
| | 0.2 | 95.8 | 95.5 | 91.8 | 61.4 | 0.1 | 93.6 | 93.9 | 86.4 | 40.5 | 0.0 |
| | 0.3 | 95.2 | 95.3 | 92.1 | 61.6 | 0.2 | 94.4 | 94.2 | 85.3 | 41.9 | 0.0 |
| | 0.4 | 95.8 | 95.8 | 91.8 | 60.4 | 0.1 | 95.2 | 95.1 | 84.4 | 41.6 | 0.0 |
| | 0.5 | 96.9 | 96.8 | 91.1 | 59.2 | 0.5 | 97.0 | 96.9 | 87.7 | 44.6 | 0.0 |
| | 0.6 | 96.4 | 96.5 | 89.7 | 59.4 | 0.7 | 96.3 | 96.2 | 85.6 | 43.7 | 0.0 |
| | 0.7 | 97.3 | 97.3 | 91.8 | 64.3 | 1.7 | 97.0 | 96.8 | 89.3 | 52.4 | 0.1 |
| | 0.8 | 97.8 | 97.8 | 92.8 | 68.4 | 6.1 | 96.7 | 96.8 | 90.0 | 57.0 | 0.1 |
| | 0.9 | 98.0 | 98.2 | 93.4 | 72.7 | 25.5 | 98.6 | 98.6 | 92.4 | 68.4 | 5.5 |

Table 4.6.6: Coverage Probability (in %) of the IPW, CC, and UN estimators for the super-population ROC curve under STSCS.

| Missing | FPR | Sampling fraction: 5% | | | | | Sampling fraction: 10% | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | IPW-M1 | IPW-M2 | IPW-M3 | CC | UN | IPW-M1 | IPW-M2 | IPW-M3 | CC | UN |
| 10% | 0.1 | 95.0 | 95.1 | 94.4 | 91.9 | 1.3 | 95.1 | 95.0 | 93.3 | 88.7 | 0.0 |
| | 0.2 | 94.3 | 94.3 | 94.7 | 91.3 | 0.3 | 95.7 | 95.8 | 93.4 | 88.3 | 0.0 |
| | 0.3 | 93.5 | 93.7 | 94.4 | 92.2 | 0.6 | 95.0 | 95.1 | 93.8 | 88.8 | 0.0 |
| | 0.4 | 95.2 | 95.2 | 95.0 | 92.2 | 1.2 | 94.9 | 94.8 | 94.3 | 88.8 | 0.0 |
| | 0.5 | 95.1 | 95.3 | 95.4 | 91.8 | 2.3 | 95.3 | 95.3 | 94.0 | 88.7 | 0.0 |
| | 0.6 | 95.6 | 95.6 | 94.9 | 91.2 | 3.3 | 94.6 | 94.6 | 93.2 | 89.1 | 0.0 |
| | 0.7 | 94.8 | 94.8 | 93.6 | 91.5 | 9.1 | 94.8 | 94.9 | 94.3 | 90.7 | 0.6 |
| | 0.8 | 94.6 | 94.6 | 93.6 | 91.4 | 21.0 | 94.9 | 94.9 | 95.0 | 91.6 | 4.2 |
| | 0.9 | 93.9 | 94.0 | 92.7 | 89.5 | 44.9 | 94.7 | 94.7 | 93.5 | 91.7 | 21.6 |
| 25% | 0.1 | 93.5 | 93.7 | 92.6 | 82.2 | 0.4 | 93.6 | 93.7 | 89.3 | 72.2 | 0.0 |
| | 0.2 | 94.2 | 94.2 | 91.3 | 78.9 | 0.4 | 95.7 | 95.7 | 89.8 | 69.9 | 0.0 |
| | 0.3 | 93.4 | 93.6 | 92.4 | 81.7 | 0.4 | 96.1 | 96.1 | 90.5 | 72.5 | 0.0 |
| | 0.4 | 95.1 | 95.5 | 93.5 | 81.7 | 0.9 | 96.0 | 96.0 | 91.4 | 71.4 | 0.0 |
| | 0.5 | 95.2 | 95.2 | 92.8 | 80.9 | 0.8 | 96.7 | 96.7 | 90.7 | 71.1 | 0.0 |
| | 0.6 | 95.6 | 95.4 | 92.2 | 81.5 | 1.6 | 95.9 | 95.9 | 90.7 | 73.5 | 0.0 |
| | 0.7 | 96.1 | 96.1 | 93.1 | 83.7 | 5.1 | 96.0 | 95.9 | 91.6 | 76.9 | 0.0 |
| | 0.8 | 96.0 | 96.0 | 93.3 | 85.6 | 13.8 | 96.2 | 96.1 | 93.2 | 81.7 | 1.7 |
| | 0.9 | 95.7 | 95.6 | 92.6 | 85.9 | 36.2 | 96.1 | 96.1 | 94.7 | 83.9 | 15.4 |
| 50% | 0.1 | 93.6 | 93.8 | 93.0 | 68.4 | 0.6 | 92.8 | 92.8 | 89.4 | 58.5 | 0.0 |
| | 0.2 | 94.7 | 94.3 | 91.6 | 67.8 | 0.0 | 93.5 | 93.6 | 89.4 | 53.6 | 0.0 |
| | 0.3 | 95.9 | 95.8 | 92.9 | 67.0 | 0.2 | 94.9 | 95.2 | 90.9 | 52.5 | 0.0 |
| | 0.4 | 96.8 | 97.0 | 92.9 | 67.8 | 0.6 | 95.6 | 95.6 | 90.4 | 52.8 | 0.0 |
| | 0.5 | 96.8 | 96.9 | 92.7 | 66.6 | 0.8 | 96.7 | 96.5 | 89.5 | 53.1 | 0.0 |
| | 0.6 | 96.7 | 96.6 | 91.5 | 67.1 | 2.4 | 96.6 | 96.5 | 89.4 | 53.6 | 0.0 |
| | 0.7 | 96.6 | 96.5 | 92.4 | 70.2 | 4.9 | 98.0 | 98.0 | 89.9 | 59.6 | 0.1 |
| | 0.8 | 96.8 | 96.8 | 93.3 | 73.0 | 13.3 | 98.1 | 98.1 | 93.8 | 66.6 | 1.8 |
| | 0.9 | 95.5 | 95.4 | 92.3 | 76.6 | 31.5 | 97.6 | 97.7 | 93.2 | 73.8 | 11.9 |

Table 4.6.7: Descriptive Statistics of adults aged 20 years or more, excluding pregnant women, that had FPG results in NHANES 1999-2006.

| Characteristic | Missing (%) | Overall N = 17,696 | Normal N = 11,798 | Pre-diabetes N = 5,328 | Undiagnosed diabetes N = 570 |
|---|---|---|---|---|---|
| **Age** | 0 (0%) | | | | |
| < 40 | | 3,359 (19%) | 2,205 (19%) | 1,072 (20%) | 82 (14%) |
| 40-49 | | 2,327 (13%) | 1,306 (11%) | 926 (17%) | 95 (17%) |
| 50-59 | | 4,637 (26%) | 2,242 (19%) | 2,045 (38%) | 350 (61%) |
| $\geq 60$ | | 9,317 (53%) | 5,573 (47%) | 3,392 (64%) | 352 (62%) |
| **FHX of diabetes** | 389 (2.2%) | 8,006 (46%) | 5,176 (45%) | 2,499 (48%) | 331 (60%) |
| **HX of hypertension** | 9 (< 0.1%) | 5,960 (34%) | 3,174 (27%) | 2,398 (45%) | 388 (68%) |
| **Obesity** | 0 (0%) | | | | |
| Extremely Obese | | 861 (4.9%) | 382 (3.2%) | 396 (7.4%) | 83 (15%) |
| Obese | | 7,123 (40%) | 4,238 (36%) | 2,532 (48%) | 353 (62%) |
| Overweight | | 5,030 (28%) | 3,443 (29%) | 1,500 (28%) | 87 (15%) |
| Not overweight or obese | | 4,489 (25%) | 3,604 (31%) | 848 (16%) | 37 (6.5%) |
| **Physical Activity** | 10,348 (58%) | 2,852 (39%) | 1,798 (38%) | 947 (40%) | 107 (34%) |
| **Risk of undiagnosed diabetes (%)** | 10,511 (59%) | | | | |
| Mean (SD) | | 88 (13) | 85 (14) | 93 (8) | 97 (5) |
| Median (IQR) | | 94 (83, 98) | 90 (77, 97) | 97 (92, 98) | 98 (97, 99) |
| Range | | 42, 100 | 42, 100 | 42, 100 | 42, 100 |
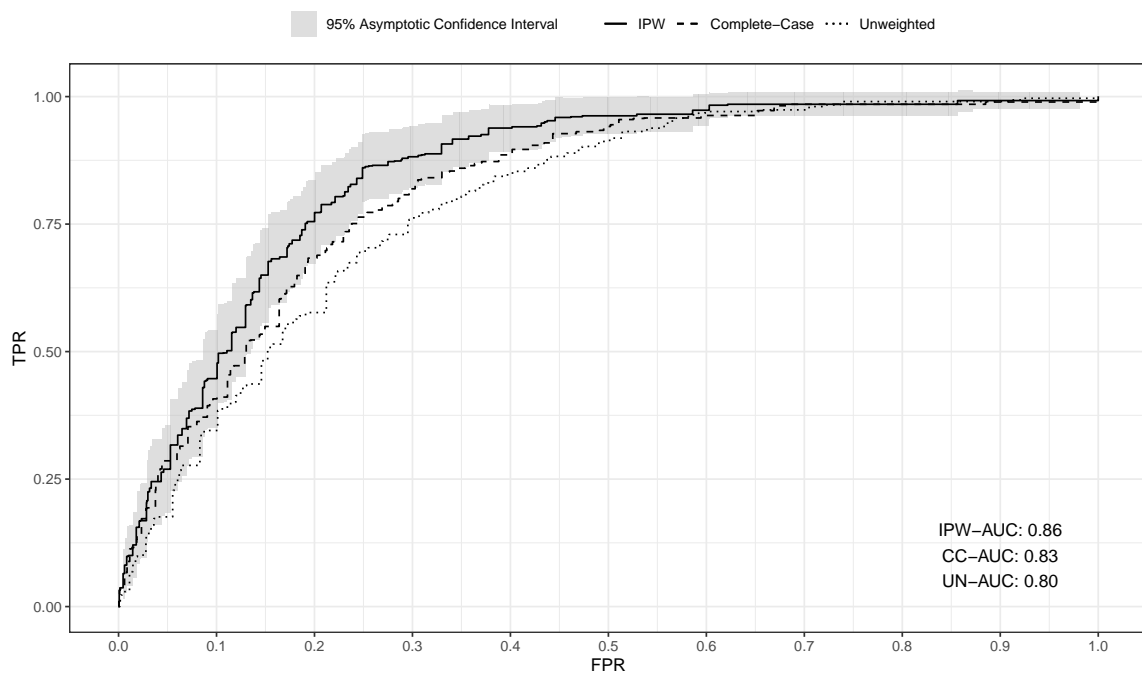
Figure 4.6.4: Estimated ROC curves using IPW, complete-case and unweighted method for the NHANES target population.

# CHAPTER 5: COVARIATE-SPECIFIC RECEIVER OPERATING CHARACTERISTIC CURVE FOR COMPLEX SURVEY DATA

## 5.1   Introduction

The receiver operating characteristic (ROC) curve is the most popular method to assess the performance of a continuous diagnostic test. The curve is defined as the plot of the false positive rate (1-specificity) versus the true-positive rate (sensitivity) across all possible cutoffs of the diagnostic test. The false-positive rate (FPR) is the proportion of non-diseased individuals that test positive for the disease based on the diagnostic test, and the true-positive rate (TPR) is the proportion of diseased individuals that test positive for the disease.

In many applications, however, the discriminating capacity of a diagnostic test may be affected by various factors. For example, disease severity may impact the marker accuracy, with less severe cases being more difficult to distinguish from controls. Currently, there are three major existing approaches to evaluate the covariate effects on the ROC curve (Pepe, 1998). In the first class of approaches, called induced methods, the distributions of the diagnostic test in the diseased and non-diseased populations are modeled separately, from which the ROC curve is computed. The second class considers regression models for the area under the ROC curve (AUC). Lastly, in the third approach, called direct methods, the covariate effects on the ROC curve are modeled directly. The latter class is often referred to as parametric distribution-free (PDF) models, since a parametric model is assumed for the ROC curve, but the distributions of the diagnostic tests remaing unspecified.

The PDF approach was originally proposed by Pepe (1997), and various estimation proposals have been made over the years, including Pepe (2000a); Alonzo and Pepe (2002); Cai and Pepe (2002); Cai (2004). Due to the similarities with the generalized linear models, this class of models is also referred to as ROC-GLM models. In our third project, we adapt the ROC-GLM model presented in Pepe (2000a) and Alonzo and Pepe (2002) to account for complex survey designs. The properties of the adapted model is evaluated using simulation studies, and the model is then applied to the National Health and Nutrition Examination Survey (NHANES) to evaluate the effect of age on the accuracy of a diabetes assessment tool.

## 5.2  ROC-GLM model for simple random samples

Let $\{(Y_i, D_i, \boldsymbol{X}_i)\}_{i=1}^n$ be i.i.d realizations of the diagnostic test measure $Y$, the disease indicator $D$, and the set of covariates $\boldsymbol{X}$. We denote the cumulative distribution function (cdf) of $Y$ conditioned on $D = 0$ as $G$, and similarly, the cdf of $Y$ conditioned on $D = 1$ as $F$. We assume that $F$ and $G$ have continuous probability density functions (pdf) $f$ and $g$, respectively. The ROC curve is defined as the plot of $\{(1 - G(c), 1 - F(c)) : c \in \mathbb{R}\}$, or equivalently, as the plot of $\{(s, R(s)) : s \in [0, 1]\}$, where $R(s) = 1 - F \circ G^{-1}(1 - s)$, with $G^{-1}(s) = \inf\{x \in \mathbb{R} : G(x) \geq s\}$, and $F \circ G^{-1}(.) \equiv F(G^{-1}(.))$. The area under the ROC curve (AUC-ROC) is $A = \int_0^1 R(s) \ ds$. One natural extension for the ROC curve to accommodate covariates is defined as

$$R(s|\boldsymbol{x}) = 1 - F\{G^{-1}(1 - s|\boldsymbol{x})|\boldsymbol{x}\}, \quad 0 \leq s \leq 1,$$

with $G(c|\boldsymbol{x}) = P(Y \leq c|D = 0, \boldsymbol{X} = \boldsymbol{x}) \equiv P(Y_{\bar{D}} \leq c|\boldsymbol{X} = \boldsymbol{x})$ and $F(c|\boldsymbol{x}) = P(Y \leq c|D = 1, \boldsymbol{X} = \boldsymbol{x}) \equiv P(Y_D \leq c|\boldsymbol{X} = \boldsymbol{x})$. The corresponding covariate-specific AUC-ROC is given by

$$A(\boldsymbol{x}) = \int_0^1 R(s|\boldsymbol{x}) \ ds.$$

The general form for the ROC-GLM regression model has the following expression:

$$g\{R(t|\boldsymbol{X})\} = h(t) + \mu(\boldsymbol{X}), \qquad t \in (0,1) \tag{5.20}$$

where $\mu(\boldsymbol{X})$ models the effects of covariate $\boldsymbol{X}$ on the ROC curve, $h(t)$ is an unknown monotonic increasing function of the FPR related to the shape of the ROC curve, and $g$ is the link function. Model (5.20) is also referred to as parametric distribution-free (PDF), which assumes a parametric model for the ROC curve, but it is distribution-free for the diagnostic test results.

Different model proposals were made by varying $g$, $h$ and $\mu$. Let $\boldsymbol{X} = (X, X_D)$, with $X$ representing the covariates that are common to diseased and nondiseased subjects, and $X_D$ representing the covariates specific to the diseased subjects. Pepe (2000a) and Alonzo and Pepe (2002) assumed the following general model:

$$g\{R(t|\boldsymbol{X})\} = \sum_{k=1}^{K} \gamma_k h_k(t) + X'\beta + X_D'\beta_D, \tag{5.21}$$

with the common model specifications given by the binormal model

$$\Phi^{-1}\{R(t|\boldsymbol{X})\} = \gamma_1 + \gamma_2 \Phi^{-1}(t) + X'\beta + X_D'\beta_D,$$

and the bilogit model:

$$\text{logit}\{R(t|\boldsymbol{X})\} = \gamma_1 + \gamma_2 \text{logit}(t) + X'\beta + X_D'\beta_D,$$

where $\text{logit}(t) = \log(t/(1-t))$. Pepe (2000a) noted that generalized linear methods for binary data could be used for estimation for model (5.21). Alonzo and Pepe (2002) simplified further the computational aspects for estimation for fitting the PDF model,

based on the fact that

$$R(s|\boldsymbol{x}) = 1 - F\{G^{-1}(1 - s|\boldsymbol{x})|\boldsymbol{x}\}$$

$$= P\{Y_D \geq G^{-1}(1 - s|\boldsymbol{x})|\boldsymbol{x}\}$$

$$= \mathrm{E}\left\{I(Y_D \geq G^{-1}(1 - s|\boldsymbol{x}))|\boldsymbol{x}\right\}$$

$$= g^{-1}\left\{\sum_{k=1}^{K} \gamma_k h_k(t) + X'\beta + X'_D\beta_D\right\},$$

so that the procedure for fitting the model (5.21) is based on the (correlated) binary variable $\{U_{it}, i = 1, \cdots, n_D; t \in T\}$ where $U_{it} = I(Y_{Di} \geq G^{-1}(1 - t|\boldsymbol{x}_i))$, $n_D$ is the number of diseased subjects in the sample, and $T \subset (0, 1)$ is a set of FPR. The algorithm to estimate $\theta = (\gamma_1, \cdots, \gamma_k, \beta, \beta_D)$ is as follows:

1. Specify a set $T \subset (0, 1)$ of FPR;

2. For each $t \in T$ estimate $G^{-1}(1 - t|x)$, the $(1 - t)$th quantile of the distribution function for the nondiseased subjects, conditional on $X = x$;

3. Caculate $U_{it} = I\left[Y_{Di} \geq \hat{G}^{-1}(1 - t|x_i)\right]$, $i = 1, \cdots, n_D$ and $t \in T$;

4. Fit the model $g\{E(U_{it})\} = \sum_{k=1}^{K} \gamma_k h_k(t) + X'\beta + X'_D\beta_D$ by solving standard estimating equations for fitting binary generalized linear model to $U_{it}$ with link function $g$ and covariates $\{h_k(t), X_i, X_{Di}; k = 1, \cdots, K\}$.

The estimators $\hat{\boldsymbol{\gamma}} = (\hat{\gamma}_1, \cdots, \hat{\gamma}_K)$ and $\hat{\boldsymbol{\beta}} = (\hat{\beta}, \hat{\beta}_D)$ are the solution of the following estimating equations:

$$\sum_{i=1}^{n_D} \sum_{t \in T} S_i(\boldsymbol{\gamma}, \boldsymbol{\beta}, t) = 0$$

with

$$S_i(\boldsymbol{\gamma}, \boldsymbol{\beta}, t) = (h(t), X_i, X_{Di})^T w(\mu_{\boldsymbol{\gamma}, \boldsymbol{\beta}}(t)) \{U_{it} - \mu_{\boldsymbol{\gamma}, \boldsymbol{\beta}}(t)\},$$

and

$$w(s) = \{g'(s)s(1-s)\}^{-1}$$

$$\mu_{\boldsymbol{\gamma},\boldsymbol{\beta}}(t) = g^{-1}\left\{\sum_{k=1}^{K}\gamma_k h_k(t) + X^T\beta + X_D^T\beta_D\right\}.$$

The simplified framework and lower computational burden of the approach in Alonzo and Pepe (2002) make this the preferred method in modern statistical software for directly modeling covariate effects on ROC curves. Although inference based on asymptotic distribution theory is desirable, the induced correlation amongst the estimating equation's components makes this task not completely straightforward (Pepe, 1997). Pepe (2000a) developed the asymptotic distribution theory for the ROC-GLM models, but the estimation based on $n_D \times n_{\bar{D}}$ observations makes this option not practical for larger datasets, as pointed in Alonzo and Pepe (2002). Consequentially, bootstrap resampling method remains the preferred method for inference on Pepe (1997) and Alonzo and Pepe (2002).

