ORIGINAL PAPER

Estimation of the volume under the ROC surface in presence of nonignorable verification bias

Khanh To Duc[1] · Monica Chiogna[2] · Gianfranco Adimari[1] · for the Alzheimer's Disease Neuroimaging Initiative

Abstract

The volume under the receiver operating characteristic surface (VUS) is useful for measuring the overall accuracy of a diagnostic test when the possible disease status belongs to one of three ordered categories. In medical studies, the VUS of a new test is typically estimated through a sample of measurements obtained by some suitable sample of patients. However, in many cases, only a subset of such patients has the true disease status assessed by a gold standard test. In this paper, for a continuous-scale diagnostic test, we propose four estimators of the VUS which accommodate for nonignorable missingness of the disease status. The estimators are based on a parametric model which jointly describes both the disease and the verification process. Identifiability of the model is discussed. Consistency and asymptotic normality of the proposed estimators are shown, and variance estimation is discussed. The finite-sample behavior is investigated by means of simulation experiments. An illustration is provided.

Keywords Diagnostic test · Nonignorable missing data mechanism · ROC analysis

---

✉ Monica Chiogna
monica.chiogna2@unibo.it

[1] Department of Statistical Sciences, University of Padova, Via Cesare Battisti, 241, 35121 Padova, Italy

[2] Department of Statistical Sciences "Paolo Fortunati", University of Bologna, Via Belle Arti, 41, 40126 Bologna, Italy

1 Introduction

For an ordinal three-category classification problem, the assessment of the performance of a diagnostic test is achieved by the analysis of the receiver operating characteristic (ROC) surface, which generalizes the ROC curve for binary diagnostic outcomes. The volume under the ROC surface (VUS) is a summary index, usually employed for measuring the overall diagnostic accuracy of the test. Under correct ordering, values of VUS vary from 1/6, suggesting the test is no better than chance alone, to 1, which implies a perfect test, i.e. a test that perfectly discriminates among the three categories. The theoretical construction of the ROC surface and VUS was introduced for the first time by Scurfield (1996).

In medical studies, the evaluation of the discriminatory ability of a diagnostic test is typically obtained by making inference about its ROC surface and VUS, based on data from some suitable sample of patients (or units). When the disease status of each patient can be exactly assessed by means of a gold standard (GS) test, a set of methods exist to estimate the ROC surface and VUS of the test in evaluation. See Nakas and Yiannoutsos (2004), Xiong et al. (2006), Li and Zhou (2009) and Kang and Tian (2013), among others. In practice, however, disease status verification via GS test could be unavailable for all units in the sample, due to the expensiveness and/or invasiveness of the GS test. Thus, often, only a subset of patients undergoes disease verification. In such situations, the implementation of the methods discussed in the above mentioned papers could only be performed on the verified subjects, typically yielding biased estimates of ROC surface and VUS. This bias is known as verification bias.

In order to correct for verification bias, the researchers often assume that the selection for disease verification does not depend on the disease status, given the test results and other observed covariates, i.e., they assume that the true disease status, when missing, is missing at random (MAR, Little and Rubin 2002). Under this assumption, there exist few methods to get bias-corrected inference in ROC surface analysis. Chi and Zhou (2008) proposed a nonparametric likelihood-based approach to obtain bias-corrected estimators for ROC surface and VUS of an ordinal diagnostic test. In case of continuous diagnostic tests, To Duc et al. (2016) discussed several solutions based on imputation and re-weighting methods, and proposed four verification bias-corrected estimators of the ROC surface and VUS: full imputation (FI), mean score imputation (MSI), inverse probability weighting (IPW) and semi-parametric efficient (SPE) estimators.

However, in some studies the decision to send a subject to verification may be directly based on the presumed subject's disease status, or, more generally, the selection mechanism may depend on some unobserved covariates related to disease; in these cases, the MAR assumption does not hold and the missing data mechanism is called nonignorable (NI).