## 5.3    ROC-GLM model for complex survey data

Consider a finite population $\mathcal{U}^N$ of size $N$, with corresponding set of indices $U = \{1, \cdots, N\}$. Each index $i \in U$ is associated with a unique vector $(y_i, w_i, z_i) \in \mathbb{R}^p \times \mathbb{R}^k \times \mathbb{R}_+^q$ representing, respectively, the characteristics of interest, the complete auxiliary information, and the sampling design information available at the time of the design of the survey on all units. We assume that $\{(y_i, w_i, z_i)\}_{i=1}^N$ are realizations of random variables $(Y, W, Z)$ defined according to a superpopulation model. We denote $\boldsymbol{y}^N = (y_1, \cdots, y_N)$, $\boldsymbol{Y}^N = (Y_1, \cdots, Y_N)$, $\boldsymbol{w}^N = (w_1, \cdots, w_N)$, and $\boldsymbol{W}^N = (W_1, \cdots, W_N)$, $\boldsymbol{z}^N = (z_1, \cdots, z_N)$, and $\boldsymbol{Z}^N = (Z_1, \cdots, Z_N)$.

For our purposes, let $\{(Y_i, W_i, Z_i) = (Y_i, D_i, X_i, X_{D_i}, Z_i)\}_{i=1}^N$ be i.i.d realizations of the diagnostic test measure $Y$, the disease indicator $D$, the covariates that are common

to diseased and nondiseased units $X$, the covariates specific to the diseased units $X_D$, and the sampling design information $Z$ available for all units in the sampling stage.

Suppose a sample $s$ of size $n$ is drawn from the finite population according to a probability sampling design $p(s) = p(s, \boldsymbol{z}^N)$. Note that since $p$ does not depend on $\boldsymbol{y}^N$, only non-informative sampling designs are being considered. Let $\xi_i = I(i \in s)$ be the sample indicators, $\pi_i = \mathbb{E}[\xi_i | \boldsymbol{Z}^N]$ the first-order inclusion probability, and $\pi_{ij} = \mathbb{E}[\xi_i \xi_j | \boldsymbol{Z}^N]$ the second-order inclusion probability.

In this paper, we propose to adapt the ROC-GLM model from Alonzo and Pepe (2002), to account for the complex-survey design:

$$g\left\{R(t|\boldsymbol{X})\right\} = \sum_{k=1}^{K} \gamma_k^\pi h_k(t) + X'\beta^\pi + X_D'\beta_D^\pi \tag{5.22}$$

The algorithm proposed by Alonzo and Pepe (2002) can be easily adapted for complex survey data since quantile regression and generalized linear models accounting for complex survey design are widely available in statistical software capable of handling survey data. The adapted algorithm is as follows:

1. Specify a set $T \subset (0, 1)$ of FPR;

2. For each $t \in T$ estimate $G^{-1}(1 - t|x)$, the $(1 - t)$th quantile of the distribution function for the nondiseased subjects, conditional on $X = x$, accounting for the survey design;

3. Caculate $U_{it}^\pi = I\left[Y_{Di} \geq \hat{G}^{-1}(1 - t|x_i)\right]$, $i = 1, \cdots, n_D$ and $t \in T$;

4. Fit the model $g\left\{E(U_{it}^\pi)\right\} = \sum_{k=1}^{K} \gamma_k^\pi h_k(t) + X'\beta^\pi + X_D'\beta_D^\pi$ by solving design-adjusted estimating equations for fitting binary generalized linear model to $U_{it}^\pi$ with link function $g$ and covariates $\{h_k(t), X_i, X_{Di}; k = 1, \cdots, K\}$

The estimators $\hat{\boldsymbol{\gamma}}^\pi = (\hat{\gamma}_1^\pi, \cdots, \hat{\gamma}_K^\pi)$ and $\hat{\boldsymbol{\beta}}^\pi = (\hat{\beta}^\pi, \hat{\beta}_D^\pi)$ are the solution of the design-adjusted estimating equations (Binder, 1983):

$$\sum_{i=1}^{N} \sum_{t \in T} D_i S_i^\pi(\boldsymbol{\gamma}^\pi, \boldsymbol{\beta}^\pi, t) = 0$$

with

$$S_i^\pi(\boldsymbol{\gamma}^\pi, \boldsymbol{\beta}^\pi, t) = \xi_i \pi_i^{-1} (h(t), X_i, X_{Di})^T w(\mu_{\boldsymbol{\gamma}^\pi, \boldsymbol{\beta}^\pi}(t)) \left\{ U_{it}^\pi - \mu_{\boldsymbol{\gamma}^\pi, \boldsymbol{\beta}^\pi}(t) \right\}. \tag{5.23}$$

As commonly done for data arising from simple random sample, variance estimation will be performed using resampling techniques suitable for complex designs. In survey sampling, those methods are usually referred as replication weights methods. Commonly implemented replicate weights methods in modern survey software are balanced repeated replication (BRR), jackknife, and bootstrap. The later being the most flexible of the methods, suitable for most sampling designs, it is the method of choice for this paper.

The bootstrap method, originally proposed by Efron (1992) for independent and identically distributed (i.i.d) data, relies on recomputing the estimate $\hat{\theta}$ a large number of times by resampling from the original sample. Since survey data are not necessarily i.i.d., many bootstrap resampling methods have been proposed in the context of survey data. Mashreghi et al. (2016) present a comprehensive review of bootstrap resampling methods, classifying them into three groups. The first one is the class of pseudo-population bootstrap methods in which a pseudo-population is first created by repeating the units of the original sample and bootstrap samples are then selected from the pseudo-population (Gross, 1980; Booth et al., 1994). The second one, called the direct bootstrap methods, consists of directly selecting bootstrap samples from the original sample or a rescaled version of it (Rao and Wu, 1984, 1988; Sitter, 1992; Canty and Davison, 1999). In the third group, called the bootstrap weights methods, an appropriate adjustment is made on the original survey weights to obtain a new set of weights called the bootstrap weights

69

(Rao et al., 1992; Beaumont and Patak, 2012). This method is attractive to users to public data files prepared by statistical agencies, which provide data sets consisting of columns with the original observations, a column with the original survey weights and $B$ columns of bootstrap weights.

For this paper, we use the function `as.svrepdesign` from the `survey` package in R, where the bootstrap method proposed in Canty and Davison (1999) is implemented and described as follows. Suppose $y_1, \cdots, y_n$ is a sample of size $n$ drawn from a finite population of size $N$, and for simplicity, suppose that the inverse of the sampling fraction $1/\lambda$ is an integer. Let $\theta$ and $\hat{\theta}$ be the parameter of interest and the parameter estimate based on the sample, respectively.

1. A bootstrap population of size $N$ is constructed by concatenating $1/\lambda$ copies of $y_1, \cdots, y_n$;

2. A bootstrap sample of size $n$ is selected without replacement. In the case of stratified population, steps 1 and 2 are applied to each stratum separately;

3. The selected bootstrap sample is reweighted through calibration so that its marginal totals match those of the original population;

4. The bootstrap parameter estimate $\hat{\theta}^*$ is computed from the bootstrap sample using the reweighted inclusion probabilities $\pi_i^*$, $i = 1, \cdots, n$;

5. Steps 1-4 are repeated $B$ times to obtain $\hat{\theta}_1^*, \cdots, \hat{\theta}_R^*$, and these are used to estimate the variance of $\hat{\theta}$ by

$$\widehat{\text{Var}}(\hat{\theta}) = \frac{1}{B-1} \sum_{b=1}^{B} (\hat{\theta}_b^* - \bar{\hat{\theta}}^*)^2,$$

with $\bar{\hat{\theta}}^* = B^{-1} \sum_{b=1}^{B} \hat{\theta}_b^*$.

## 5.4 Simulation studies

In this simulation study, we investigate the performance of the proposed survey-weighted ROC-GLM under stratified simple random sampling (SSRS) and stratified two-stage cluster sampling (STSCS). For each sampling scheme, a total of 4 scenarios are considered according to different disease proportions $p = 5\%$, $15\%$ and sampling fractions $\lambda = 5\%$, $10\%$.

We generated populations of size $N = 100,000$ subdivided in five strata containing $5\%$, $10\%$, $25\%$, $30\%$ and $30\%$ of the observations. For each stratum $h$, we generated $X_h = \alpha_0 + \alpha_h D + \beta_h D \times W + \epsilon$, where $W \sim N(0,1)$, $D \sim \text{Ber}(p)$ and $\epsilon \sim N(0,1)$. The induced ROC curve in each stratum $h$ is given by the binormal model $R(s) = \Phi(\alpha_h + \beta_h W + \Phi^{-1}(s))$, where $\Phi(.)$ is the standard normal cdf. We set $\alpha_0 = 1$, $\alpha_1 = 2$, $\alpha_2 = 1.75$, $\alpha_3 = 1.5$, $\alpha_4 = 1.25$, $\alpha_5 = 1$, $\beta_1 = 0.5$, $\beta_2 = 1$, $\beta_3 = 1.5$, $\beta_4 = 2$, and $\beta_5 = 2.5$ such that the expected AUC for each strata were 0.9, 0.85, 0.75, 0.7, 0.6 respectively. For STSCS, 10,000 clusters of sizes 5, 10 and 15 were generated using quantiles of $X_h + \tau$, $\tau \sim N(0,1)$, in addition to the steps already described.

Samples with size determined by the sampling fraction were drawn assuming uniform allocation. For the first stage of STSCS, $m = 125$ and 250 clusters were sampled from each stratum, and $80\%$ of the observations were sampled from each cluster at the second stage. For each sample, we fitted the ROC-GLM model

$$R_W(s) = \Phi(\alpha_0 + \alpha_1 \Phi^{-1}(s) + \beta W),$$

using our proposed design-adjusted method and the unadjusted method proposed by Alonzo and Pepe (2002).

We evaluated the performance of the proposed method in terms of the model parameters $\alpha_0$, $\alpha_1$ and $\beta$, and the estimated covariate-specific ROC curve with two scenarios $W = -0.67$ and $W = 0$. The estimates are compared with the finite population pa-

rameters, with the results presented in terms of relative bias (RB), empirical standard error (ESE), bootstrap standard error (BSE), and coverage probability (CP) for the 95% bootstrap confidence intervals. We compare our method (SVYW) to the unadjusted method (UN), where the sampling weights are ignored. The results are based on 1,000 simulation samples, and 250 bootstrap replicates drawn from each simulation sample to compute the bootstrap standard error.

### 5.4.1 Results for the model parameters

The simulation results for the model parameters are reported in Table 5.7.1, and Figures 5.7.1, 5.7.2, and 5.7.3. As expected, the relative bias for the ROC-GLM model without accounting for the sampling design leads to biased estimates, especially for the covariate effect parameter $\beta$, with absolute relative biases reaching more than 20%. In contrast, the relative bias from the adapted ROC-GLM model does not exceed 1.6%, with higher biases for the case with a smaller sample size and disease proportion.

In terms of standard errors, we see that the standard errors obtained using bootstrap are very close to the empirical standard error. As expected, the bootstrap standard error that does not account for the sampling design is underestimated. Finally, in terms of coverage probability, our design-adjusted method has coverage probabilities close to 95%. In contrast, the unadjusted model has poor coverage due to the significant bias and underestimated variances observed previously.

### 5.4.2 Results for the ROC curve

The results for the relative biases are reported in Table 5.7.2 and Figure 5.7.4. As expected, the relative bias for the ROC curve from the unadjusted method is quite large, especially at the beginning of the curve, with relative biases close to 70% and a decreasing trend along the curve. In contrast, the values for the SW and WT estimators never exceed 1%.

In terms of standard errors of the ROC curve presented in Tables 5.7.3, 5.7.4, and Figure 5.7.5, we see that the standard error obtained without accounting for the survey design is far from the empirical standard error. When the survey design is considered, the bootstrap standard error is very close to the empirical standard error, with mild departures for smaller sample sizes and disease proportions.

Finally, Table 5.7.5 and Figure 5.7.6 report the coverage probability of the 95% bootstrap confidence interval for both approaches. As expected, not taking the sampling design into account leads to inferior performance, while our adapted ROC-GLM model presents coverage probabilities close to 95%.

## 5.5  Application

Diabetes and its complications are major causes of morbidity and mortality worldwide. Currently, clinical practice guidelines recommend screening for pre-diabetes and type 2 diabetes with an informal assessment of risk factors or validated risk calculator in asymptomatic adults to guide providers on whether performing a definitive diagnostic test is necessary (Draznin et al., 2022). The current risk assessment tool used by the American Diabetes Association (ADA) to screen for pre-diabetes and type 2 diabetes is adapted from the algorithm developed in Bang et al. (2009) to estimate the risk of undiagnosed diabetes.

According to the Centers for Disease Control and Prevention's (CDC) 2020 National Diabetes Statistics Report (https://www.cdc.gov/diabetes/data/statistics-report/index.html - Accessed 2022-06-02), approximately 20% of the diabetes cases in adults aged 65 years or more are undiagnosed. Older individuals with diabetes have higher rates of premature death, functional disability, and coexisting illnesses such as hypertension, coronary heart disease, and stroke than those without diabetes. Older adults with diabetes are also at greater risk than other older adults for several common geriatric syndromes, such as

polypharmacy, depression, cognitive impairment, urinary incontinence, injurious falls, and persistent pain (American Diabetes Association, 2011).

According to the ADA, older adults are defined as those aged 65 years or more. In this application, we compare the discrimination of the algorithm developed by Bang et al. (2009) between working-age adults and older adults. We use data from the National Health and Nutrition Examination Survey (NHANES), which is an annual survey conducted by the CDC's National Center for Health Statistics (NCHS) that utilizes a complex, multistage probability sampling design to select a representative sample of the non-institutionalized resident population of the United States, between 1999-2006.

Similarly as presented in Bang et al. (2009), we consider participants aged 20 years or more, excluding pregnant women, that had available fasting plasma glucose (FPG) results. The participants are classified into four groups of diabetes status: known diabetes (if answered "yes" to the question "Other than during pregnancy, have you ever been told by a doctor or health professional that you have diabetes or sugar diabetes?"), normal glucose metabolism (FPG < 100 mg/dL), pre-diabetes (FPG 100-125 mg/dL), and undiagnosed diabetes (FPG > 125 mg/dL). The participants classified as "known diabetes" were not considered in the analysis, and the undiagnosed diabetes was used as the binary outcome. The diabetes risk score was computed using age (< 40, 40-49, 50-59, > 59), sex (female, male), family history of diabetes (yes, no), history of hypertension (yes, no), obesity (not overweight, overweight, obese, extremely obese), physically active (yes, no), according to Bang et al. (2009).

In the 1999-2006 NHANES, 20,159 non-pregnant adults aged 20 years or more were enrolled. Out of this sample, 17,696 observations were classified as either normal glucose metabolism, pre-diabetes, and undiagnosed diabetes, and 7,185 observations had information for all variables needed to compute the risk score. In this final analytic sample, 1,761 observations were classified as older adults, with 7.5% (95% CI: 5.7%, 9.4%) of undiagnosed diabetes in this group.

For this application, we consider a binary covariate $X$ that assumes one when the respondent is an older adult, and zero otherwise. We fit the following ROC-GLM model:

$$\text{ROC}(t) = \Phi(\alpha_0 + \alpha_1 \Phi^{-1}(t) + \beta X),$$

Table 5.7.6 presents the estimated parameters for the model above using the design-adjusted and unadjusted ROC-GLM model. From Table 5.7.6, we see that the design-adjusted point and interval estimates of $\beta$ indicate that the risk assessment tool has a lower accuracy in detecting undiagnosed diabetes among older adults. The estimate and standard error for $\hat{\beta}$ are smaller when the sampling design is not taken into account. Figure 5.7.7 show both survey-weighted and unweighted estimates of the ROC curve and their corresponding AUC.

From Figure 5.7.7 we see that the estimated area under the ROC curve from the unweighted ROC-GLM model is smaller, with a greater departure among old adults (design-adjusted: 0.715, unadjusted: 0.669). As discussed previously, the risk tool performs worse among older adults, with a design-adjusted AUC of 0.715, compared to the design-adjusted AUC of 0.824 among adults. This result seems to be more indirect evidence for supporting the ADA recommendation of screening all adults aged 45 years or more every 1-3 years using either FPG, A1C, or oral glucose test (American Diabetes Association, 2022).

## 5.6 Discussion

In this paper, we adapted the ROC-GLM model proposed by Pepe (2000a) and Alonzo and Pepe (2002) to account for complex survey designs. Similarly as done for data arising from simple random sample, variance estimation was performed using the bootstrap resampling technique in the context of complex survey data. The properties of the proposed method were evaluated using simulation studies and compared to the

original ROC-GLM model, which does not account for the complex survey sampling. As expected, using ROC-GLM for i.i.d. data leads to poor performance in bias, variance estimation, and coverage probabilities. Our adapted method performed well, especially for larger sample sizes and disease proportions. Finally, the proposed method was applied to a national-level health survey to evaluate the discriminatory ability of a traditional risk calculator for undiagnosed diabetes.

The method presented in this paper serves as a basis for ROC regression in the context of complex survey data. The simplicity of the algorithm proposed by Alonzo and Pepe (2002), combined with widely available software for quantile regression and generalized linear model for complex survey data, makes this method especially useful in practice. This method is also attractive to be used with the bootstrap sampling weights provided by the statistical agencies in large public surveys.

Our method relies on the bootstrap resampling method to obtain the variance of the estimated parameters of the model. Although the original method also relies on resampling methods, the development of theoretical properties of the proposed method using theory for generalized estimating equations for correlated binary data will be worth investigating.

## 5.7 Figures and Tables
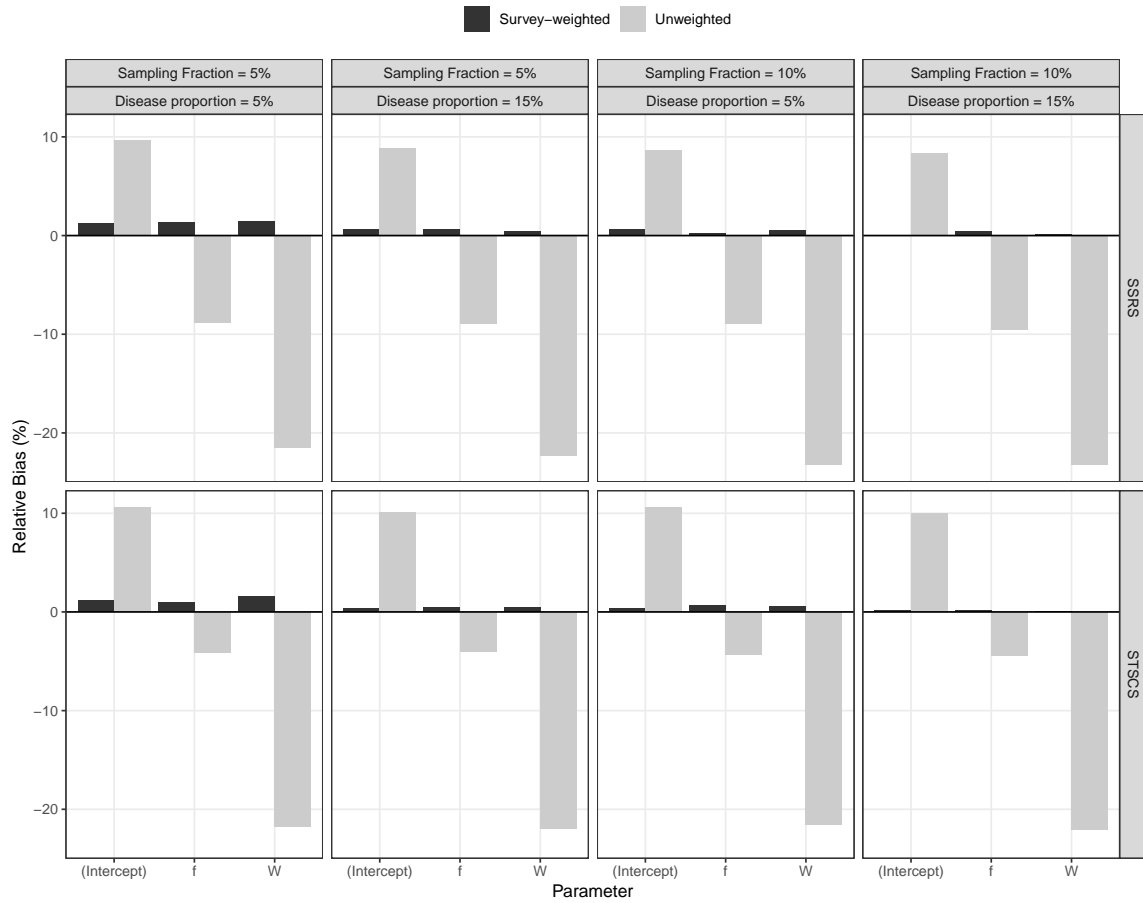


Figure 5.7.1: Relative Bias (in %) for the design-adjusted and unadjusted estimated parameters from the ROC-GLM model by sampling fraction, disease proportion, and sampling design.
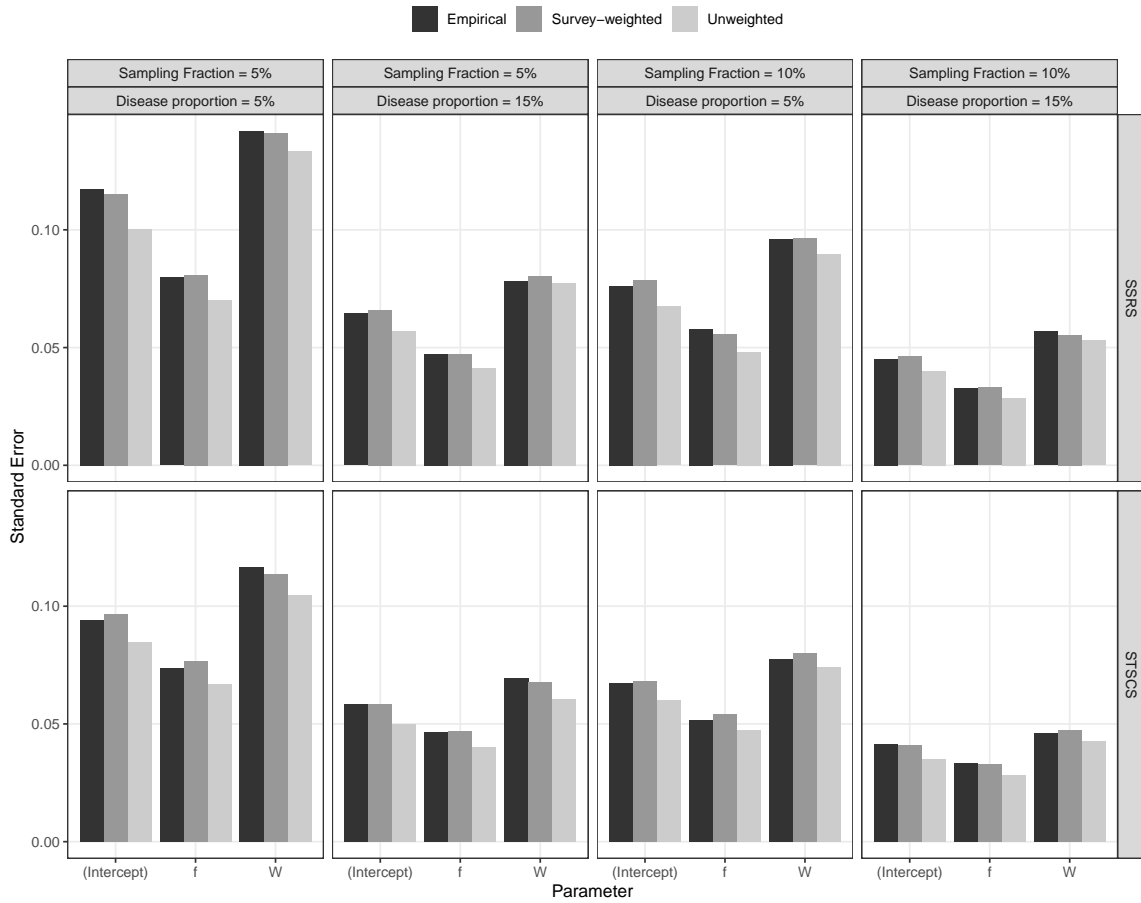
Figure 5.7.2: Empirical and bootstrap standard error for the design-adjusted and unadjusted estimated parameters from the ROC-GLM model by sampling fraction, disease proportion, and sampling design.
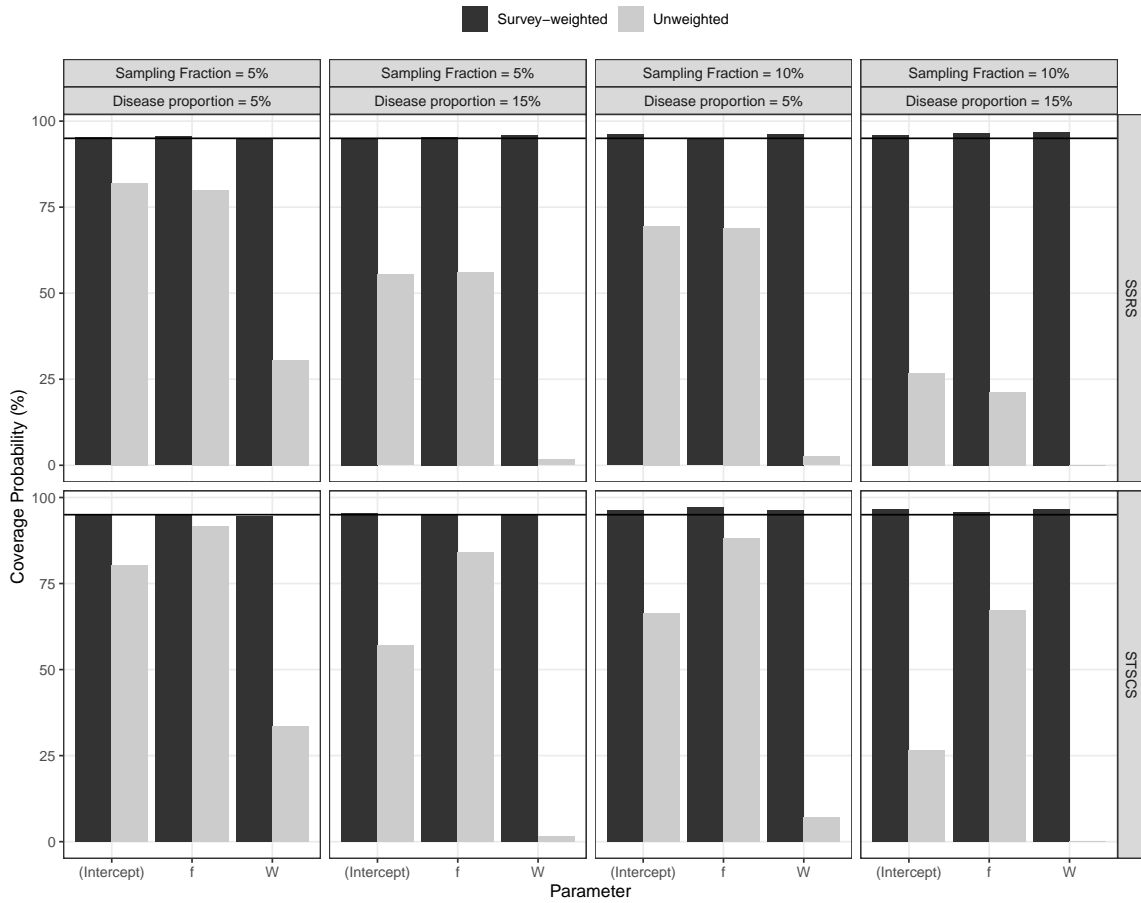
Figure 5.7.3: Empirical coverage probability for nominal 95% confidence intervals of the design-adjusted and unadjusted parameters from the ROC-GLM model by sampling fraction, disease proportion, and sampling design.
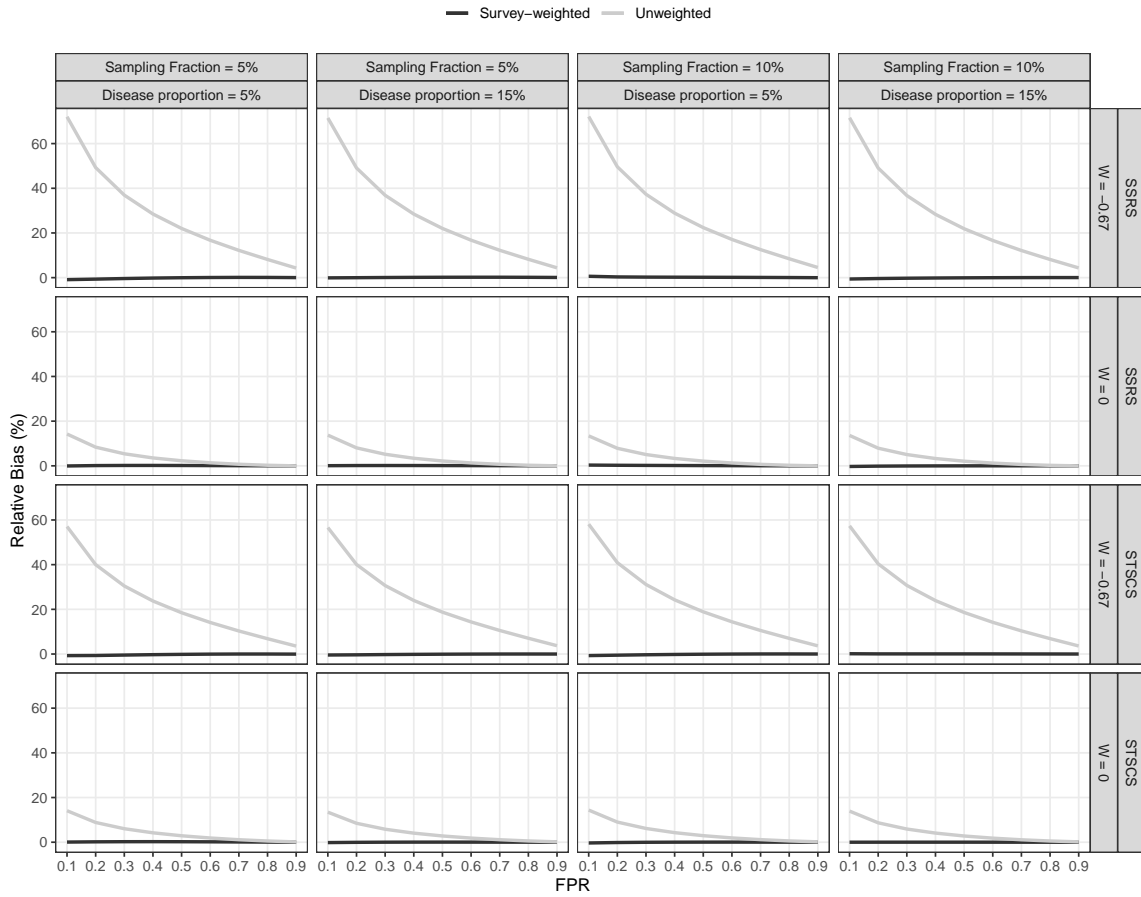
Figure 5.7.4: Relative Bias (in %) for the estimated covariate-specific ROC curve from the design-adjusted and unadjusted ROC-GLM models by sampling fraction, disease proportion, and sampling design.
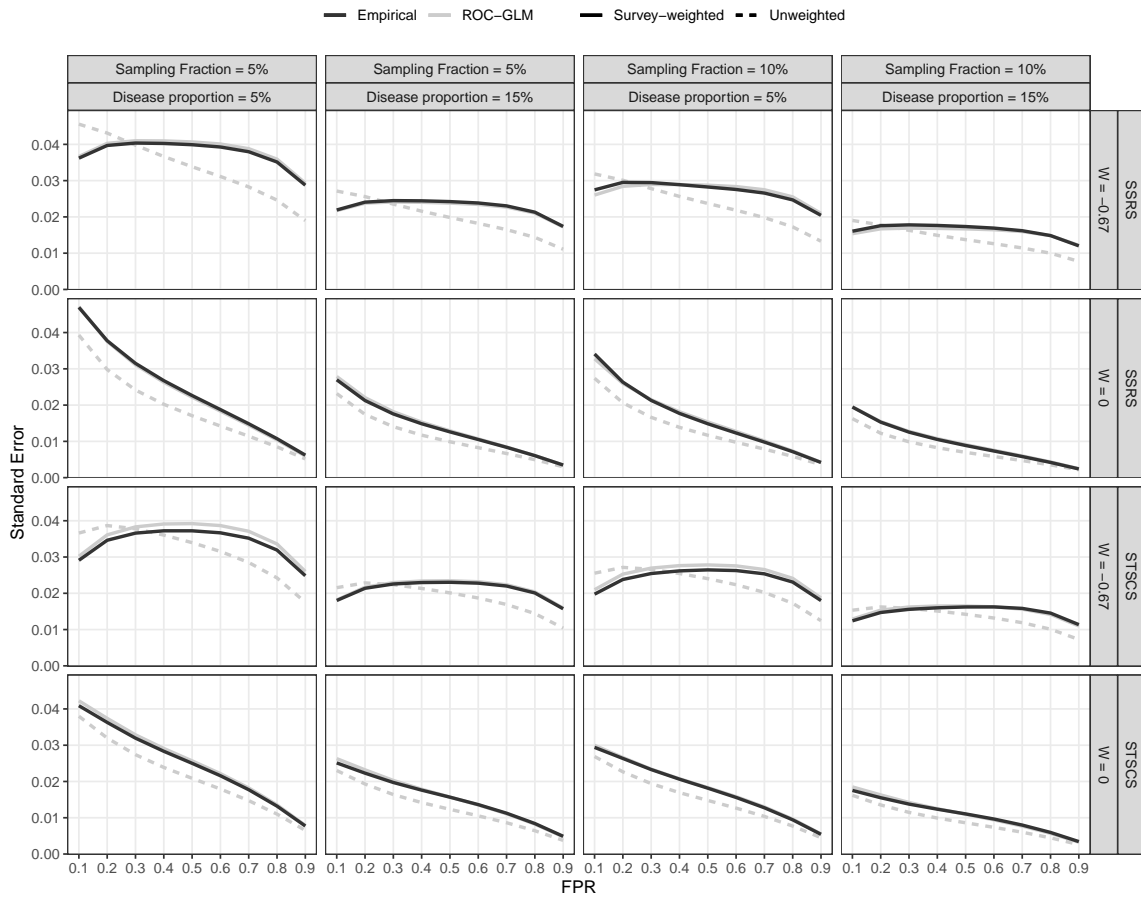
Figure 5.7.5: Empirical and bootstrap standard error for the estimated covariate-specific ROC curve from the design-adjusted and unadjusted ROC-GLM models by sampling fraction, disease proportion, and sampling design.
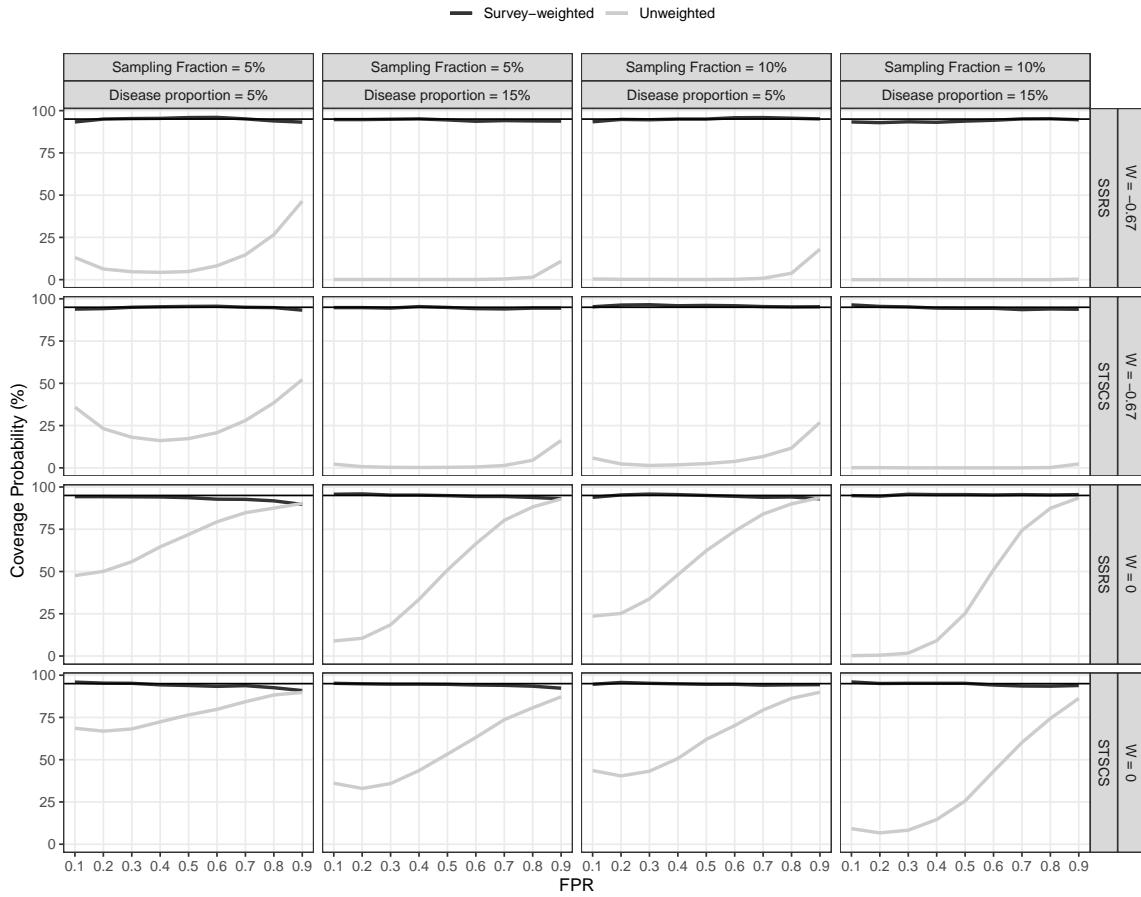
Figure 5.7.6: Empirical coverage probability for nominal 95% confidence intervals of the covariate-specific ROC curve from the design-adjusted and unadjusted ROC-GLM models by sampling fraction, disease proportion, and sampling design.