For two-class problems, methods to deal with NI verification bias have been developed, for instance, in Baker (1995), Zhou and Rodenberg (1998), Zhou and Castelluccio (2003), Zhou and Castelluccio (2004), Rotnitzky et al. (2006), Fluss et al. (2009), Fluss et al. (2012) and Liu and Zhou (2010). However, the issue of correcting for NI verification bias in ROC surface analysis is very scarcely considered in the statistical literature. To the best of our knowledge, only Zhang and Alonzo (2018)

gave a contribution in this direction. In detail, the Authors: (i) extended to the VUS the doubly robust estimator for the area under the ROC curve of Rotnitzky et al. (2006); (ii) proposed an IPW estimator of the VUS again based on the model of Rotnitzky et al. (2006); and (iii) considered an extension of the model in Liu and Zhou (2010) to obtain a pseudo doubly robust (PDR) estimator of the VUS.

However, on the one hand, techniques discussed by Zhang and Alonzo (2018) appear to lack generality. Indeed, with reference to the verification process, in proposals (i) and (ii) the planned adjustment for selection to verification is assumed to be user-specified, i.e., fixed and known in practice; moreover, in all proposals, i.e., (i)–(iii), the disease variable is treated in an unnatural way as a numerical variable. On the other hand, asymptotic properties of the proposed estimators are not theoretically investigated, nor a proof of identifiability for the model is provided in case (iii), in which the adjustment for selection to verification is assumed to be unknown.

In this work, we try to overcome limitations of current proposals in developing bias-corrected methods for continuous diagnostic tests with three-class disease status, under a NI missing data mechanism. In particular, we adopt parametric regression models for the disease and the verification processes, properly extending the selection model of Liu and Zhou (2010) to match the case of three-class disease status. Then, we use likelihood-based estimators of model parameters to derive four estimators (FI, MSI, IPW, PDR) of the VUS. Within this approach, we prove identifiability of the model as well as consistency and asymptotic normality of the proposed estimators. Estimation of their variance is also discussed.

The rest of the paper is organized as follows. In Sect. 2, we set the working model and discuss its identifiability. In Sect. 3 we present our proposed bias-corrected VUS estimators, along with theoretical results about consistency and asymptotic normality. The results of a simulation study are presented in Sect. 4 and an illustration based on data from the Alzheimer's Disease Neuroimaging Initiative is provided in Sect. 5. Concluding remarks are left to Sect. 6. Technical details and a discussion about variance estimation can be found in the appendices.

2 Model for NI missing data mechanism

2.1 Background

Suppose we need to evaluate the predictive ability of a new continuous diagnostic test in a context where the disease status of a patient can be described by three ordered categories, "non-diseased", "intermediate" and "diseased", say. Consider a sample of n subjects and let T, $\mathcal{D}$ and A denote the test result, the disease status and a vector of covariates for each subject, respectively. In this framework, $\mathcal{D}$ can be modeled as a trinomial random vector $\mathcal{D} = (D_1, D_2, D_3)^\top$, such that D_k is a Bernoulli random variable having mean $\theta_k = \Pr(D_k = 1)$ where $\theta_1 + \theta_2 + \theta_3 = 1$. Hence, θ_k represents the probability that a generic subject, classified according to its disease status, belong to the class k. We are interested in estimating the VUS of the test T, say μ, which is defined as (Nakas and Yiannoutsos 2004)

$$\mu = \Pr\left(T_i < T_\ell < T_r | D_{1i} = 1, D_{2\ell} = 1, D_{3r} = 1\right)$$
$$+ \frac{1}{2}\Pr\left(T_i < T_\ell = T_r | D_{1i} = 1, D_{2\ell} = 1, D_{3r} = 1\right)$$
$$+ \frac{1}{2}\Pr\left(T_i = T_\ell < T_r | D_{1i} = 1, D_{2\ell} = 1, D_{3r} = 1\right)$$
$$+ \frac{1}{6}\Pr\left(T_i = T_\ell = T_r | D_{1i} = 1, D_{2\ell} = 1, D_{3r} = 1\right)$$

or, equivalently,

$$\mu = \frac{\mathbb{E}\left(D_{1i} D_{2\ell} D_{3r} I_{i\ell r}\right)}{\mathbb{E}\left(D_{1i} D_{2\ell} D_{3r}\right)}, \tag{1}$$

where the indices i, ℓ, r refer to three different subjects, $I_{i\ell r} = I(T_i < T_\ell < T_r) + \frac{1}{2}I(T_i < T_\ell = T_r) + \frac{1}{2}I(T_i = T_\ell < T_r) + \frac{1}{6}I(T_i = T_\ell = T_r)$ and $I(\cdot)$ is the indicator function.