Table 5.7.1: Simulation results for design-adjusted (SVYW) and unadjusted (UN) ROC-GLM model parameters for stratified simple random sample (SSRS), and stratified two stage cluster sampling (STSCS). $\lambda$: sampling fraction; $p$: disease proportion; RB: relative bias (in %); EM: empirical standard error; SE: standard error; CP: empirical coverage probability for nominal 95% confidence intervals (in %).

| Design | $\lambda$ | $p$ | Parameter | RB | | SE | | | CP | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | SVYW | UN | EM | SVYW | UN | SVYW | UN |
| SSRS | 5% | 5% | $\alpha_0$ | 1.2 | 9.7 | 0.117 | 0.115 | 0.100 | 95.3 | 82.0 |
| | | | $\alpha_1$ | 1.4 | -8.9 | 0.080 | 0.081 | 0.070 | 95.5 | 80.0 |
| | | | $\beta$ | 1.4 | -21.6 | 0.142 | 0.141 | 0.133 | 94.8 | 30.6 |
| | | 15% | $\alpha_0$ | 0.6 | 8.9 | 0.065 | 0.066 | 0.057 | 95.1 | 55.4 |
| | | | $\alpha_1$ | 0.6 | -9.0 | 0.047 | 0.047 | 0.041 | 95.4 | 56.2 |
| | | | $\beta$ | 0.5 | -22.4 | 0.078 | 0.080 | 0.077 | 95.9 | 1.7 |
| | 10% | 5% | $\alpha_0$ | 0.7 | 8.7 | 0.076 | 0.079 | 0.067 | 96.1 | 69.5 |
| | | | $\alpha_1$ | 0.3 | -9.0 | 0.058 | 0.056 | 0.048 | 95.0 | 69.0 |
| | | | $\beta$ | 0.5 | -23.3 | 0.096 | 0.096 | 0.090 | 96.2 | 2.5 |
| | | 15% | $\alpha_0$ | 0.1 | 8.4 | 0.045 | 0.046 | 0.040 | 96.0 | 26.8 |
| | | | $\alpha_1$ | 0.5 | -9.5 | 0.033 | 0.033 | 0.029 | 96.5 | 21.3 |
| | | | $\beta$ | 0.2 | -23.3 | 0.057 | 0.055 | 0.053 | 96.9 | 0.0 |
| STSCS | 5% | 5% | $\alpha_0$ | 1.2 | 10.6 | 0.094 | 0.096 | 0.085 | 94.9 | 80.3 |
| | | | $\alpha_1$ | 1.0 | -4.2 | 0.074 | 0.077 | 0.067 | 95.3 | 91.7 |
| | | | $\beta$ | 1.6 | -21.7 | 0.117 | 0.114 | 0.105 | 94.5 | 33.7 |
| | | 15% | $\alpha_0$ | 0.3 | 10.0 | 0.058 | 0.059 | 0.050 | 95.4 | 57.1 |
| | | | $\alpha_1$ | 0.5 | -4.0 | 0.047 | 0.047 | 0.040 | 95.3 | 84.2 |
| | | | $\beta$ | 0.5 | -21.9 | 0.070 | 0.068 | 0.060 | 95.2 | 1.6 |
| | 10% | 5% | $\alpha_0$ | 0.4 | 10.6 | 0.067 | 0.068 | 0.060 | 96.4 | 66.5 |
| | | | $\alpha_1$ | 0.7 | -4.4 | 0.051 | 0.054 | 0.047 | 97.1 | 88.1 |
| | | | $\beta$ | 0.5 | -21.6 | 0.078 | 0.080 | 0.074 | 96.2 | 7.0 |
| | | 15% | $\alpha_0$ | 0.1 | 10.0 | 0.041 | 0.041 | 0.035 | 96.5 | 26.5 |
| | | | $\alpha_1$ | 0.1 | -4.4 | 0.033 | 0.033 | 0.028 | 95.6 | 67.3 |
| | | | $\beta$ | 0.1 | -22.0 | 0.046 | 0.048 | 0.043 | 96.7 | 0.0 |

Table 5.7.2: Simulation results of relative bias (in %) for estimated covariate-specific ROC curve using design-adjusted (SVYW) and unadjusted (UN) ROC-GLM models for stratified simple random sample (SSRS), and stratified two stage cluster sampling (STSCS). $W$: covariate value; $\lambda$: sampling fraction; $p$: disease proportion.

| Design | FPR | Method | $W = -0.67$ | | | | $W = 0$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\lambda = 5\%$ | | $\lambda = 10\%$ | | $\lambda = 5\%$ | | $\lambda = 10\%$ | |
| | | | $p = 5\%$ | $p = 15\%$ | $p = 5\%$ | $p = 15\%$ | $p = 5\%$ | $p = 15\%$ | $p = 5\%$ | $p = 15\%$ |
| SSRS | 0.1 | SVYW | -0.9 | -0.1 | 0.6 | -0.6 | -0.1 | 0.1 | 0.3 | -0.3 |
| | | UN | 72.0 | 71.5 | 72.1 | 71.6 | 14.2 | 13.7 | 13.4 | 13.6 |
| | 0.2 | SVYW | -0.7 | 0.0 | 0.3 | -0.4 | 0.1 | 0.1 | 0.2 | -0.1 |
| | | UN | 49.3 | 49.1 | 49.6 | 49.1 | 8.4 | 8.0 | 7.8 | 7.9 |
| | 0.3 | SVYW | -0.4 | 0.1 | 0.2 | -0.3 | 0.2 | 0.1 | 0.2 | -0.1 |
| | | UN | 36.9 | 36.9 | 37.3 | 36.8 | 5.4 | 5.2 | 5.1 | 5.0 |
| | 0.4 | SVYW | -0.2 | 0.1 | 0.2 | -0.2 | 0.2 | 0.1 | 0.2 | 0.0 |
| | | UN | 28.5 | 28.5 | 28.9 | 28.3 | 3.5 | 3.4 | 3.3 | 3.3 |
| | 0.5 | SVYW | 0.0 | 0.2 | 0.2 | -0.1 | 0.2 | 0.1 | 0.1 | 0.0 |
| | | UN | 22.0 | 22.0 | 22.4 | 21.9 | 2.3 | 2.2 | 2.1 | 2.1 |
| | 0.6 | SVYW | 0.1 | 0.2 | 0.1 | 0.0 | 0.1 | 0.1 | 0.1 | 0.0 |
| | | UN | 16.7 | 16.8 | 17.1 | 16.6 | 1.4 | 1.3 | 1.3 | 1.2 |
| | 0.7 | SVYW | 0.1 | 0.2 | 0.1 | 0.0 | 0.1 | 0.1 | 0.0 | 0.0 |
| | | UN | 12.2 | 12.3 | 12.5 | 12.2 | 0.7 | 0.7 | 0.7 | 0.6 |
| | 0.8 | SVYW | 0.1 | 0.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | | UN | 8.1 | 8.2 | 8.4 | 8.1 | 0.3 | 0.3 | 0.2 | 0.2 |
| | 0.9 | SVYW | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | | UN | 4.3 | 4.4 | 4.5 | 4.3 | 0.0 | 0.0 | 0.0 | 0.0 |
| STSCS | 0.1 | SVYW | -0.7 | -0.4 | -0.7 | 0.1 | 0.1 | -0.2 | -0.3 | 0.0 |
| | | UN | 57.1 | 56.6 | 58.1 | 57.4 | 14.1 | 13.4 | 14.4 | 13.9 |
| | 0.2 | SVYW | -0.7 | -0.4 | -0.5 | 0.1 | 0.2 | -0.1 | -0.1 | 0.0 |
| | | UN | 40.0 | 40.1 | 40.9 | 40.4 | 8.8 | 8.5 | 9.0 | 8.7 |
| | 0.3 | SVYW | -0.5 | -0.2 | -0.3 | 0.1 | 0.2 | 0.0 | 0.0 | 0.0 |
| | | UN | 30.4 | 30.7 | 31.1 | 30.7 | 6.0 | 5.8 | 6.2 | 5.9 |
| | 0.4 | SVYW | -0.3 | -0.2 | -0.2 | 0.1 | 0.2 | 0.0 | 0.0 | 0.0 |
| | | UN | 23.7 | 24.0 | 24.3 | 23.9 | 4.2 | 4.1 | 4.3 | 4.1 |
| | 0.5 | SVYW | -0.1 | -0.1 | -0.1 | 0.1 | 0.2 | 0.0 | 0.0 | 0.0 |
| | | UN | 18.5 | 18.7 | 18.9 | 18.6 | 2.9 | 2.8 | 2.9 | 2.8 |
| | 0.6 | SVYW | -0.1 | 0.0 | 0.0 | 0.0 | 0.2 | 0.0 | 0.0 | 0.0 |
| | | UN | 14.1 | 14.4 | 14.4 | 14.2 | 1.9 | 1.8 | 1.9 | 1.8 |
| | 0.7 | SVYW | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 |
| | | UN | 10.3 | 10.5 | 10.5 | 10.4 | 1.1 | 1.1 | 1.1 | 1.1 |
| | 0.8 | SVYW | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | | UN | 6.9 | 7.1 | 7.0 | 6.9 | 0.5 | 0.6 | 0.6 | 0.5 |
| | 0.9 | SVYW | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | | UN | 3.6 | 3.7 | 3.7 | 3.6 | 0.1 | 0.2 | 0.2 | 0.2 |

Table 5.7.3: Simulation results of empirical (EM) and boostrap standard errors for the covariate-specific ROC curve using design-adjusted (SVY) and unadjusted (UN) ROC-GLM models for stratified simple random sample (SSRS). $W$: covariate value; $\lambda$: sampling fraction; $p$: disease proportion.

| FPR | Method | $W = -0.67$ | | | | $W = 0$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $\lambda = 5\%$ | | $\lambda = 10\%$ | | $\lambda = 5\%$ | | $\lambda = 10\%$ | |
| | | $p = 5\%$ | $p = 15\%$ | $p = 5\%$ | $p = 15\%$ | $p = 5\%$ | $p = 15\%$ | $p = 5\%$ | $p = 15\%$ |
| 0.1 | EMP | 0.036 | 0.022 | 0.027 | 0.016 | 0.047 | 0.027 | 0.034 | 0.019 |
| | SVY | 0.037 | 0.022 | 0.026 | 0.015 | 0.047 | 0.028 | 0.033 | 0.020 |
| | UN | 0.046 | 0.027 | 0.032 | 0.019 | 0.039 | 0.023 | 0.027 | 0.016 |
| 0.2 | EMP | 0.040 | 0.024 | 0.030 | 0.018 | 0.038 | 0.021 | 0.026 | 0.015 |
| | SVY | 0.040 | 0.024 | 0.028 | 0.017 | 0.038 | 0.022 | 0.026 | 0.015 |
| | UN | 0.043 | 0.026 | 0.030 | 0.018 | 0.030 | 0.017 | 0.021 | 0.012 |
| 0.3 | EMP | 0.040 | 0.024 | 0.029 | 0.018 | 0.032 | 0.018 | 0.021 | 0.013 |
| | SVY | 0.041 | 0.024 | 0.029 | 0.017 | 0.031 | 0.018 | 0.021 | 0.013 |
| | UN | 0.040 | 0.024 | 0.028 | 0.016 | 0.024 | 0.014 | 0.017 | 0.010 |
| 0.4 | EMP | 0.040 | 0.024 | 0.029 | 0.018 | 0.027 | 0.015 | 0.018 | 0.011 |
| | SVY | 0.041 | 0.024 | 0.029 | 0.017 | 0.026 | 0.015 | 0.018 | 0.011 |
| | UN | 0.037 | 0.022 | 0.026 | 0.015 | 0.020 | 0.012 | 0.014 | 0.008 |
| 0.5 | EMP | 0.040 | 0.024 | 0.028 | 0.017 | 0.023 | 0.013 | 0.015 | 0.009 |
| | SVY | 0.041 | 0.024 | 0.029 | 0.017 | 0.022 | 0.013 | 0.015 | 0.009 |
| | UN | 0.034 | 0.020 | 0.024 | 0.014 | 0.017 | 0.010 | 0.012 | 0.007 |
| 0.6 | EMP | 0.039 | 0.024 | 0.028 | 0.017 | 0.019 | 0.011 | 0.012 | 0.007 |
| | SVY | 0.040 | 0.023 | 0.028 | 0.016 | 0.018 | 0.011 | 0.013 | 0.008 |
| | UN | 0.031 | 0.018 | 0.022 | 0.013 | 0.014 | 0.008 | 0.010 | 0.006 |
| 0.7 | EMP | 0.038 | 0.023 | 0.027 | 0.016 | 0.015 | 0.008 | 0.010 | 0.006 |
| | SVY | 0.039 | 0.023 | 0.027 | 0.016 | 0.014 | 0.008 | 0.010 | 0.006 |
| | UN | 0.028 | 0.016 | 0.020 | 0.011 | 0.011 | 0.007 | 0.008 | 0.005 |
| 0.8 | EMP | 0.035 | 0.021 | 0.025 | 0.015 | 0.011 | 0.006 | 0.007 | 0.004 |
| | SVY | 0.036 | 0.021 | 0.026 | 0.015 | 0.010 | 0.006 | 0.007 | 0.004 |
| | UN | 0.025 | 0.014 | 0.017 | 0.010 | 0.008 | 0.005 | 0.006 | 0.004 |
| 0.9 | EMP | 0.029 | 0.017 | 0.020 | 0.012 | 0.006 | 0.003 | 0.004 | 0.002 |
| | SVY | 0.029 | 0.017 | 0.021 | 0.012 | 0.006 | 0.003 | 0.004 | 0.002 |
| | UN | 0.019 | 0.011 | 0.013 | 0.008 | 0.005 | 0.003 | 0.004 | 0.002 |

Table 5.7.4: Simulation results of empirical (EM) and boostrap standard errors for the covariate-specific ROC curve using design-adjusted (SVY) and unadjusted (UN) ROC-GLM models for stratified two stage cluster sampling (STSCS). $W$: covariate value; $\lambda$: sampling fraction; $p$: disease proportion.

| FPR | Method | $W = -0.67$ | | | | $W = 0$ | | | |
| | | $\lambda = 5\%$ | | $\lambda = 10\%$ | | $\lambda = 5\%$ | | $\lambda = 10\%$ | |
| | | $p = 5\%$ | $p = 15\%$ | $p = 5\%$ | $p = 15\%$ | $p = 5\%$ | $p = 15\%$ | $p = 5\%$ | $p = 15\%$ |
|---|---|---|---|---|---|---|---|---|---|
| 0.1 | EMP | 0.029 | 0.018 | 0.020 | 0.012 | 0.041 | 0.025 | 0.029 | 0.018 |
| | SVY | 0.030 | 0.018 | 0.021 | 0.013 | 0.042 | 0.026 | 0.030 | 0.019 |
| | UN | 0.037 | 0.022 | 0.026 | 0.015 | 0.038 | 0.023 | 0.027 | 0.016 |
| 0.2 | EMP | 0.035 | 0.021 | 0.024 | 0.015 | 0.036 | 0.022 | 0.026 | 0.016 |
| | SVY | 0.036 | 0.022 | 0.025 | 0.015 | 0.037 | 0.023 | 0.027 | 0.016 |
| | UN | 0.039 | 0.023 | 0.027 | 0.016 | 0.032 | 0.019 | 0.023 | 0.014 |
| 0.3 | EMP | 0.037 | 0.023 | 0.025 | 0.016 | 0.032 | 0.020 | 0.023 | 0.014 |
| | SVY | 0.038 | 0.023 | 0.027 | 0.016 | 0.033 | 0.020 | 0.023 | 0.014 |
| | UN | 0.038 | 0.022 | 0.027 | 0.016 | 0.027 | 0.016 | 0.019 | 0.011 |
| 0.4 | EMP | 0.037 | 0.023 | 0.026 | 0.016 | 0.028 | 0.018 | 0.021 | 0.012 |
| | SVY | 0.039 | 0.023 | 0.028 | 0.016 | 0.029 | 0.018 | 0.021 | 0.012 |
| | UN | 0.036 | 0.021 | 0.025 | 0.015 | 0.024 | 0.014 | 0.017 | 0.010 |
| 0.5 | EMP | 0.037 | 0.023 | 0.026 | 0.016 | 0.025 | 0.016 | 0.018 | 0.011 |
| | SVY | 0.039 | 0.023 | 0.028 | 0.016 | 0.026 | 0.016 | 0.018 | 0.011 |
| | UN | 0.034 | 0.020 | 0.024 | 0.014 | 0.021 | 0.012 | 0.015 | 0.009 |
| 0.6 | EMP | 0.037 | 0.023 | 0.026 | 0.016 | 0.022 | 0.014 | 0.016 | 0.010 |
| | SVY | 0.039 | 0.023 | 0.028 | 0.016 | 0.022 | 0.014 | 0.016 | 0.009 |
| | UN | 0.032 | 0.019 | 0.022 | 0.013 | 0.018 | 0.011 | 0.013 | 0.007 |
| 0.7 | EMP | 0.035 | 0.022 | 0.025 | 0.016 | 0.018 | 0.011 | 0.013 | 0.008 |
| | SVY | 0.037 | 0.022 | 0.027 | 0.016 | 0.018 | 0.011 | 0.013 | 0.008 |
| | UN | 0.028 | 0.017 | 0.020 | 0.012 | 0.015 | 0.009 | 0.010 | 0.006 |
| 0.8 | EMP | 0.032 | 0.020 | 0.023 | 0.015 | 0.013 | 0.008 | 0.009 | 0.006 |
| | SVY | 0.034 | 0.020 | 0.024 | 0.014 | 0.014 | 0.008 | 0.010 | 0.006 |
| | UN | 0.024 | 0.014 | 0.017 | 0.010 | 0.011 | 0.006 | 0.008 | 0.004 |
| 0.9 | EMP | 0.025 | 0.016 | 0.018 | 0.011 | 0.008 | 0.005 | 0.005 | 0.003 |
| | SVY | 0.026 | 0.016 | 0.019 | 0.011 | 0.008 | 0.005 | 0.006 | 0.003 |
| | UN | 0.018 | 0.010 | 0.012 | 0.007 | 0.007 | 0.004 | 0.005 | 0.003 |

Table 5.7.5: Simulation results of empirical coverage probability (in %) for nominal 95% boostrap confidence intervals for the estimated covariate-specific ROC curve using design-adjusted (SVYW) and unadjusted (UN) ROC-GLM models for stratified simple random sample (SSRS), and stratified two stage cluster sampling (STSCS). $W$: covariate value; $\lambda$: sampling fraction; $p$: disease proportion.

| | | | $W = -0.67$ | | | | $W = 0$ | | | |
| | | | $\lambda = 5\%$ | | $\lambda = 10\%$ | | $\lambda = 5\%$ | | $\lambda = 10\%$ | |
| Design | FPR | Method | $p = 5\%$ | $p = 15\%$ | $p = 5\%$ | $p = 15\%$ | $p = 5\%$ | $p = 15\%$ | $p = 5\%$ | $p = 15\%$ |
|---|---|---|---|---|---|---|---|---|---|---|
| SSRS | 0.1 | SVYW | 93.3 | 94.7 | 93.4 | 93.3 | 94.3 | 95.7 | 93.9 | 94.9 |
| | | UN | 13.1 | 0.1 | 0.4 | 0.0 | 47.6 | 8.9 | 23.6 | 0.2 |
| | 0.2 | SVYW | 95.0 | 94.7 | 94.8 | 92.9 | 94.3 | 95.9 | 95.3 | 94.6 |
| | | UN | 6.3 | 0.1 | 0.2 | 0.0 | 50.1 | 10.5 | 25.2 | 0.5 |
| | 0.3 | SVYW | 95.3 | 94.9 | 94.6 | 93.4 | 94.2 | 95.2 | 95.8 | 95.7 |
| | | UN | 4.7 | 0.1 | 0.2 | 0.0 | 55.8 | 18.5 | 33.8 | 1.7 |
| | 0.4 | SVYW | 95.4 | 95.1 | 95.0 | 93.1 | 94.1 | 95.2 | 95.5 | 95.5 |
| | | UN | 4.3 | 0.1 | 0.1 | 0.0 | 64.6 | 33.5 | 48.1 | 9.1 |
| | 0.5 | SVYW | 95.9 | 94.5 | 95.0 | 93.8 | 93.7 | 94.9 | 95.0 | 95.5 |
| | | UN | 4.8 | 0.1 | 0.1 | 0.0 | 71.9 | 50.9 | 62.3 | 25.2 |
| | 0.6 | SVYW | 96.0 | 93.7 | 95.8 | 94.3 | 92.8 | 94.4 | 94.5 | 95.3 |
| | | UN | 8.2 | 0.1 | 0.2 | 0.0 | 79.4 | 66.4 | 73.9 | 51.0 |
| | 0.7 | SVYW | 95.1 | 94.1 | 95.9 | 95.1 | 92.7 | 94.4 | 93.9 | 95.5 |
| | | UN | 14.7 | 0.4 | 0.8 | 0.0 | 84.8 | 80.3 | 84.0 | 74.3 |
| | 0.8 | SVYW | 93.9 | 93.9 | 95.5 | 95.2 | 91.8 | 93.8 | 94.1 | 95.3 |
| | | UN | 26.6 | 1.4 | 3.8 | 0.0 | 87.5 | 88.2 | 90.0 | 87.4 |
| | 0.9 | SVYW | 93.2 | 93.8 | 95.1 | 94.6 | 89.7 | 93.1 | 93.0 | 95.5 |
| | | UN | 46.4 | 10.9 | 18.0 | 0.3 | 90.3 | 92.9 | 93.7 | 93.6 |
| STSCS | 0.1 | SVYW | 93.9 | 94.8 | 95.3 | 96.4 | 95.9 | 95.2 | 94.6 | 96.0 |
| | | UN | 35.9 | 2.2 | 5.8 | 0.1 | 68.6 | 36.1 | 43.6 | 9.2 |
| | 0.2 | SVYW | 94.2 | 94.8 | 96.3 | 95.5 | 95.3 | 94.9 | 95.7 | 95.1 |
| | | UN | 23.2 | 0.7 | 2.3 | 0.1 | 66.9 | 33.0 | 40.4 | 6.7 |
| | 0.3 | SVYW | 95.0 | 94.5 | 96.5 | 95.2 | 95.2 | 94.7 | 95.2 | 95.2 |
| | | UN | 18.1 | 0.3 | 1.4 | 0.0 | 68.2 | 35.9 | 43.2 | 8.3 |
| | 0.4 | SVYW | 95.3 | 95.4 | 95.9 | 94.5 | 94.3 | 94.7 | 94.9 | 95.2 |
| | | UN | 16.1 | 0.2 | 1.8 | 0.0 | 72.4 | 43.6 | 50.7 | 14.6 |
| | 0.5 | SVYW | 95.5 | 94.9 | 96.1 | 94.4 | 93.9 | 94.6 | 94.6 | 95.2 |
| | | UN | 17.3 | 0.3 | 2.5 | 0.0 | 76.5 | 53.3 | 62.0 | 25.6 |
| | 0.6 | SVYW | 95.6 | 94.2 | 95.9 | 94.4 | 93.4 | 94.2 | 94.6 | 94.2 |
| | | UN | 20.8 | 0.5 | 3.8 | 0.0 | 79.8 | 63.2 | 70.2 | 43.1 |
| | 0.7 | SVYW | 95.0 | 94.0 | 95.4 | 93.6 | 93.8 | 94.0 | 94.1 | 93.6 |
| | | UN | 28.0 | 1.4 | 6.7 | 0.0 | 84.3 | 73.7 | 79.4 | 60.2 |
| | 0.8 | SVYW | 94.8 | 94.5 | 95.2 | 94.0 | 92.6 | 93.5 | 94.3 | 93.5 |
| | | UN | 38.5 | 4.5 | 11.7 | 0.2 | 88.3 | 80.7 | 86.3 | 74.4 |
| | 0.9 | SVYW | 93.3 | 94.5 | 95.3 | 93.8 | 90.9 | 92.3 | 94.4 | 93.9 |
| | | UN | 52.2 | 16.2 | 26.8 | 2.3 | 89.8 | 87.2 | 89.9 | 86.3 |
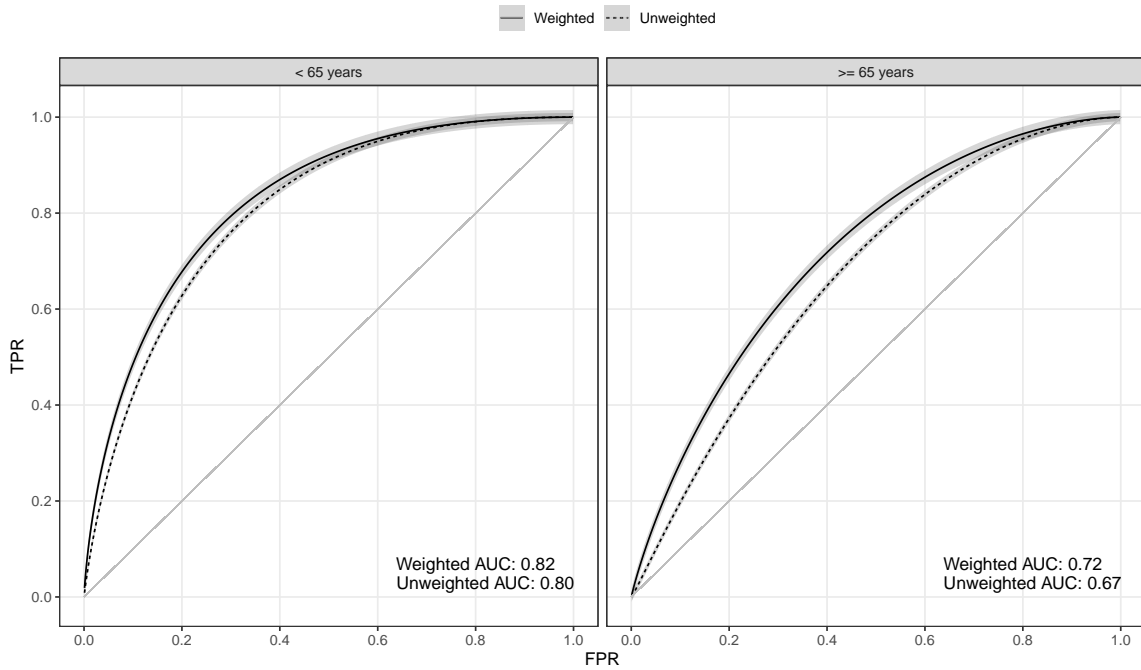
Figure 5.7.7: Estimated ROC curves with 95% boostrap confidence interval (shaded) from unweighted ROC-GLM model (dashed) and the proposed design-adjusted ROC-GLM model (solid) according to age group in the NHANES target population.