When the disease status $\mathcal{D}$ is available for all subjects, a natural nonparametric estimator of μ is given by

$$\hat{\mu}_{\mathrm{NP}} = \frac{\sum_{i=1}^n \sum_{\ell=1, \ell \neq i}^n \sum_{\substack{r=1 \\ r \neq \ell, r \neq i}}^n I_{i\ell r} D_{1i} D_{2\ell} D_{3r}}{\sum_{i=1}^n \sum_{\ell=1, \ell \neq i}^n \sum_{\substack{r=1 \\ r \neq \ell, r \neq i}}^n D_{1i} D_{2\ell} D_{3r}}. \tag{2}$$

However, in many situations not all subjects undergo the verification process, and hence, the disease status $\mathcal{D}$ is missing in a subset of patients in the study. Let V_i be the verification status for the ith subject: $V_i = 1$ if $\mathcal{D}_i$ is observed and $V_i = 0$ otherwise. We define the observed data as the set $\{O_i = (\mathcal{D}_i^\top, V_i, T_i, A_i^\top)^\top, i = 1, \ldots, n\}$. When the true disease status is subject to NI missingness, estimators working under the MAR assumption cannot be applied tout court. Our goal is to adjust FI, MSI, IPW and SPE estimators discussed in To Duc et al. (2016) to the framework of NI missingness.

2.2 Model settings

To deal with NI missing data mechanism, in what follows we extend parametric models adopted in Liu and Zhou (2010) for the two-class problem to the three-class case. More precisely, with three disease categories, we fix the model for the verification process as follows

$$\pi = \Pr(V = 1 | D_1, D_2, T, A) = \frac{\exp\{h(T, A; \boldsymbol{\tau}_\pi) + \lambda_1 D_1 + \lambda_2 D_2\}}{1 + \exp\{h(T, A; \boldsymbol{\tau}_\pi) + \lambda_1 D_1 + \lambda_2 D_2\}}, \tag{3}$$

where D_1 and D_2 are defined in the previous section, $h(T, A; \boldsymbol{\tau}_\pi)$ is, in general, an arbitrary working function, and $\boldsymbol{\tau}_\pi$ is a set of parameters. Here, $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)^\top$ is the non-ignorable parameter: the missing data mechanism is MAR if $\lambda_1 = \lambda_2 = 0$; NI, otherwise. As for the disease model, we employ the multinomial logistic regression for the whole sample, i.e.,

$$\rho_k = \Pr(D_k = 1 | T, A) = \frac{\exp\left\{f(T, A; \tau_{\rho_k})\right\}}{1 + \exp\left\{f(T, A; \tau_{\rho_1})\right\} + \exp\left\{f(T, A; \tau_{\rho_2})\right\}}, \quad (4)$$

where $f(T, A; \tau_{\rho_k})$ is an arbitrary working function, and τ_{ρ_k} is a set of parameters, for $k = 1, 2$. The parameters $\lambda, \tau_\pi, \tau_\rho$, with $\tau_\rho = (\tau_{\rho_1}^\top, \tau_{\rho_2}^\top)^\top$, can be estimated jointly by using a likelihood-based approach.

It is worth noting that, under (3), an application of Bayes' rule gives that

$$\frac{\Pr(D_1 = 1 | V = 1, T, A)}{\Pr(D_1 = 1 | V = 0, T, A)} = \frac{\Pr(V = 0 | T, A)}{\Pr(V = 1 | T, A)} \exp\left\{h(T, A; \tau_\pi) + \lambda_1\right\},$$

$$\frac{\Pr(D_2 = 1 | V = 1, T, A)}{\Pr(D_2 = 1 | V = 0, T, A)} = \frac{\Pr(V = 0 | T, A)}{\Pr(V = 1 | T, A)} \exp\left\{h(T, A; \tau_\pi) + \lambda_2\right\},$$

$$\frac{\Pr(D_3 = 1 | V = 1, T, A)}{\Pr(D_3 = 1 | V = 0, T, A)} = \frac{\Pr(V = 0 | T, A)}{\Pr(V = 1 | T, A)} \exp\left\{h(T, A; \tau_\pi)\right\}.$$

Therefore,

$$\frac{\Pr(D_1 = 1 | V = 1, T, A)}{\Pr(D_1 = 1 | V = 0, T, A)} \bigg/ \frac{\Pr(D_3 = 1 | V = 1, T, A)}{\Pr(D_3 = 1 | V = 0, T, A)} = \exp(\lambda_1), \quad (5)$$

$$\frac{\Pr(D_2 = 1 | V = 1, T, A)}{\Pr(D_2 = 1 | V = 0, T, A)} \bigg/ \frac{\Pr(D_3 = 1 | V = 1, T, A)}{\Pr(D_3 = 1 | V = 0, T, A)} = \exp(\lambda_2), \quad (6)$$

so that, according to (5) and (6), λ_1 and λ_2 can also be interpreted as log-odds ratios of belonging to class 1 (instead of class 3) and to class 2 (instead of class 3), respectively, for a verified subject compared to an unverified subject with the same test result and covariates.

2.3 Parameter estimation

As in Liu and Zhou (2010), in our model, for simplicity, we take $h(T, A; \tau_\pi) = \tau_{\pi_1} + \tau_{\pi_2} T + A^\top \tau_{\pi_3}$ and $f(T, A; \tau_{\rho_k}) = \tau_{\rho_{1k}} + \tau_{\rho_{2k}} T + A^\top \tau_{\rho_{3k}}$, which is a natural choice in practice. For fixed T and A, the observed distribution is fully determined by the three probabilities $\Pr(D_1 = 1, D_2 = 0, V = 1 | T, A)$, $\Pr(D_1 = 0, D_2 = 1, V = 1 | T, A)$ and $\Pr(D_1 = 0, D_2 = 0, V = 1 | T, A)$. It is easy to show that

$$\begin{aligned}
&\Pr(D_1 = 1, D_2 = 0, V = 1 | T, A) \\
&= \Pr(D_1 = 1, D_2 = 0 | T, A)\Pr(V = 1 | D_1 = 1, D_2 = 0, T, A) \\
&= \Pr(D_1 = 1 | T, A)\Pr(V = 1 | D_1 = 1, D_2 = 0, T, A) \\
&= \rho_1 \pi_{10}.
\end{aligned}$$

Similarly, we have that

$$\Pr(D_1 = 0, D_2 = 1, V = 1 | T, A) = \rho_2 \pi_{01},$$
$$\Pr(D_1 = 0, D_2 = 0, V = 1 | T, A) = (1 - \rho_1 - \rho_2)\pi_{00} = \rho_3 \pi_{00},$$

with $\pi_{01} = \Pr(V = 1 | D_1 = 0, D_2 = 1, T, A)$, $\pi_{00} = \Pr(V = 1 | D_1 = 0, D_2 = 0, T, A)$ and $\rho_3 = 1 - \rho_1 - \rho_2$. Then,

$$\Pr(V = 1 | T, A) = \rho_1 \pi_{10} + \rho_2 \pi_{01} + \rho_3 \pi_{00},$$

and $\Pr(V = 0 | T, A) = 1 - \Pr(V = 1 | T, A) = 1 - \rho_1 \pi_{10} + \rho_2 \pi_{01} + \rho_3 \pi_{00}$. It follows that the log-likelihood function can be written as

$$
\begin{aligned}
\log L(\lambda, \tau_\pi, \tau_\rho) \\
= \sum_{i=1}^{n} \Bigg\{ & D_{1i} V_i \log(\rho_{1i} \pi_{10i}) + D_{2i} V_i \log(\rho_{2i} \pi_{01i}) + D_{3i} V_i \log(\rho_{3i} \pi_{00i}) \\
& + (1 - V_i) \log(1 - \rho_{1i} \pi_{10i} - \rho_{2i} \pi_{01i} - \rho_{3i} \pi_{00i}) \Bigg\}.
\end{aligned}
\tag{7}
$$