Table 5.7.6: Survey-weighted and weighted fitted ROC-GLM model with probit link.

| Parameter | Survey-weighted | | | Unweighted | | |
| | Estimate | Std. Error | 95% CI | Estimate | Std. Error | 95% CI |
| --- | --- | --- | --- | --- | --- | --- |
| $\alpha_0$ | 1.41 | 0.19 | (1.05, 1.78) | 1.34 | 0.13 | (1.08, 1.59) |
| $\alpha_1$ | 1.13 | 0.10 | (0.93, 1.33) | 1.20 | 0.08 | (1.05, 1.35) |
| $\beta$ | -0.55 | 0.20 | (-0.94, -0.16) | -0.65 | 0.15 | (-0.95, -0.35) |

## CHAPTER 6: DISCUSSION AND FUTURE WORK

This dissertation focused on developing methods for the receiver operating characteristic (ROC) curve in the context of complex survey data, motivated by many applications using ROC curves in large-scale surveys, but ignoring their complex survey designs. First, we proposed a non-parametric estimator for the ROC curve based on the Horwitz-Thompson estimator. We used the general results developed in Han and Wellner (2021) to study the theoretical properties of our proposed estimator using empirical process arguments. The proposed estimator was developed for the complete-case scenario. Since nonresponse is a common issue in large-scale surveys, we proposed an IPW estimator for the ROC curve to accommodate the case where the diagnostic test is missing. We developed the theoretical aspects of the estimator following Han and Wellner (2021) using empirical process arguments. Finally, to address the common desire to study covariate effects on the accuracy of a diagnostic test, we adapted the ROC-GLM model proposed originally by Pepe (2000a) to account for the complex survey design. Our work contributes toward the broader availability of methods and software for the ROC curve for complex survey sampling.

Our proposed estimator in Chapter 3 is a discrete rather than a continuous function as the true ROC curve. Thus, the study of semiparametric and parametric estimators for the ROC curve in the context of complex survey data deserves attention. In addition to smoothness, these alternative approaches might be more efficient in estimating the ROC curve if the model is correctly specified.

A natural extension for the proposed IPW estimator in Chapter 4 is the Augmented IPW (AIPW) estimator, where a model is proposed for the diagnostic test in addition

to the model for the missing mechanism. In this context, one can obtain an estimator that is doubly robust to model specification. Alternative methods using imputation for complex survey data might also be a possibility.

Our adapted ROC-GLM model from Chapter 5 relies on bootstrap resampling for variance estimation. Although the original method also relies on resampling methods for variance estimation, the development of theoretical properties of the proposed method using design-based theory for generalized estimation equations might be worth investigating.

To close the gap in the availability of software to handle the estimation of ROC curves in the context of complex survey data, an R package with the methods presented in this work is under development.

# APPENDIX 1: TECHNICAL DETAILS FOR CHAPTER 3

## A.1 Proofs

*Proof of Theorem 2.1 (a).* Let $\mathcal{F} = \{f_{s,l} \equiv f_{s,l}(x,d) = I(x \leq s, d = l) : s \in \mathbb{R}, l \in \{0,1\}\}$, and $f_{s,d}, f_{u,d'} \in \mathcal{F}$, with $s, u \in \mathbb{R}$, and $d, d' \in \{0,1\}$. From Corollary 3.13 in Han and Wellner (2021), it follows that

$$\sqrt{n}(\mathbb{P}_N^\pi - \mathbb{P}_N) \rightsquigarrow \mathbb{G}^\pi \quad \text{in} \quad \ell^\infty(\mathcal{F}),$$

where $\mathbb{G}^\pi$ is a tight Gaussian process with covariance function

$$\text{Cov}(\mathbb{G}^\pi(f_{s,d}), \mathbb{G}^\pi(f_{u,d'})) = \lambda \left( \mu_{\pi_1} P(f_{s,d} f_{u,d'}) + \mu_{\pi_2} (P f_{s,d})(P f_{u,d'}) \right)$$

$$= \begin{cases} \lambda(1-p) \left[ \mu_{\pi_1} G(s \wedge u) + \mu_{\pi_2}(1-p)G(s)G(u) \right], & d = d' = 0 \\[2mm] \lambda p \left[ \mu_{\pi_1} F(s \wedge u) + \mu_{\pi_2} p F(s) F(u) \right], & d = d' = 1 \\[2mm] \lambda \mu_{\pi_2} p(1-p)G(s)F(u), & 0 = d \neq d' = 1 \\[2mm] \lambda \mu_{\pi_2} p(1-p)G(u)F(s), & 1 = d \neq d' = 0 \end{cases}$$

For $s \in \mathbb{R}$, the estimators for cdfs $G(s)$ and $F(s)$ are

$$G_n(s) = \frac{\mathbb{P}_N^\pi(f_{s,0})}{\mathbb{P}_N^\pi(f_{\infty,0})} \qquad F_n(s) = \frac{\mathbb{P}_N^\pi(f_{s,1})}{\mathbb{P}_N^\pi(f_{\infty,1})}$$

The ratio map $\phi(A, B) = A/B$ is Hadamard-differentiable with derivative $\phi'(\alpha, \beta) = \alpha/B - (A\beta)/B^2$. It follows from Functional Delta Method (Vaart and Wellner, 1996) that

$$\sqrt{n} \begin{bmatrix} G_n(s) - G_N(s) \\ F_n(s) - F_N(s) \end{bmatrix} \rightsquigarrow \begin{bmatrix} \mathbb{G}_0^\pi(s) \\ \mathbb{G}_1^\pi(s) \end{bmatrix} = \begin{bmatrix} (1-p)^{-1}(\mathbb{G}^\pi(f_{s,0}) - G(s)\mathbb{G}^\pi(f_{\infty,0})) \\ p^{-1}(\mathbb{G}^\pi(f_{s,1}) - F(s)\mathbb{G}^\pi(f_{\infty,1})) \end{bmatrix}.$$

The covariance structure of $\mathbb{G}_0^\pi$ and $\mathbb{G}_1^\pi$ can be computed as follows:

$$
\begin{aligned}
\operatorname{Cov}(\mathbb{G}_0^\pi(s), \mathbb{G}_0^\pi(u)) =& (1-p)^{-2}\Big[\operatorname{Cov}(\mathbb{G}^\pi f_{s,0}, \mathbb{G}^\pi f_{u,0}) + G(s)G(u)\operatorname{Cov}(\mathbb{G}^\pi f_{\infty,0}, \mathbb{G}^\pi f_{\infty,0}) \\
& - G(u)\operatorname{Cov}(\mathbb{G}^\pi f_{s,0}, \mathbb{G}^\pi f_{\infty,0}) - G(s)\operatorname{Cov}(\mathbb{G}^\pi f_{\infty,0}, \mathbb{G}^\pi f_{u,0})\Big] \\
=& \lambda(1-p)^{-1}\Big[\mu_{\pi_1}G(s \wedge u) + \mu_{\pi_2}(1-p)G(s)G(u)+ \\
& + \mu_{\pi_1}G(s)G(u) + \mu_{\pi_2}(1-p)G(s)G(u) \\
& - \mu_{\pi_1}G(u)G(s) - \mu_{\pi_2}(1-p)G(u)G(s)- \\
& - \mu_{\pi_1}G(s)G(u) - \mu_{\pi_2}(1-p)G(s)G(u)\Big] \\
=& \lambda(1-p)^{-1}\mu_{\pi_1}\Big[G(s \wedge u) - G(s)G(u)\Big]
\end{aligned}
$$

$$
\begin{aligned}
\operatorname{Cov}(\mathbb{G}_1^\pi(s), \mathbb{G}_1^\pi(u)) =& p^{-2}\Big[\operatorname{Cov}(\mathbb{G}^\pi f_{s,1}, \mathbb{G}^\pi f_{u,1}) + F(s)F(u)\operatorname{Cov}(\mathbb{G}^\pi f_{\infty,1}, \mathbb{G}^\pi f_{\infty,1}) \\
& - F(u)\operatorname{Cov}(\mathbb{G}^\pi f_{s,1}, \mathbb{G}^\pi f_{\infty,1}) - F(s)\operatorname{Cov}(\mathbb{G}^\pi f_{\infty,1}, \mathbb{G}^\pi f_{u,1})\Big] \\
=& \lambda p^{-1}\Big[\mu_{\pi_1}F(s \wedge u) + \mu_{\pi_2}pF(s)F(u) + \mu_{\pi_1}F(s)F(u) + \mu_{\pi_2}pF(s)F(u) \\
& - \mu_{\pi_1}F(u)F(s) - \mu_{\pi_2}pF(u)F(s) - \mu_{\pi_1}F(s)F(u) + \mu_{\pi_2}pF(s)F(u)\Big] \\
=& \lambda p^{-1}\mu_{\pi_1}\Big[F(s \wedge u) - F(s)F(u)\Big]
\end{aligned}
$$

$$
\begin{aligned}
\operatorname{Cov}(\mathbb{G}_0^\pi(s), \mathbb{G}_1^\pi(u)) =& p^{-1}(1-p)^{-1}\Big[\operatorname{Cov}(\mathbb{G}^\pi f_{s,0}, \mathbb{G}^\pi f_{u,1}) + G(s)F(u)\operatorname{Cov}(\mathbb{G}^\pi f_{\infty,0}, \mathbb{G}^\pi f_{\infty,1}) \\
& - F(u)\operatorname{Cov}(\mathbb{G}^\pi f_{s,0}, \mathbb{G}^\pi f_{\infty,1}) - G(s)\operatorname{Cov}(\mathbb{G}^\pi f_{\infty,0}, \mathbb{G}^\pi f_{u,1})\Big] \\
=& \lambda\mu_{\pi_2}G(s)F(u) + \lambda\mu_{\pi_2}G(s)F(u) \\
& - \lambda\mu_{\pi_2}G(s)F(u) - \lambda\mu_{\pi_2}G(s)F(u) = 0
\end{aligned}
$$

So, the covariance function is given by

$$\text{Cov}(\mathbb{G}_d^\pi(s), \mathbb{G}_{d'}^\pi(u)) = \begin{cases} \lambda(1-p)^{-1}\mu_{\pi_1}(G(u \wedge s) - G(u)G(s)), & d = d' = 0 \\ \lambda p^{-1}\mu_{\pi_1}(F(u \wedge s) - F(u)F(s)), & d = d' = 1 \\ 0, & d \neq d' \end{cases} \qquad (7.24)$$

which implies that

$$\sqrt{n}\begin{bmatrix} G_n(s) - G_N(s) \\ F_n(s) - F_N(s) \end{bmatrix} \rightsquigarrow \begin{bmatrix} \mathbb{G}_0^\pi(s) \\ \mathbb{G}_1^\pi(s) \end{bmatrix} = \begin{bmatrix} \{\lambda(1-p)^{-1}\mu_{\pi_1}\}^{1/2}B_1(G(s)) \\ \{\lambda p^{-1}\mu_{\pi_1}\}^{1/2}B_2(F(s)) \end{bmatrix}, \qquad (7.25)$$

where $B_1(.)$ and $B_2(.)$ denote two independent Brownian bridges and $p = P(D = 1)$.

■

*Proof of Theorem 2.1 (b).* The ROC estimator depends on the pair $(G_n, F_n)$ through the map $\psi(A, B) = B(A^{-1})$, where $A^{-1}$ is the inverse map of $A$. The map $\psi(G, F)$ is Hadamard-differentiable (Lemma 12.2 and Lemma 12.7 from Kosorok (2008)) with derivative

$$\psi'(\alpha, \beta) = \beta(G^{-1}) - \frac{f(G^{-1})}{g(G^{-1})}\alpha(G^{-1})$$

It follows from (7.25) and Functional Delta Method that

$$\sqrt{n}(F_n \circ G_n^{-1}(s) - F_N \circ G_N^{-1}(s))$$

$$\rightsquigarrow \sqrt{\lambda\mu_{\pi_1}}\left\{ p^{-1/2}B_1(F \circ G^{-1}(s)) - (1-p)^{-1/2}\frac{f(G^{-1})}{g(G^{-1})}B_2(s) \right\}$$

where $B_1(.)$ and $B_2(.)$ denote two independent Brownian bridges and $p = P(D = 1)$.

■

*Proof of Theorem 2.1 (c).* It follows from Theorem 2.1 (b) and Continuous Mapping Theorem that

$$\sqrt{n}(A_n - A_N) = \int_0^1 \sqrt{n}(R_n(s) - R_N(s)) \ ds$$

$$\rightsquigarrow \int_0^1 \mathbb{W}\{G^{-1}(1-s)\}ds \sim N(0, \sigma^2)$$

where

$$\sigma^2 = \int_0^1 \int_0^1 \sigma^2\{G^{-1}(1-s), G^{-1}(1-t)\} \ ds \ dt$$

$$= \int_0^1 \int_0^1 \sigma^2\{G^{-1}(s), G^{-1}(t)\} \ ds \ dt$$

$$= \int_{-\infty}^\infty \int_{-\infty}^\infty \sigma^2(s,t) \ dG(s) \ dG(t)$$

∎

## A.2 Supplemental Simulation Tables

Table A.2.1: Relative Bias (in %) of the SVY, WT, and UN, estimators for the super-population ROC curve with finite population size $N$, disease proportion $p$, and sampling fraction $\lambda$ under SSRS.

| | | $N = 50,000$ | | | | $N = 100,000$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $p = 5\%$ | | $p = 25\%$ | | $p = 5\%$ | | $p = 25\%$ | |
| FPR | Method | $\lambda = 5\%$ | $\lambda = 10\%$ | $\lambda = 5\%$ | $\lambda = 10\%$ | $\lambda = 5\%$ | $\lambda = 10\%$ | $\lambda = 5\%$ | $\lambda = 10\%$ |
| | SVY | 0.8 | 0.2 | 0.1 | 0.2 | 0.8 | 0.6 | 0.0 | 0.2 |
| 0.1 | WT | 0.8 | 0.2 | 0.1 | 0.2 | 0.8 | 0.6 | 0.0 | 0.2 |
| | UN | 32.5 | 32.0 | 31.7 | 31.8 | 32.4 | 32.4 | 31.8 | 31.8 |
| | SVY | 0.7 | 0.4 | -0.2 | -0.4 | 0.6 | 0.6 | -0.4 | -0.2 |
| 0.2 | WT | 0.7 | 0.4 | -0.2 | -0.4 | 0.6 | 0.6 | -0.4 | -0.2 |
| | UN | 20.4 | 20.2 | 19.4 | 19.3 | 20.4 | 20.4 | 19.4 | 19.4 |
| | SVY | 0.2 | 0.1 | 0.0 | 0.0 | 0.1 | 0.1 | 0.0 | 0.1 |
| 0.3 | WT | 0.2 | 0.1 | 0.0 | 0.0 | 0.1 | 0.1 | 0.0 | 0.1 |
| | UN | 13.6 | 13.5 | 13.5 | 13.5 | 13.6 | 13.6 | 13.5 | 13.5 |
| | SVY | 0.1 | 0.2 | 0.0 | 0.0 | 0.1 | 0.1 | 0.0 | 0.1 |
| 0.4 | WT | 0.1 | 0.2 | 0.0 | 0.0 | 0.1 | 0.1 | 0.0 | 0.1 |
| | UN | 9.6 | 9.6 | 9.4 | 9.4 | 9.6 | 9.6 | 9.4 | 9.5 |
| | SVY | 0.0 | 0.0 | 0.1 | 0.0 | -0.1 | 0.0 | 0.0 | 0.1 |
| 0.5 | WT | 0.0 | 0.0 | 0.1 | 0.0 | -0.1 | 0.0 | 0.0 | 0.1 |
| | UN | 6.5 | 6.5 | 6.6 | 6.6 | 6.5 | 6.6 | 6.6 | 6.6 |
| | SVY | 0.0 | 0.0 | 0.1 | 0.1 | -0.1 | 0.0 | 0.0 | 0.1 |
| 0.6 | WT | 0.0 | 0.0 | 0.1 | 0.1 | -0.1 | 0.0 | 0.0 | 0.1 |
| | UN | 4.4 | 4.4 | 4.5 | 4.5 | 4.4 | 4.4 | 4.5 | 4.5 |
| | SVY | -0.1 | -0.1 | 0.0 | 0.0 | -0.1 | -0.1 | -0.1 | 0.0 |
| 0.7 | WT | -0.1 | -0.1 | 0.0 | 0.0 | -0.1 | -0.1 | -0.1 | 0.0 |
| | UN | 2.7 | 2.7 | 2.8 | 2.8 | 2.6 | 2.7 | 2.7 | 2.7 |
| | SVY | -0.2 | -0.1 | 0.0 | 0.0 | -0.2 | -0.1 | 0.0 | 0.0 |
| 0.8 | WT | -0.2 | -0.1 | 0.0 | 0.0 | -0.2 | -0.1 | 0.0 | 0.0 |
| | UN | 1.3 | 1.4 | 1.5 | 1.5 | 1.4 | 1.4 | 1.5 | 1.5 |
| | SVY | -0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 0.9 | WT | -0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | UN | 0.5 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 |