The estimates $\hat{\lambda}$, $\hat{\tau}_\pi$, and $\hat{\tau}_\rho$ can be obtained by maximizing $\log L(\lambda, \tau_\pi, \tau_\rho)$ or by solving the score equations

$$
0 = \sum_{i=1}^{n} \left\{ D_{1i} V_i (1 - \pi_{10i}) - \frac{(1 - V_i) \rho_{1i} \pi_{10i} (1 - \pi_{10i})}{1 - \rho_{1i} \pi_{10i} - \rho_{2i} \pi_{01i} - \rho_{3i} \pi_{00i}} \right\},
$$

$$
0 = \sum_{i=1}^{n} \left\{ D_{2i} V_i (1 - \pi_{01i}) - \frac{(1 - V_i) \rho_{2i} \pi_{01i} (1 - \pi_{01i})}{1 - \rho_{1i} \pi_{10i} - \rho_{2i} \pi_{01i} - \rho_{3i} \pi_{00i}} \right\},
$$

$$
\begin{aligned}
0 = \sum_{i=1}^{n} U_i \Bigg\{ & D_{1i} V_i (1 - \pi_{10i}) + D_{2i} V_i (1 - \pi_{01i}) + D_{3i} V_i (1 - \pi_{00i}) \\
& - (1 - V_i) \frac{\rho_{1i} \pi_{10i} (1 - \pi_{10i}) + \rho_{2i} \pi_{01i} (1 - \pi_{01i}) + \rho_{3i} \pi_{00i} (1 - \pi_{00i})}{1 - \rho_{1i} \pi_{10i} - \rho_{2i} \pi_{01i} - \rho_{3i} \pi_{00i}} \Bigg\},
\end{aligned}
$$

$$
\begin{aligned}
0 = \sum_{i=1}^{n} U_i \Bigg\{ & V_i (D_{1i} - \rho_{1i}) \\
& - (1 - V_i) \frac{(\pi_{10i} - \pi_{00i}) \rho_{1i} (1 - \rho_{1i}) - (\pi_{01i} - \pi_{00i}) \rho_{1i} \rho_{2i}}{1 - \rho_{1i} \pi_{10i} - \rho_{2i} \pi_{01i} - \rho_{3i} \pi_{00i}} \Bigg\},
\end{aligned}
$$

$$
\begin{aligned}
0 = \sum_{i=1}^{n} U_i \Bigg\{ & V_i (D_{2i} - \rho_{2i}) \\
& - (1 - V_i) \frac{(\pi_{01i} - \pi_{00i}) \rho_{2i} (1 - \rho_{2i}) - (\pi_{10i} - \pi_{00i}) \rho_{1i} \rho_{2i}}{1 - \rho_{1i} \pi_{10i} - \rho_{2i} \pi_{01i} - \rho_{3i} \pi_{00i}} \Bigg\},
\end{aligned}
$$

where $U_i = (1, T_i, A_i^\top)^\top$. The above equations are obtained by using the following results: $\frac{\partial}{\partial \lambda_1} \pi_{10i} = \pi_{10i}(1 - \pi_{10i})$, $\frac{\partial}{\partial \lambda_2} \pi_{01i} = \pi_{01i}(1 - \pi_{01i})$, $\frac{\partial}{\partial \tau_\pi} \pi_{d_1 d_2 i} = U_i(1 - \pi_{d_1 d_2 i}) \pi_{d_1 d_2 i}$ (here (d_1, d_2) is a pair in the set $\{(1, 0), (0, 1), (0, 0)\}$), and

$$\frac{\partial}{\partial \boldsymbol{\tau}_{\rho_1}^\top} \rho_{1i} = \boldsymbol{U}_i \rho_{1i}(1 - \rho_{1i}); \qquad \frac{\partial}{\partial \boldsymbol{\tau}_{\rho_2}^\top} \rho_{1i} = -\boldsymbol{U}_i \rho_{1i} \rho_{2i};$$

$$\frac{\partial}{\partial \boldsymbol{\tau}_{\rho_2}^\top} \rho_{2i} = \boldsymbol{U}_i \rho_{2i}(1 - \rho_{2i}); \qquad \frac{\partial}{\partial \boldsymbol{\tau}_{\rho_1}^\top} \rho_{2i} = -\boldsymbol{U}_i \rho_{1i} \rho_{2i}.$$