Table A.2.2: Relative Bias (in %) of the SVY, WT, and UN estimators for the super-population ROC curve with finite population size $N$, disease proportion $p$, and sampling fraction $\lambda$ under STSCS.

| | | $N = 50{,}000$ | | | | $N = 100{,}000$ | | | |
| | | $p = 5\%$ | | $p = 25\%$ | | $p = 5\%$ | | $p = 25\%$ | |
| FPR | Method | $\lambda = 5\%$ | $\lambda = 10\%$ | $\lambda = 5\%$ | $\lambda = 10\%$ | $\lambda = 5\%$ | $\lambda = 10\%$ | $\lambda = 5\%$ | $\lambda = 10\%$ |
|---|---|---|---|---|---|---|---|---|---|
| 0.1 | SVY | -0.1 | 0.1 | 0.2 | 0.0 | -0.2 | -0.2 | 0.2 | 0.0 |
| | WT | -0.1 | 0.1 | 0.2 | 0.0 | -0.2 | -0.2 | 0.2 | 0.0 |
| | UN | 30.6 | 30.8 | 30.9 | 30.8 | 30.8 | 30.6 | 31.1 | 30.9 |
| 0.2 | SVY | 0.2 | 0.5 | 0.0 | -0.1 | 0.3 | 0.0 | 0.1 | -0.1 |
| | WT | 0.2 | 0.5 | 0.0 | -0.1 | 0.3 | 0.0 | 0.1 | -0.1 |
| | UN | 20.6 | 20.9 | 20.3 | 20.4 | 20.7 | 20.4 | 20.5 | 20.3 |
| 0.3 | SVY | 0.0 | 0.3 | -0.4 | -0.4 | 0.2 | 0.1 | -0.2 | -0.4 |
| | WT | 0.0 | 0.3 | -0.4 | -0.4 | 0.2 | 0.1 | -0.2 | -0.4 |
| | UN | 14.6 | 15.0 | 14.2 | 14.2 | 14.8 | 14.7 | 14.3 | 14.2 |
| 0.4 | SVY | -0.3 | 0.0 | -0.3 | -0.3 | -0.4 | -0.3 | -0.3 | -0.3 |
| | WT | -0.3 | 0.0 | -0.3 | -0.3 | -0.4 | -0.3 | -0.3 | -0.3 |
| | UN | 10.3 | 10.5 | 10.3 | 10.3 | 10.3 | 10.3 | 10.4 | 10.3 |
| 0.5 | SVY | -0.4 | -0.2 | -0.4 | -0.3 | -0.5 | -0.4 | -0.2 | -0.3 |
| | WT | -0.4 | -0.2 | -0.4 | -0.3 | -0.5 | -0.4 | -0.2 | -0.3 |
| | UN | 7.2 | 7.4 | 7.4 | 7.4 | 7.2 | 7.3 | 7.5 | 7.4 |
| 0.6 | SVY | -0.3 | -0.1 | -0.2 | -0.2 | -0.2 | -0.2 | -0.2 | -0.2 |
| | WT | -0.3 | -0.1 | -0.2 | -0.2 | -0.2 | -0.2 | -0.2 | -0.2 |
| | UN | 5.0 | 5.3 | 5.2 | 5.2 | 5.2 | 5.2 | 5.3 | 5.2 |
| 0.7 | SVY | -0.1 | -0.1 | -0.1 | -0.1 | -0.1 | -0.2 | -0.1 | -0.1 |
| | WT | -0.1 | -0.1 | -0.1 | -0.1 | -0.1 | -0.2 | -0.1 | -0.1 |
| | UN | 3.4 | 3.4 | 3.4 | 3.5 | 3.4 | 3.4 | 3.5 | 3.5 |
| 0.8 | SVY | -0.1 | 0.0 | 0.0 | 0.0 | -0.1 | -0.1 | 0.0 | 0.0 |
| | WT | -0.1 | 0.0 | 0.0 | 0.0 | -0.1 | -0.1 | 0.0 | 0.0 |
| | UN | 2.0 | 2.0 | 2.1 | 2.1 | 2.0 | 2.0 | 2.1 | 2.0 |
| 0.9 | SVY | 0.1 | 0.1 | 0.0 | 0.0 | 0.1 | 0.1 | 0.0 | 0.0 |
| | WT | 0.1 | 0.1 | 0.0 | 0.0 | 0.1 | 0.1 | 0.0 | 0.0 |
| | UN | 0.9 | 1.0 | 0.9 | 0.9 | 1.0 | 1.0 | 0.9 | 0.9 |

Table A.2.3: Estimates of empirical (EMP) and asymptotic standard error of the SVY, WT, and UN estimators for the super-population ROC curve with finite population size $N$, disease proportion $p$, and sampling fraction $\lambda$ under SSRS.

| | | $N = 50{,}000$ | | | | $N = 100{,}000$ | | | |
| | | $p = 5\%$ | | $p = 25\%$ | | $p = 5\%$ | | $p = 25\%$ | |
| FPR | Method | $\lambda = 5\%$ | $\lambda = 10\%$ | $\lambda = 5\%$ | $\lambda = 10\%$ | $\lambda = 5\%$ | $\lambda = 10\%$ | $\lambda = 5\%$ | $\lambda = 10\%$ |
|---|---|---|---|---|---|---|---|---|---|
| 0.1 | EMP | 0.049 | 0.035 | 0.026 | 0.019 | 0.035 | 0.025 | 0.018 | 0.013 |
|     | SVY | 0.051 | 0.036 | 0.026 | 0.019 | 0.036 | 0.025 | 0.019 | 0.013 |
|     | WT  | 0.045 | 0.032 | 0.023 | 0.016 | 0.032 | 0.023 | 0.016 | 0.012 |
|     | UN  | 0.046 | 0.032 | 0.023 | 0.017 | 0.033 | 0.023 | 0.017 | 0.012 |
| 0.2 | EMP | 0.051 | 0.036 | 0.026 | 0.018 | 0.036 | 0.026 | 0.018 | 0.013 |
|     | SVY | 0.051 | 0.036 | 0.026 | 0.018 | 0.036 | 0.026 | 0.018 | 0.013 |
|     | WT  | 0.046 | 0.032 | 0.023 | 0.016 | 0.032 | 0.023 | 0.016 | 0.012 |
|     | UN  | 0.043 | 0.031 | 0.021 | 0.015 | 0.031 | 0.022 | 0.015 | 0.011 |
| 0.3 | EMP | 0.050 | 0.035 | 0.025 | 0.017 | 0.036 | 0.025 | 0.017 | 0.012 |
|     | SVY | 0.049 | 0.035 | 0.024 | 0.017 | 0.035 | 0.025 | 0.017 | 0.012 |
|     | WT  | 0.043 | 0.031 | 0.022 | 0.015 | 0.031 | 0.022 | 0.015 | 0.011 |
|     | UN  | 0.040 | 0.028 | 0.019 | 0.014 | 0.028 | 0.020 | 0.014 | 0.010 |
| 0.4 | EMP | 0.047 | 0.033 | 0.023 | 0.016 | 0.034 | 0.023 | 0.016 | 0.011 |
|     | SVY | 0.045 | 0.032 | 0.022 | 0.016 | 0.032 | 0.023 | 0.016 | 0.011 |
|     | WT  | 0.040 | 0.028 | 0.020 | 0.014 | 0.028 | 0.020 | 0.014 | 0.010 |
|     | UN  | 0.035 | 0.025 | 0.017 | 0.012 | 0.025 | 0.018 | 0.012 | 0.009 |
| 0.5 | EMP | 0.042 | 0.030 | 0.020 | 0.014 | 0.031 | 0.021 | 0.015 | 0.010 |
|     | SVY | 0.040 | 0.028 | 0.020 | 0.014 | 0.028 | 0.020 | 0.014 | 0.010 |
|     | WT  | 0.035 | 0.025 | 0.017 | 0.012 | 0.025 | 0.018 | 0.012 | 0.009 |
|     | UN  | 0.031 | 0.022 | 0.015 | 0.010 | 0.022 | 0.016 | 0.010 | 0.007 |
| 0.6 | EMP | 0.037 | 0.026 | 0.018 | 0.013 | 0.027 | 0.018 | 0.013 | 0.008 |
|     | SVY | 0.034 | 0.024 | 0.017 | 0.012 | 0.024 | 0.017 | 0.012 | 0.008 |
|     | WT  | 0.031 | 0.022 | 0.015 | 0.010 | 0.022 | 0.015 | 0.010 | 0.007 |
|     | UN  | 0.026 | 0.019 | 0.012 | 0.009 | 0.019 | 0.013 | 0.009 | 0.006 |
| 0.7 | EMP | 0.031 | 0.021 | 0.015 | 0.010 | 0.022 | 0.015 | 0.010 | 0.007 |
|     | SVY | 0.028 | 0.020 | 0.014 | 0.010 | 0.020 | 0.014 | 0.010 | 0.007 |
|     | WT  | 0.025 | 0.018 | 0.012 | 0.009 | 0.018 | 0.013 | 0.009 | 0.006 |
|     | UN  | 0.021 | 0.015 | 0.010 | 0.007 | 0.015 | 0.011 | 0.007 | 0.005 |
| 0.8 | EMP | 0.023 | 0.016 | 0.011 | 0.008 | 0.016 | 0.011 | 0.008 | 0.005 |
|     | SVY | 0.021 | 0.015 | 0.010 | 0.007 | 0.015 | 0.011 | 0.007 | 0.005 |
|     | WT  | 0.019 | 0.013 | 0.009 | 0.006 | 0.013 | 0.010 | 0.006 | 0.005 |
|     | UN  | 0.016 | 0.011 | 0.008 | 0.005 | 0.011 | 0.008 | 0.005 | 0.004 |
| 0.9 | EMP | 0.015 | 0.010 | 0.007 | 0.005 | 0.011 | 0.007 | 0.005 | 0.003 |
|     | SVY | 0.015 | 0.010 | 0.006 | 0.005 | 0.010 | 0.007 | 0.005 | 0.003 |
|     | WT  | 0.013 | 0.008 | 0.006 | 0.004 | 0.009 | 0.006 | 0.004 | 0.003 |
|     | UN  | 0.011 | 0.007 | 0.005 | 0.003 | 0.007 | 0.005 | 0.003 | 0.002 |

Table A.2.4: Estimates of empirical (EMP) and asymptotic standard error of the SVY, WT, and UN estimators for the super-population ROC curve with finite population size $N$, disease proportion $p$, and sampling fraction $\lambda$ under STSCS.

| | | $N = 50{,}000$ | | | | $N = 100{,}000$ | | | |
| | | $p = 5\%$ | | $p = 25\%$ | | $p = 5\%$ | | $p = 25\%$ | |
| FPR | Method | $\lambda = 5\%$ | $\lambda = 10\%$ | $\lambda = 5\%$ | $\lambda = 10\%$ | $\lambda = 5\%$ | $\lambda = 10\%$ | $\lambda = 5\%$ | $\lambda = 10\%$ |
|---|---|---|---|---|---|---|---|---|---|
| | EMP | 0.047 | 0.034 | 0.025 | 0.018 | 0.033 | 0.023 | 0.018 | 0.012 |
| | SVY | 0.047 | 0.033 | 0.025 | 0.017 | 0.033 | 0.024 | 0.017 | 0.012 |
| 0.1 | WT | 0.042 | 0.030 | 0.022 | 0.015 | 0.030 | 0.021 | 0.015 | 0.011 |
| | UN | 0.045 | 0.032 | 0.023 | 0.016 | 0.032 | 0.022 | 0.016 | 0.012 |
| | EMP | 0.051 | 0.037 | 0.027 | 0.019 | 0.037 | 0.026 | 0.019 | 0.013 |
| | SVY | 0.052 | 0.036 | 0.026 | 0.019 | 0.036 | 0.026 | 0.019 | 0.013 |
| 0.2 | WT | 0.046 | 0.032 | 0.023 | 0.016 | 0.032 | 0.023 | 0.016 | 0.012 |
| | UN | 0.046 | 0.032 | 0.023 | 0.016 | 0.032 | 0.023 | 0.016 | 0.012 |
| | EMP | 0.052 | 0.038 | 0.027 | 0.019 | 0.037 | 0.026 | 0.019 | 0.013 |
| | SVY | 0.051 | 0.036 | 0.026 | 0.018 | 0.036 | 0.026 | 0.018 | 0.013 |
| 0.3 | WT | 0.046 | 0.032 | 0.023 | 0.016 | 0.032 | 0.023 | 0.016 | 0.011 |
| | UN | 0.044 | 0.031 | 0.022 | 0.015 | 0.031 | 0.022 | 0.015 | 0.011 |
| | EMP | 0.050 | 0.036 | 0.026 | 0.019 | 0.035 | 0.025 | 0.018 | 0.012 |
| | SVY | 0.049 | 0.034 | 0.024 | 0.017 | 0.035 | 0.024 | 0.017 | 0.012 |
| 0.4 | WT | 0.043 | 0.031 | 0.022 | 0.015 | 0.031 | 0.022 | 0.015 | 0.011 |
| | UN | 0.041 | 0.029 | 0.020 | 0.014 | 0.029 | 0.020 | 0.014 | 0.010 |
| | EMP | 0.047 | 0.033 | 0.024 | 0.017 | 0.033 | 0.024 | 0.017 | 0.011 |
| | SVY | 0.045 | 0.032 | 0.022 | 0.016 | 0.032 | 0.022 | 0.016 | 0.011 |
| 0.5 | WT | 0.040 | 0.028 | 0.020 | 0.014 | 0.028 | 0.020 | 0.014 | 0.010 |
| | UN | 0.036 | 0.026 | 0.018 | 0.012 | 0.026 | 0.018 | 0.012 | 0.009 |
| | EMP | 0.042 | 0.030 | 0.021 | 0.015 | 0.029 | 0.021 | 0.015 | 0.010 |
| | SVY | 0.040 | 0.028 | 0.020 | 0.014 | 0.028 | 0.020 | 0.014 | 0.010 |
| 0.6 | WT | 0.035 | 0.025 | 0.017 | 0.012 | 0.025 | 0.018 | 0.012 | 0.009 |
| | UN | 0.032 | 0.022 | 0.015 | 0.011 | 0.022 | 0.016 | 0.011 | 0.008 |
| | EMP | 0.036 | 0.025 | 0.018 | 0.012 | 0.025 | 0.017 | 0.013 | 0.009 |
| | SVY | 0.034 | 0.024 | 0.017 | 0.012 | 0.024 | 0.017 | 0.012 | 0.008 |
| 0.7 | WT | 0.030 | 0.021 | 0.015 | 0.010 | 0.021 | 0.015 | 0.010 | 0.007 |
| | UN | 0.026 | 0.019 | 0.013 | 0.009 | 0.019 | 0.013 | 0.009 | 0.006 |
| | EMP | 0.027 | 0.020 | 0.014 | 0.010 | 0.020 | 0.014 | 0.010 | 0.007 |
| | SVY | 0.026 | 0.019 | 0.013 | 0.009 | 0.019 | 0.013 | 0.009 | 0.006 |
| 0.8 | WT | 0.023 | 0.016 | 0.011 | 0.008 | 0.016 | 0.012 | 0.008 | 0.006 |
| | UN | 0.020 | 0.014 | 0.010 | 0.007 | 0.014 | 0.010 | 0.007 | 0.005 |
| | EMP | 0.018 | 0.013 | 0.009 | 0.006 | 0.013 | 0.009 | 0.006 | 0.005 |
| | SVY | 0.018 | 0.012 | 0.008 | 0.006 | 0.012 | 0.009 | 0.006 | 0.004 |
| 0.9 | WT | 0.016 | 0.011 | 0.007 | 0.005 | 0.011 | 0.008 | 0.005 | 0.004 |
| | UN | 0.013 | 0.009 | 0.006 | 0.004 | 0.009 | 0.006 | 0.004 | 0.003 |

Table A.2.5: Coverage Probabilities (in %) for 95% confidence intervals of the UN, WT, and SVY estimators for the super-population ROC curve with finite population size $N$, disease proportion $p$, and sampling fraction $\lambda$ under SSRS.

| FPR | Method | $N = 50,000$ | | | | $N = 100,000$ | | | |
| | | $p = 5\%$ | | $p = 25\%$ | | $p = 5\%$ | | $p = 25\%$ | |
| | | $\lambda = 5\%$ | $\lambda = 10\%$ | $\lambda = 5\%$ | $\lambda = 10\%$ | $\lambda = 5\%$ | $\lambda = 10\%$ | $\lambda = 5\%$ | $\lambda = 10\%$ |
|---|---|---|---|---|---|---|---|---|---|
| 0.1 | SVY | 95.9 | 95.3 | 95.2 | 94.8 | 95.4 | 95.2 | 95.9 | 95.6 |
| | WT | 92.5 | 92.1 | 92.1 | 91.8 | 92.6 | 92.3 | 92.4 | 92.6 |
| | UN | 22.9 | 4.1 | 0.0 | 0.0 | 2.6 | 0.1 | 0.0 | 0.0 |
| 0.2 | SVY | 95.3 | 94.6 | 94.7 | 95.0 | 95.1 | 95.3 | 95.5 | 95.2 |
| | WT | 92.2 | 91.7 | 92.1 | 91.2 | 91.3 | 91.8 | 91.9 | 92.1 |
| | UN | 29.0 | 6.6 | 0.2 | 0.0 | 5.7 | 0.1 | 0.0 | 0.0 |
| 0.3 | SVY | 94.3 | 94.1 | 94.4 | 95.3 | 93.7 | 94.7 | 94.7 | 95.3 |
| | WT | 90.9 | 90.7 | 91.1 | 92.1 | 90.8 | 90.8 | 91.4 | 92.3 |
| | UN | 40.1 | 12.5 | 0.6 | 0.0 | 13.7 | 0.8 | 0.0 | 0.0 |
| 0.4 | SVY | 94.1 | 94.2 | 94.2 | 94.4 | 92.7 | 93.9 | 94.2 | 95.3 |
| | WT | 90.4 | 90.6 | 90.7 | 91.4 | 89.4 | 90.5 | 91.0 | 91.7 |
| | UN | 48.8 | 22.2 | 2.5 | 0.0 | 20.8 | 2.6 | 0.0 | 0.0 |
| 0.5 | SVY | 92.6 | 93.5 | 94.1 | 94.2 | 93.2 | 94.1 | 93.3 | 94.4 |
| | WT | 89.1 | 89.7 | 91.2 | 90.8 | 88.4 | 91.1 | 89.3 | 91.4 |
| | UN | 57.4 | 34.1 | 5.9 | 0.2 | 34.5 | 10.5 | 0.0 | 0.0 |
| 0.6 | SVY | 92.1 | 92.8 | 93.8 | 93.7 | 92.2 | 94.3 | 92.7 | 94.5 |
| | WT | 88.9 | 89.5 | 90.3 | 90.6 | 88.9 | 91.0 | 89.6 | 91.1 |
| | UN | 64.2 | 45.9 | 15.4 | 1.2 | 47.2 | 19.9 | 1.6 | 0.0 |
| 0.7 | SVY | 90.4 | 93.0 | 92.9 | 93.7 | 92.8 | 94.0 | 93.2 | 93.4 |
| | WT | 86.9 | 89.7 | 88.6 | 90.1 | 88.5 | 90.5 | 89.6 | 89.9 |
| | UN | 70.7 | 59.2 | 33.6 | 7.0 | 59.2 | 39.2 | 7.3 | 0.3 |
| 0.8 | SVY | 91.9 | 91.2 | 92.5 | 94.2 | 92.2 | 93.3 | 93.8 | 93.4 |
| | WT | 89.3 | 87.9 | 89.0 | 90.9 | 88.5 | 89.6 | 89.8 | 89.6 |
| | UN | 78.3 | 71.5 | 50.2 | 25.2 | 72.2 | 59.2 | 25.5 | 4.3 |
| 0.9 | SVY | 98.6 | 86.9 | 91.1 | 92.3 | 88.0 | 89.1 | 92.2 | 93.6 |
| | WT | 97.8 | 84.0 | 87.1 | 88.1 | 84.5 | 85.1 | 88.2 | 89.7 |
| | UN | 99.8 | 81.1 | 67.7 | 54.4 | 82.0 | 71.6 | 53.9 | 29.9 |