2.4 Identifiability

In this section, we verify that the working model based on (3), with $h(T, \boldsymbol{A}; \boldsymbol{\tau}_\pi)$ $= \tau_{\pi_1} + \tau_{\pi_2} T + \boldsymbol{A}^\top \boldsymbol{\tau}_{\pi_3}$, and (4), with $f(T, \boldsymbol{A}; \boldsymbol{\tau}_{\rho_k}) = \tau_{\rho_{1k}} + \tau_{\rho_{2k}} T + \boldsymbol{A}^\top \boldsymbol{\tau}_{\rho_{3k}}$, is identifiable. Since the log-likelihood (7) is fully determined by the three probabilities $\Pr(D_1 = 1, D_2 = 0, V = 1 | T, \boldsymbol{A})$, $\Pr(D_1 = 0, D_2 = 1, V = 1 | T, \boldsymbol{A})$ and $\Pr(D_1 = 0, D_2 = 0, V = 1 | T, \boldsymbol{A})$, we have to show that such probabilities are uniquely determined by the parameters for all possible T and $\boldsymbol{A}$. For the sake of simplicity, in the remainder of this section the auxiliary covariate $\boldsymbol{A}$ is omitted (actually, we can always view $\boldsymbol{A}$ as fixed while varying T).

Let $\boldsymbol{\xi} = (\lambda_1, \lambda_2, \tau_{\pi_1}, \tau_{\pi_2}, \tau_{\rho_{11}}, \tau_{\rho_{21}}, \tau_{\rho_{12}}, \tau_{\rho_{22}})^\top$ be the set of parameters. For given $T = t$, we can write

$$\log(\rho_1 \pi_{10}) = (\tau_{\rho_{11}} + \tau_{\rho_{21}} t) - \log\left\{1 + \exp(\tau_{\rho_{11}} + \tau_{\rho_{21}} t) + \exp(\tau_{\rho_{12}} + \tau_{\rho_{22}} t)\right\}$$
$$+ (\tau_{\pi_1} + \tau_{\pi_2} t) + \lambda_1 - \log\left\{1 + \exp(\tau_{\pi_1} + \tau_{\pi_2} t) \exp(\lambda_1)\right\},$$
$$\log(\rho_2 \pi_{01}) = (\tau_{\rho_{12}} + \tau_{\rho_{22}} t) - \log\left\{1 + \exp(\tau_{\rho_{11}} + \tau_{\rho_{21}} t) + \exp(\tau_{\rho_{12}} + \tau_{\rho_{22}} t)\right\}$$
$$+ (\tau_{\pi_1} + \tau_{\pi_2} t) + \lambda_2 - \log\left\{1 + \exp(\tau_{\pi_1} + \tau_{\pi_2} t) \exp(\lambda_2)\right\},$$
$$\log(\rho_3 \pi_{00}) = -\log\left\{1 + \exp(\tau_{\rho_{11}} + \tau_{\rho_{21}} t) + \exp(\tau_{\rho_{12}} + \tau_{\rho_{22}} t)\right\} + (\tau_{\pi_1} + \tau_{\pi_2} t)$$
$$- \log\left\{1 + \exp(\tau_{\pi_1} + \tau_{\pi_2} t)\right\}.$$

Let $x(t) = \tau_{\pi_1} + \tau_{\pi_2} t$, $y(t) = \tau_{\rho_{11}} + \tau_{\rho_{21}} t$ and $z(t) = \tau_{\rho_{12}} + \tau_{\rho_{22}} t$, for each $t \in \mathbb{R}$. The above expressions, which refer to the quantities characterizing the log-likelihood function (7), can be rewritten as