Table A.2.6: Coverage Probabilities (in %) for 95% confidence intervals of the UN, WT, and SVY estimators for the super-population ROC curve with finite population size $N$, disease proportion $p$, and sampling fraction $\lambda$ under STSCS.

| FPR | Method | $N = 50{,}000$ | | | | $N = 100{,}000$ | | | |
| | | $p = 5\%$ | | $p = 25\%$ | | $p = 5\%$ | | $p = 25\%$ | |
| | | $\lambda = 5\%$ | $\lambda = 10\%$ | $\lambda = 5\%$ | $\lambda = 10\%$ | $\lambda = 5\%$ | $\lambda = 10\%$ | $\lambda = 5\%$ | $\lambda = 10\%$ |
|---|---|---|---|---|---|---|---|---|---|
| 0.1 | SVY | 94.4 | 94.3 | 94.2 | 93.6 | 94.8 | 95.0 | 94.0 | 94.5 |
| | WT | 91.3 | 90.8 | 91.6 | 89.7 | 91.9 | 92.2 | 91.2 | 91.1 |
| | UN | 50.2 | 20.1 | 3.8 | 0.0 | 20.2 | 2.8 | 0.1 | 0.0 |
| 0.2 | SVY | 94.8 | 95.1 | 94.3 | 93.6 | 95.0 | 95.3 | 94.0 | 94.5 |
| | WT | 91.4 | 91.3 | 91.1 | 90.1 | 91.4 | 92.0 | 90.6 | 90.5 |
| | UN | 49.9 | 19.0 | 4.4 | 0.1 | 20.4 | 2.6 | 0.0 | 0.0 |
| 0.3 | SVY | 94.3 | 94.2 | 93.9 | 93.3 | 95.0 | 95.3 | 93.8 | 94.4 |
| | WT | 90.8 | 90.1 | 90.5 | 89.0 | 90.9 | 91.6 | 90.3 | 90.8 |
| | UN | 53.5 | 23.6 | 5.8 | 0.1 | 25.4 | 4.1 | 0.2 | 0.0 |
| 0.4 | SVY | 93.9 | 94.8 | 94.3 | 92.0 | 94.4 | 94.7 | 93.5 | 94.6 |
| | WT | 90.8 | 91.0 | 90.3 | 88.7 | 90.6 | 91.2 | 89.9 | 91.1 |
| | UN | 59.1 | 33.1 | 10.1 | 0.6 | 34.6 | 9.0 | 0.8 | 0.0 |
| 0.5 | SVY | 93.2 | 93.5 | 93.5 | 92.9 | 93.9 | 93.8 | 93.5 | 94.0 |
| | WT | 89.4 | 89.7 | 90.0 | 88.7 | 90.6 | 89.4 | 90.0 | 90.5 |
| | UN | 65.1 | 42.0 | 16.7 | 1.2 | 44.5 | 17.2 | 1.5 | 0.1 |
| 0.6 | SVY | 92.8 | 93.6 | 94.0 | 93.7 | 94.2 | 93.2 | 92.6 | 94.1 |
| | WT | 89.6 | 90.5 | 90.6 | 90.2 | 90.7 | 89.5 | 88.6 | 90.1 |
| | UN | 69.9 | 50.1 | 24.1 | 3.4 | 51.1 | 24.9 | 4.0 | 0.1 |
| 0.7 | SVY | 92.2 | 93.0 | 93.5 | 93.3 | 92.8 | 94.3 | 92.4 | 94.1 |
| | WT | 88.8 | 89.0 | 89.7 | 90.2 | 89.0 | 90.4 | 88.8 | 90.0 |
| | UN | 72.5 | 59.3 | 36.4 | 11.1 | 59.7 | 38.6 | 10.9 | 0.6 |
| 0.8 | SVY | 91.6 | 92.0 | 91.8 | 92.8 | 92.7 | 92.8 | 93.1 | 93.0 |
| | WT | 88.5 | 88.7 | 87.5 | 89.4 | 89.0 | 89.7 | 89.2 | 88.6 |
| | UN | 77.4 | 68.2 | 48.7 | 22.4 | 68.0 | 54.0 | 23.7 | 4.8 |
| 0.9 | SVY | 95.6 | 88.5 | 91.3 | 92.6 | 88.1 | 89.4 | 91.1 | 91.7 |
| | WT | 91.5 | 85.8 | 87.7 | 88.4 | 85.9 | 86.5 | 87.3 | 87.6 |
| | UN | 81.1 | 73.6 | 65.2 | 47.4 | 72.1 | 63.3 | 46.7 | 24.3 |

# APPENDIX 2: TECHNICAL DETAILS FOR CHAPTER 4

## B.1  Proofs

*Proof of Theorem 4.2.1 (a).* By definition, $U_N(\phi_N) = 0$. Thus,

$$N^{-1/2}U_N^\pi(\phi_N) = N^{-1/2}(U_N^\pi(\phi_N) - U_N(\phi_N))$$

$$= N^{-1/2}\sum_{i=1}^{N} \left(\frac{\xi_i - \pi_i}{\pi_i}\right)\{R_i - p(X_i; \phi_N)\}h_i(\phi_N)$$

$$= N^{-1/2}\sum_{i=1}^{N} \left(\frac{\xi_i - \pi_i}{\pi_i}\right)u_i(\phi_N).$$

From the central limit theorem for normalised Horvitz-Thompson estimator,

$$N^{-1/2}U_N^\pi(\phi_N) \to_{\mathbb{P}_d} \mathrm{N}(0, V(\phi_N)),$$

with

$$V(\phi) = \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N} \frac{\pi_{ij} - \pi_i\pi_j}{\pi_i\pi_j}u_i(\phi)u_j^\top(\phi).$$

By Taylor Expansion,

$$0 = U_N^\pi(\phi_N^\pi) = U_N^\pi(\phi_N) + \left(\frac{\partial U_N^\pi(\phi)}{\partial \phi}\Big|_{\phi=\phi_N}\right)(\phi_N^\pi - \phi_N)$$

$$\Rightarrow \sqrt{N}(\phi_N^\pi - \phi_N) = \left(-N^{-1}\frac{\partial U_N^\pi(\phi)}{\partial \phi}\Big|_{\phi=\phi_N}\right)^{-1}N^{-1/2}U_N^\pi(\phi_N)$$

Similarly as argued in Lin (2000), we have that $N^{-1}U_N^\pi(\phi)$ is asymptotically equivalent to $N^{-1}U_N(\phi)$, since $\pi_i = P(\xi_i = 1)$.

Thus,

$$\sqrt{N}(\phi_N^\pi - \phi_N) \to_{\mathbb{P}_d} I(\phi_N)^{-1}\mathrm{N}(0, V(\phi_N)), \tag{8.26}$$

where $I(\phi_0) = \lim_{N\to\infty} -N^{-1}\partial U_N(\phi)/\partial\phi|_{\phi=\phi_0}$. Combining the consistency of $\phi_N$ from classical results for generalized linear models and Theorem 5.1 (ii) from Rubin-Bleuer et al. (2005), we have that

$$\sqrt{N}(\phi_N^\pi - \phi_N) \to_{\mathbb{P}_{d,m}} I(\phi^*)^{-1} N(0, V(\phi^*)), \qquad (8.27)$$

where $\phi^*$ is the true value for $\phi$ in the model probability space.

$\blacksquare$

*Proof of Theorem 4.2.1 (b).* For $f \in \mathcal{F}$, with $\mathcal{F} = \{f_{s,l} \equiv f_{s,l}(y,d) = I(y \leq s, d = l) : s \in \mathbb{R}, l \in \{0,1\}\}$, and $\mathbb{P}_{N,\phi}^\pi(f)$ and $\mathbb{P}_{N,\phi}(f)$ as defined in (4.12) and (4.14) we have that

$$G_{\text{IPW}}(s) = \frac{\mathbb{P}_{N,\phi_N}(f_{s,0})}{\mathbb{P}_{N,\phi_N}(f_{\infty,0})} = \left[\sum_{i=1}^{N} \frac{R_i}{p(X_i;\phi_N)} I(D_i = 0)\right]^{-1} \sum_{i=1}^{N} \frac{R_i}{p(X_i;\phi_N)} I(Y_i \leq s, D_i = 0)$$

$$F_{\text{IPW}}(s) = \frac{\mathbb{P}_{N,\phi_N}(f_{s,1})}{\mathbb{P}_{N,\phi_N}(f_{\infty,1})} = \left[\sum_{i=1}^{N} \frac{R_i}{p(X_i;\phi_N)} I(D_i = 1)\right]^{-1} \sum_{i=1}^{N} \frac{R_i}{p(X_i;\phi_N)} I(Y_i \leq s, D_i = 1)$$

$$G_{\text{IPW}}^\pi(s) = \frac{\mathbb{P}_{N,\phi_N^\pi}^\pi(f_{s,0})}{\mathbb{P}_{N,\phi_N^\pi}^\pi(f_{\infty,0})} = \left[\sum_{i=1}^{N} \frac{\xi_i}{\pi_i} \frac{R_i}{p(X_i;\phi_N^\pi)} I(D_i = 0)\right]^{-1} \sum_{i=1}^{N} \frac{\xi_i}{\pi_i} \frac{R_i}{p(X_i;\phi_N^\pi)} I(Y_i \leq s, D_i = 0)$$

$$F_{\text{IPW}}^\pi(s) = \frac{\mathbb{P}_{N,\phi_N^\pi}^\pi(f_{s,1})}{\mathbb{P}_{N,\phi_N^\pi}^\pi(f_{\infty,1})} = \left[\sum_{i=1}^{N} \frac{\xi_i}{\pi_i} \frac{R_i}{p(X_i;\phi_N^\pi)} I(D_i = 1)\right]^{-1} \sum_{i=1}^{N} \frac{\xi_i}{\pi_i} \frac{R_i}{p(X_i;\phi_N^\pi)} I(Y_i \leq s, D_i = 1)$$

We are interested in the asymptotic distribution of

$$\sqrt{N} \begin{bmatrix} G_{\text{IPW}}^\pi - G_{\text{IPW}} \\ F_{\text{IPW}}^\pi - F_{\text{IPW}} \end{bmatrix} = \sqrt{N} \begin{bmatrix} \frac{\mathbb{P}_{N,\phi_N^\pi}^\pi(f_{s,0})}{\mathbb{P}_{N,\phi_N^\pi}^\pi(f_{\infty,0})} - \frac{\mathbb{P}_{N,\phi_N}(f_{s,0})}{\mathbb{P}_{N,\phi_N}(f_{\infty,0})} \\ \frac{\mathbb{P}_{N,\phi_N^\pi}^\pi(f_{s,1})}{\mathbb{P}_{N,\phi_N^\pi}^\pi(f_{\infty,1})} - \frac{\mathbb{P}_{N,\phi_N}(f_{s,1})}{\mathbb{P}_{N,\phi_N}(f_{\infty,1})} \end{bmatrix}$$

First, note that

$$\sqrt{N}(\mathbb{P}_{N,\phi_N^\pi}^\pi - \mathbb{P}_{N,\phi_N})(f) = \sqrt{N}(\mathbb{P}_{N,\phi_N^\pi}^\pi - \mathbb{P}_{N,\phi_N}^\pi)(f) + \sqrt{N}(\mathbb{P}_{N,\phi_N}^\pi - \mathbb{P}_{N,\phi_N})(f) \quad (8.28)$$

Let $p_i(\phi) = p(X_i; \phi)$, $w_i(\phi) = p_i(\phi)^{-1}$ and $\partial_\phi p_i(\phi_0) = \partial p_i(\phi)/\partial \phi|_{\phi=\phi_0}$. For the first term of the right hand side of (8.28), we have by Taylor expansion:

$$
\sqrt{N}(\mathbb{P}^\pi_{N,\phi^\pi_N} - \mathbb{P}^\pi_{N,\phi_N})(f_{s,l}) =
$$

$$
= N^{-1/2} \sum_{i=1}^N \frac{\xi_i}{\pi_i}[w_i(\phi^\pi_N) - w_i(\phi_N)]R_i f_{s,l}(Y_i, D_i)
$$

$$
= N^{-1/2} \sum_{i=1}^N \frac{\xi_i}{\pi_i}\left(\frac{\partial w_i(\phi)}{\partial \phi}\Big|_{\phi=\phi_N}\right)(\phi^\pi_N - \phi_N)R_i f_{s,l}(Y_i, D_i)
$$

$$
= \left(-N^{-1}\sum_{i=1}^N \frac{\xi_i}{\pi_i}\frac{R_i}{p_i(\phi_N)^2}\partial_\phi p_i(\phi_N)f_{s,l}(Y_i, D_i)\right)\sqrt{N}(\phi^\pi_N - \phi_N)
$$

$$
= \left(-N^{-1}\sum_{i=1}^N \frac{\xi_i}{\pi_i}\frac{R_i}{p_i(\phi_N)^2}\partial_\phi p_i(\phi_N)f_{s,l}(Y_i, D_i)\right)\left(-N^{-1}\frac{\partial U^\pi_N(\phi)}{\partial \phi}\Big|_{\phi=\phi_N}\right)^{-1}N^{-1/2}U^\pi_N(\phi_N)
$$

$$(8.29)$$

Similarly as argued in Lin (2000), because $\pi_i = P(\xi_i = 1)$, we have that (8.29) is asymptotically equivalent to

$$
N^{-1/2} \sum_{i=1}^N \left(\frac{\xi_i - \pi_i}{\pi_i}\right) r_{s,l}(\phi_N)I(\phi_N)^{-1}u_i(\phi_N), \tag{8.30}
$$

where

$$
r_{s,l}(\phi_N) = \lim_{N\to\infty} -N^{-1}\sum_{i=1}^N \frac{R_i}{p_i(\phi_N)^2}\partial_\phi p_i(\phi_N)f_{s,l}(Y_i, D_i)
$$

For the second term of the right hand side of (8.28), we have:

$$
\sqrt{N}(\mathbb{P}^\pi_{N,\phi_N} - \mathbb{P}_{N,\phi_N}) = N^{-1/2}\sum_{i=1}^N \left(\frac{\xi_i - \pi_i}{\pi_i}\right)\frac{R_i}{p_i(\phi_N)}f_{s,l}(Y_i, D_i) \tag{8.31}
$$

Combining (8.30) and (8.31), we have:

$$\sqrt{N}(\mathbb{P}^{\pi}_{N,\phi^{\pi}_N} - \mathbb{P}_{N,\phi_N})(f_{s,l}) =$$

$$= N^{-1/2} \sum_{i=1}^{N} \left( \frac{\xi_i - \pi_i}{\pi_i} \right) \left\{ \frac{R_i}{p_i(\phi_N)} f_{s,l}(Y_i, D_i) + r_{s,l}(\phi_N) I^{-1}(\phi_N) u_i(\phi_N) \right\} + o_P(1)$$

Therefore, $\sqrt{N}(\mathbb{P}^{\pi}_{N,\phi^{\pi}_N} - \mathbb{P}_{N,\phi_N})(f_{s,l})$ converges weakly in the design probability space to a zero-mean Gaussian process $\mathbb{G}^{\pi}$ with covariance function

$$\lim_{N \to \infty} N^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} h_i(s, d; \phi_N) h_j^{\top}(u, d'; \phi_N),$$

with $h_i(s, l; \phi_N) = R_i p_i(\phi_N)^{-1} f_{s,l}(Y_i, D_i) + r_{s,l}(\phi_N) I^{-1}(\phi_N) u_i(\phi_N)$.

From Theorem 5.1 (ii) from Rubin-Bleuer et al. (2005), we have that $\sqrt{N}(\mathbb{P}^{\pi}_{N,\phi^{\pi}_N} - \mathbb{P}_{N,\phi_N})(f_{s,l})$ converges in the product probability space to a zero mean gaussian process $\mathbb{G}^{\pi}$ with covariance function

$$\text{Cov}(\mathbb{G}^{\pi}(f_{s,d}), \mathbb{G}^{\pi}(f_{u,d'})) = \mathbb{E}_m \mathbb{G}^{\pi}(f_{s,d}) \mathbb{G}^{\pi}(f_{u,d'})$$

$$= \lim_{N \to \infty} N^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} \mathbb{E}_m \left[ \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} h_i(s, d; \phi_N) h_j^{\top}(u, d'; \phi_N) \right].$$

(8.32)

Let $h_{s,l} = Rp(\phi)^{-1} f_{s,l}(Y, D) + r_{s,l}(\phi) I^{-1}(\phi) u(\phi)$, $\delta = P(D = 1)$, and $\partial_{\phi} p = \partial p(X, \phi)/\partial \phi$. To further simplify the covariance function (8.32), we use similar steps as in Han and

Wellner (2021), following the result from Lemma B.1 in Boistard et al. (2017).

$$\mathrm{Cov}(\mathbb{G}^{\pi}(f_{s,d}), \mathbb{G}^{\pi}(f_{u,d'})) = \mu_{\pi_1} P(h_{s,d} h_{u,d'}) + \mu_{\pi_2}(P h_{s,d})(P h_{u,d'})$$

$$= \begin{cases} \mu_{\pi_1}(1-\delta)\left[p(\phi)^{-1}G(s \wedge u) + (1-\delta)G(s)G(u)\partial_{\phi}pI^{-1}(\phi)\partial_{\phi}p^{\top}\right] + \\ \quad + \mu_{\pi_2}(1-\delta)^2 G(u)G(s), \quad d = d' = 0 \\ \mu_{\pi_1}\delta\left[p(\phi)^{-1}F(s \wedge u) + \delta F(s)F(u)\partial_{\phi}pI^{-1}(\phi)\partial_{\phi}p^{\top}\right] + \mu_{\pi_2}\delta^2 F(u)F(s), \quad d = d' = 1 \\ \delta(1-\delta)G(s)F(u)\left[\mu_{\pi_1}\partial_{\phi}pI^{-1}(\phi)\partial_{\phi}p^{\top} + \mu_{\pi_2}\right], \quad 0 = d \neq d' = 1 \\ \delta(1-\delta)G(u)F(s)\left[\mu_{\pi_1}\partial_{\phi}pI^{-1}(\phi)\partial_{\phi}p^{\top} + \mu_{\pi_2}\right], \quad 1 = d \neq d' = 0 \end{cases}$$

The ratio map $\varphi(A, B) = A/B$ is Hadamard-differentiable with derivative $\varphi'(\alpha, \beta) = \alpha/B - (A\beta)/B^2$. It follows from Functional Delta Method (Vaart and Wellner, 1996) that

$$\sqrt{N}\begin{bmatrix} G_{\mathrm{IPW}}^{\pi} - G_{\mathrm{IPW}} \\ F_{\mathrm{IPW}}^{\pi} - F_{\mathrm{IPW}} \end{bmatrix} \rightsquigarrow \begin{bmatrix} \mathbb{G}_0^{\pi} \\ \mathbb{G}_1^{\pi} \end{bmatrix} = \begin{bmatrix} (1-\delta)^{-1}\left(\mathbb{G}^{\pi}(f_{s,0}) - G(s)\mathbb{G}^{\pi}(f_{\infty,0})\right) \\ \delta^{-1}\left(\mathbb{G}^{\pi}(f_{s,1}) - F(s)\mathbb{G}^{\pi}(f_{\infty,1})\right) \end{bmatrix},$$

The covariance structure of $\mathbb{G}_0^\pi$ and $\mathbb{G}_1^\pi$ can be computed as follows:

$$\mathrm{Cov}(\mathbb{G}_0^\pi(s), \mathbb{G}_0^\pi(u))$$

$$= (1-\delta)^{-2}[\mathrm{Cov}(\mathbb{G}^\pi f_{s,0}, \mathbb{G}^\pi f_{u,0})$$

$$+ G(s)G(u)\mathrm{Cov}(\mathbb{G}^\pi f_{\infty,0}, \mathbb{G}^\pi f_{\infty,0})$$

$$- G(u)\mathrm{Cov}(\mathbb{G}^\pi f_{s,0}, \mathbb{G}^\pi f_{\infty,0})$$

$$- G(s)\mathrm{Cov}(\mathbb{G}^\pi f_{\infty,0}, \mathbb{G}^\pi f_{u,0})]$$

$$= (1-\delta)^{-2}\left\{\mu_{\pi_1}(1-\delta)\left[\frac{G(s\wedge u)}{p(\phi)} + (1-\delta)G(s)G(u)\partial_\phi p I^{-1}(\phi)\partial_\phi p^\top + (1-\delta)G(u)G(s)\right]\right.$$

$$+ G(s)G(u)\left[\mu_{\pi_1}(1-\delta)\left[\frac{1}{p(\phi)} + (1-\delta)\partial_\phi p I^{-1}(\phi)\partial_\phi p^\top\right] + \mu_{\pi_2}(1-\delta)^2\right]$$

$$- G(u)\left[\mu_{\pi_1}(1-\delta)\left[\frac{G(s)}{p(\phi)} + (1-\delta)G(s)\partial_\phi p I^{-1}(\phi)\partial_\phi p^\top\right] + \mu_{\pi_2}(1-\delta)^2 G(s)\right]$$

$$\left.- G(s)\left[\mu_{\pi_1}(1-\delta)\left[\frac{G(u)}{p(\phi)} + (1-\delta)G(u)\partial_\phi p I^{-1}(\phi)\partial_\phi p^\top\right] + \mu_{\pi_2}(1-\delta)^2 G(u)\right]\right\}$$

$$= \frac{\mu_{\pi_1}}{(1-\delta)p(\phi)}[G(s\wedge u) - G(u)G(s)]$$

$$\mathrm{Cov}(\mathbb{G}_1^\pi(s), \mathbb{G}_1^\pi(u)) = \delta^{-2}[\mathrm{Cov}(\mathbb{G}^\pi f_{s,1}, \mathbb{G}^\pi f_{u,1})$$

$$+ F(s)F(u)\mathrm{Cov}(\mathbb{G}^\pi f_{\infty,1}, \mathbb{G}^\pi f_{\infty,1})$$

$$- F(u)\mathrm{Cov}(\mathbb{G}^\pi f_{s,1}, \mathbb{G}^\pi f_{\infty,1})$$

$$- F(s)\mathrm{Cov}(\mathbb{G}^\pi f_{\infty,1}, \mathbb{G}^\pi f_{u,1})]$$

$$= \delta^{-2}\left\{ \mu_{\pi_1}\delta\left[\frac{F(s \wedge u)}{p(\phi)} + \delta F(s)F(u)\partial_\phi p I^{-1}(\phi)\partial_\phi p^\top\right] + \mu_{\pi_2}\delta^2 F(u)F(s)\right.$$

$$+ F(s)F(u)\left[\mu_{\pi_1}\delta\left[\frac{1}{p(\phi)} + \delta\partial_\phi p I^{-1}(\phi)\partial_\phi p^\top\right] + \mu_{\pi_2}\delta^2\right]$$

$$- F(u)\left[\mu_{\pi_1}\delta\left[\frac{F(s)}{p(\phi)} + \delta F(s)\partial_\phi p I^{-1}(\phi)\partial_\phi p^\top\right] + \mu_{\pi_2}\delta^2 F(s)\right]$$

$$\left. - F(s)\left[\mu_{\pi_1}\delta\left[\frac{F(u)}{p(\phi)} + \delta F(u)\partial_\phi p I^{-1}(\phi)\partial_\phi p^\top\right] + \mu_{\pi_2}\delta^2 F(u)\right]\right\}$$

$$= \frac{\mu_{\pi_1}}{\delta p(\phi)}\left[F(s \wedge u) - F(u)F(s)\right]$$

$$\mathrm{Cov}(\mathbb{G}_0^\pi(s), \mathbb{G}_1^\pi(u)) = \delta^{-1}(1-\delta)^{-1}[\mathrm{Cov}(\mathbb{G}^\pi f_{s,0}, \mathbb{G}^\pi f_{u,1})$$

$$+ G(s)F(u)\mathrm{Cov}(\mathbb{G}^\pi f_{\infty,0}, \mathbb{G}^\pi f_{\infty,1})$$

$$- F(u)\mathrm{Cov}(\mathbb{G}^\pi f_{s,0}, \mathbb{G}^\pi f_{\infty,1})$$

$$- G(s)\mathrm{Cov}(\mathbb{G}^\pi f_{\infty,0}, \mathbb{G}^\pi f_{u,1})]$$

$$= \mu_{\pi_1}G(s)F(u)\partial_\phi p I^{-1}(\phi)\partial_\phi p^\top + \mu_{\pi_2}G(s)F(u)$$

$$+ G(s)F(u)\left[\mu_{\pi_1}\partial_\phi p I^{-1}(\phi)\partial_\phi p^\top + \mu_{\pi_2}\right]$$

$$- F(u)\left[\mu_{\pi_1}G(s)\partial_\phi p I^{-1}(\phi)\partial_\phi p^\top + \mu_{\pi_2}G(s)\right]$$

$$- G(s)\left[\mu_{\pi_1}F(u)\partial_\phi p I^{-1}(\phi)\partial_\phi p^\top + \mu_{\pi_2}F(u)\right]$$

$$= 0$$

So, the covariance function is given by

$$\text{Cov}(\mathbb{G}_d^\pi(s), \mathbb{G}_{d'}^\pi(u)) =$$

$$= \begin{cases} \mu_{\pi_1}(1-\delta)^{-1}p(\phi^*)^{-1}\left[G(s \wedge u) - G(u)G(s)\right], & d = d' = 0 \\ \mu_{\pi_1}\delta^{-1}p(\phi^*)^{-1}\left[F(s \wedge u) - F(u)F(s)\right], & d = d' = 1 \\ 0, & d \neq d' \end{cases}$$

■

*Proof of Theorem 4.2.1 (c).* The ROC estimator depends on the pair $(G_{\text{IPW}}^\pi, F_{\text{IPW}}^\pi)$ through the map $\psi(A, B) = B(A^{-1})$, where $A^{-1}$ is the inverse map of $A$. The map $\psi(G, F)$ is Hadamard-differentiable (Lemma 12.2 and Lemma 12.7 from Kosorok (2008)) with derivative

$$\psi'(\alpha, \beta) = \beta(G^{-1}) - \frac{f(G^{-1})}{g(G^{-1})}\alpha(G^{-1})$$

It follows from Functional Delta Method that

$$\sqrt{n}(F_{\text{IPW}}^\pi \circ (G_{\text{IPW}}^\pi)^{-1} - F_{\text{IPW}} \circ (G_{\text{IPW}})^{-1})$$

$$\rightsquigarrow \sqrt{\frac{\lambda\mu_{\pi_1}}{p(\phi^*)}}\left\{\delta^{-1/2}B_1(F \circ G^{-1}) - (1-\delta)^{-1/2}\frac{f(G^{-1})}{g(G^{-1})}B_2\right\}$$

where $B_1$ and $B_2$ are two independent Brownian bridges.

■

# BIBLIOGRAPHY

Alonzo, T. A. and Pepe, M. S. (2002). Distribution-free roc analysis using binary regression techniques. *Biostatistics*, 3(3):421–432.

Alonzo, T. A. and Pepe, M. S. (2005). Assessing accuracy of a continuous screening test in the presence of verification bias. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(1):173–190.

American Diabetes Association (2011). Standards of medical care in diabetes—2012. *Diabetes care*, 35(Supplement_1):S11–S63.

American Diabetes Association (2022). 16. diabetes care in the hospital: Standards of medical care in diabetes—2022. *Diabetes Care*, 45(Supplement_1):S244–S253.

Bang, H., Edwards, A. M., Bomback, A. S., Ballantyne, C. M., Brillon, D., Callahan, M. A., Teutsch, S. M., Mushlin, A. I., and Kern, L. M. (2009). A patient self-assessment diabetes screening score:: development, validation, and comparison to other diabetes risk assessment scores. *Annals of internal medicine*, 151(11):775.

Beaumont, J.-F. and Patak, Z. (2012). On the generalized bootstrap for sample surveys with special attention to poisson sampling. *International Statistical Review*, 80(1):127–148.

Begg, C. B. and Greenes, R. A. (1983). Assessment of diagnostic tests when disease verification is subject to selection bias. *Biometrics*, pages 207–215.

Binder, D. A. (1983). On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review/Revue Internationale de Statistique*, pages 279–292.

Bisoffi, G., Mazzi, M. A., and Dunn, G. (2000). Evaluating screening questionnaires using receiver operating characteristic (roc) curves from two-phase (double) samples. *International Journal of Methods in Psychiatric Research*, 9(3):121–133.

Boistard, H., Lopuhaä, H. P., Ruiz-Gazen, A., et al. (2017). Functional central limit theorems for single-stage sampling designs. *The Annals of Statistics*, 45(4):1728–1758.

Booth, J. G., Butler, R. W., and Hall, P. (1994). Bootstrap methods for finite populations. *Journal of the American Statistical Association*, 89(428):1282–1289.

Breidt, F. J. and Opsomer, J. D. (2009). Nonparametric and semiparametric estimation in complex surveys. In *Handbook of statistics*, volume 29, pages 103–119. Elsevier.

Brick, J. M. and Montaquila, J. M. (2009). Nonresponse and weighting. In *Handbook of statistics*, volume 29, pages 163–185. Elsevier.

Cai, T. (2004). Semi-parametric roc regression analysis with placement values. *Biostatistics*, 5(1):45–60.

Cai, T. and Pepe, M. S. (2002). Semiparametric receiver operating characteristic analysis to evaluate biomarkers for disease. *Journal of the American statistical Association*, 97(460):1099–1107.

Canty, A. J. and Davison, A. C. (1999). Resampling-based variance estimation for labour force surveys. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 48(3):379–391.

DeBoer, M. D. and Gurka, M. J. (2014). Low sensitivity of the metabolic syndrome to identify adolescents with impaired glucose tolerance: an analysis of nhanes 1999–2010. *Cardiovascular diabetology*, 13(1):1–8.

Draznin, B., Aroda, V. R., Bakris, G., Benson, G., Brown, F. M., Freeman, R., Green, J., Huang, E., Isaacs, D., Kahan, S., et al. (2022). 2. classification and diagnosis of diabetes: Standards of medical care in diabetes-2022. *Diabetes Care*, 45(Supplement_1):S17–S38.

Efron, B. (1992). Bootstrap methods: another look at the jackknife. In *Breakthroughs in statistics*, pages 569–593. Springer.

Faraggi, D. (2003). Adjusting receiver operating characteristic curves and related indices for covariates. *Journal of the Royal Statistical Society: Series D (the Statistician)*, 52(2):179–192.

Ferraro, P. M., Costanzi, S., Naticchia, A., Sturniolo, A., and Gambaro, G. (2010). Low level exposure to cadmium increases the risk of chronic kidney disease: analysis of the nhanes 1999-2006. *BMC public health*, 10(1):1–8.

Fluss, R., Reiser, B., Faraggi, D., and Rotnitzky, A. (2009). Estimation of the roc curve under verification bias. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 51(3):475–490.

Gaziano, T. A., Young, C. R., Fitzmaurice, G., Atwood, S., and Gaziano, J. M. (2008). Laboratory-based versus non-laboratory-based method for assessment of cardiovascular disease risk: the nhanes i follow-up study cohort. *The Lancet*, 371(9616):923–931.

González-Manteiga, W., Pardo-Fernández, J. C., and Keilegom, I. v. (2011). Roc curves in non-parametric location-scale regression models. *Scandinavian Journal of Statistics*, 38(1):169–184.

Gray, R., Begg, C. B., and Greenes, R. A. (1984). Construction of receiver operating characteristic curves when disease verification is subject to selection bias. *Medical Decision Making*, 4(2):151–164.

Gross, S. (1980). Median estimation in sample surveys. In *Proceedings of the Section on Survey Research Methods*, volume 1814184. American Statistical Association Alexandria, VA.

Han, Q. and Wellner, J. A. (2021). Complex sampling designs: Uniform limit theorems and applications. *The Annals of Statistics*, 49(1):459–485.

Heeringa, S. G., West, B. T., and Berglund, P. A. (2017). *Applied survey data analysis*. chapman and hall/CRC.

Higgins, T. N., Tran, D., Cembrowski, G. S., Shalapay, C., Steele, P., and Wiley, C. (2011). Is hba1c a good screening test for diabetes mellitus? *Clinical biochemistry*, 44(17-18):1469–1472.

Hsieh, F., Turnbull, B. W., et al. (1996). Nonparametric and semiparametric estimation of the receiver operating characteristic curve. *The annals of statistics*, 24(1):25–40.

Inácio, V., Rodríguez-Álvarez, M. X., and Gayoso-Diz, P. (2021). Statistical evaluation of medical tests. *Annual Review of Statistics and Its Application*, 8:41–67.

Karakaya, J., Karabulut, E., and Yucel, R. M. (2015). Sensitivity to imputation models and assumptions in receiver operating characteristic analysis with incomplete data. *Journal of statistical computation and simulation*, 85(17):3498–3511.

Kim, J. K. and Riddles, M. K. (2012). Some theory for propensity-score-adjustment estimators in survey sampling. *Survey Methodology*, 38(2):157.

Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference*. Springer.

Laurson, K. R., Eisenmann, J. C., and Welk, G. J. (2011). Development of youth percent body fat standards using receiver operating characteristic curves. *American journal of preventive medicine*, 41(4):S93–S99.

Li, P., Taylor, J. M., Spratt, D. E., Karnes, R. J., and Schipper, M. J. (2021). Evaluation of predictive model performance of an existing model in the presence of missing data. *Statistics in Medicine*.

Li, S. and Ning, Y. (2015). Estimation of covariate-specific time-dependent roc curves in the presence of missing biomarkers. *Biometrics*, 71(3):666–676.

Lin, D. (2000). On fitting cox's proportional hazards models to survey data. *Biometrika*, 87(1):37–47.

Liu, D. and Zhou, X.-H. (2010). A model for adjusting for nonignorable verification bias in estimation of the roc curve and its area with likelihood-based approach. *Biometrics*, 66(4):1119–1128.

Liu, X. and Zhao, Y. (2012). Semi-empirical likelihood inference for the roc curve with missing data. *Journal of Statistical Planning and Inference*, 142(12):3123–3133.

Lohr, S. L. (2019). *Sampling: design and analysis*. Chapman and Hall/CRC.

Long, Q., Zhang, X., and Hsu, C.-H. (2011a). Nonparametric multiple imputation for receiver operating characteristics analysis when some biomarker values are missing at random. *Statistics in medicine*, 30(26):3149–3161.

Long, Q., Zhang, X., and Johnson, B. A. (2011b). Robust estimation of area under roc curve using auxiliary variables in the presence of missing biomarker values. *Biometrics*, 67(2):559–567.

Mashreghi, Z., Haziza, D., and Léger, C. (2016). A survey of bootstrap methods in finite population sampling. *Statistics Surveys*, 10:1–52.

Metz, C. E., Herman, B. A., and Shen, J.-H. (1998). Maximum likelihood estimation of receiver operating characteristic (roc) curves from continuously-distributed data. *Statistics in medicine*, 17(9):1033–1053.

Molenberghs, G., Fitzmaurice, G., Kenward, M. G., Tsiatis, A., and Verbeke, G. (2014). *Handbook of missing data methodology*. CRC Press.

Olson, D. E., Rhee, M. K., Herrick, K., Ziemer, D. C., Twombly, J. G., and Phillips, L. S. (2010). Screening for diabetes and pre-diabetes with proposed a1c-based diagnostic criteria. *Diabetes care*, 33(10):2184–2189.

Pandya, A., Weinstein, M. C., and Gaziano, T. A. (2011). A comparative assessment of non-laboratory-based versus commonly used laboratory-based cardiovascular disease risk scores in the nhanes iii population. *PloS one*, 6(5):e20416.

Pepe, M., Leisenring, W., and Rutter, C. (2000). 12 evaluating diagnostic tests in public health. *Handbook of statistics*, 18:397–422.

Pepe, M. S. (1997). A regression modelling framework for receiver operating characteristic curves in medical diagnostic testing. *Biometrika*, 84(3):595–608.

Pepe, M. S. (1998). Three approaches to regression analysis of receiver operating characteristic curves for continuous test results. *Biometrics*, pages 124–135.

Pepe, M. S. (2000a). An interpretation for the roc curve and inference using glm procedures. *Biometrics*, 56(2):352–359.

Pepe, M. S. (2000b). Receiver operating characteristic methodology. *Journal of the American Statistical Association*, 95(449):308–311.

Pfeffermann, D. (2000). *Handbook of Statistics_29B: Sample Surveys: Inference and Analysis*, volume 29. Elsevier.

Pfeffermann, D. and Rao, C. R. (2009). *Sample surveys: design, methods and applications*. Elsevier.

Qin, G. and Wang, B. (2012). Imputation-based empirical likelihood inference for the area under the roc curve with missing data. *Statistics and Its Interface*, 5(3):319–329.

Rao, J., Wu, C., and Yue, K. (1992). Some recent work on resampling methods for complex surveys. *Survey methodology*, 18(2):209–217.

Rao, J. N. and Wu, C. (1988). Resampling inference with complex survey data. *Journal of the american statistical association*, 83(401):231–241.

Rao, J. N. K. and Wu, C. (1984). Boostrap inference for sample surveys. *JSM Proceedings, Survey Research Methods Section*, pages 106–112.

Rodenberg, C. and Zhou, X.-H. (2000). Roc curve estimation when covariates affect the verification process. *Biometrics*, 56(4):1256–1262.

Rodríguez-Álvarez, M. X., Roca-Pardiñas, J., and Cadarso-Suárez, C. (2011). Roc curve and covariates: extending induced methodology to the non-parametric framework. *Statistics and Computing*, 21(4):483–499.

Rotnitzky, A., Faraggi, D., and Schisterman, E. (2006). Doubly robust estimation of the area under the receiver-operating characteristic curve in the presence of verification bias. *Journal of the American Statistical Association*, 101(475):1276–1288.

Rubin-Bleuer, S., Kratina, I. S., et al. (2005). On the two-phase framework for joint model and design-based inference. *The Annals of Statistics*, 33(6):2789–2810.

Särndal, C.-E. and Lundström, S. (2005). *Estimation in surveys with nonresponse.* John Wiley & Sons.

Sitter, R. R. (1992). A resampling procedure for complex survey data. *Journal of the American Statistical Association*, 87(419):755–765.

Toledano, A. Y. and Gatsonis, C. (1996). Ordinal regression methodology for roc curves derived from correlated data. *Statistics in Medicine*, 15(16):1807–1826.

U.S. Department of Health and Human Services (2008). 2008 Physical Activity Guidelines for Americans. https://health.gov/sites/default/files/2019-09/paguide.pdf. [Accessed: 2022-04-12].

Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes: with applications to statistics.* Springer.

Yang, H. and Zhao, Y. (2015). Smoothed jackknife empirical likelihood inference for roc curves with missing data. *Journal of Multivariate Analysis*, 140:123–138.

Yao, F., Craiu, R. V., and Reiser, B. (2010). Nonparametric covariate adjustment for receiver operating characteristic curves. *Canadian Journal of Statistics*, 38(1):27–46.

Yao, W., Li, Z., and Graubard, B. I. (2015). Estimation of roc curve with complex survey data. *Statistics in medicine*, 34(8):1293–1303.

Zhang, L., Zhang, Z., Zhang, Y., Hu, G., and Chen, L. (2014). Evaluation of finnish diabetes risk score in screening undiagnosed diabetes and prediabetes among us adults by gender and race: Nhanes 1999-2010. *PloS one*, 9(5):e97865.

Zhou, X. H. (1996). A nonparametric maximum likelihood estimator for the receiver operating characteristic curve area in the presence of verification bias. *Biometrics*, pages 299–305.

Zhou, X.-H. (1998). Correcting for verification bias in studies of a diagnostic test's accuracy. *Statistical methods in medical research*, 7(4):337–353.

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

Zou, K. H. and Hall, W. (2000). Two transformation models for estimating an roc curve derived from continuous data. *Journal of Applied Statistics*, 27(5):621–631.

Zou, K. H., Hall, W., and Shapiro, D. E. (1997). Smooth non-parametric receiver operating characteristic (roc) curves for continuous diagnostic tests. *Statistics in medicine*, 16(19):2143–2156.

Zou, K. H., Tempany, C. M., Fielding, J. R., and Silverman, S. G. (1998). Original smooth receiver operating characteristic curve estimation from continuous data: statistical methods for analyzing the predictive value of spiral ct of ureteral stones. *Academic radiology*, 5(10):680–687.