$$\log(\rho_3 \pi_{00}) = -\log\left\{1 + \exp(y(t)) + \exp(z(t))\right\} + x(t) - \log\left\{1 + \exp(x(t))\right\},$$
$$\log(\rho_1 \pi_{10}) = y(t) - \log\left\{1 + \exp(y(t)) + \exp(z(t))\right\} + x(t) + \lambda_1$$
$$- \log\left\{1 + \exp(x(t)) \exp(\lambda_1)\right\}$$
$$= \log(\rho_3 \pi_{00}) + \log\left\{1 + \exp(x(t))\right\} + y(t) + \lambda_1$$
$$- \log\left\{1 + \exp(x(t)) \exp(\lambda_1)\right\}$$
$$= \log(\rho_3 \pi_{00}) + y(t) + \log\left\{1 + \exp(x(t))\right\}$$
$$- \log\left\{\exp(-\lambda_1) + \exp(x(t))\right\}$$
$$= \log(\rho_3 \pi_{00}) + y(t) + \log\left\{\frac{1 + \exp(x(t))}{\exp(-\lambda_1) + \exp(x(t))}\right\},$$
$$\log(\rho_2 \pi_{01}) = z(t) - \log\left\{1 + \exp(y(t)) + \exp(z(t))\right\} + x(t) + \lambda_2$$
$$- \log\left\{1 + \exp(x(t)) \exp(\lambda_2)\right\}$$
$$= \log(\rho_3 \pi_{00}) + z(t) + \log\left\{\frac{1 + \exp(x(t))}{\exp(-\lambda_2) + \exp(x(t))}\right\}.$$

Now, assume that there are two distinct points $\boldsymbol{\xi}$ and $\boldsymbol{\xi}^*$ ($\boldsymbol{\xi} \neq \boldsymbol{\xi}^*$) in the parameter space, such that the following equations (with obvious notation) hold:

$$\rho_1 \pi_{10} = \rho_1^* \pi_{10}^*, \tag{8}$$

$$\rho_2 \pi_{01} = \rho_2^* \pi_{01}^*, \tag{9}$$

$$\rho_3 \pi_{00} = \rho_3^* \pi_{00}^*, \tag{10}$$

for all $t \in \mathbb{R}$. By using (10), the Eqs. (8) and (9) are equivalent to

$$
\begin{aligned}
&y(t) - y^*(t) \\
&= \log \left\{ \frac{1 + \exp(x^*(t))}{\exp(-\lambda_1^*) + \exp(x^*(t))} \right\} - \log \left\{ \frac{1 + \exp(x(t))}{\exp(-\lambda_1) + \exp(x(t))} \right\}, \quad (11)
\end{aligned}
$$

$$
\begin{aligned}
&z(t) - z^*(t) \\
&= \log \left\{ \frac{1 + \exp(x^*(t))}{\exp(-\lambda_2^*) + \exp(x^*(t))} \right\} - \log \left\{ \frac{1 + \exp(x(t))}{\exp(-\lambda_2) + \exp(x(t))} \right\}, \quad (12)
\end{aligned}
$$

respectively. In (11) and (12) the left hand sides are straight lines. Thus, in order to (11) and (12) hold for all t, the right hand sides must be constants. If these constants were 0 (because $\lambda_1 = \lambda_1^* = \lambda_2 = \lambda_2^* = 0$), then (10) would no longer hold for $\boldsymbol{\xi} \neq \boldsymbol{\xi}^*$ and all t. Alternatively, the right hand sides of (11) and (12) are non-zero constants if $\tau_{\pi_2} = \tau_{\pi_2}^* = 0$. Then, as a consequence, (10) still is valid, for $\boldsymbol{\xi} \neq \boldsymbol{\xi}^*$ and all t, eventually if $\tau_{\rho_{21}} = \tau_{\rho_{21}}^* = 0$ and $\tau_{\rho_{22}} = \tau_{\rho_{22}}^* = 0$. This allows us to state that: if $\Pr(D_k|T) \neq \Pr(D_k)$, with $k = 1, 2$, then the considered model (with the particular choice for the functions h and f) is identifiable, i.e., the joint probabilities $\Pr(D_1 = 1, D_2 = 0, V = 1|T = t)$, $\Pr(D_1 = 0, D_2 = 1, V = 1|T = t)$ and $\Pr(D_1 = 0, D_2 = 0, V = 1|T = t)$ are determined by a unique set of parameters. Of course, this claim can be easily extended to handle the presence of a covariate vector, $\boldsymbol{A}$.

3 The proposal

3.1 VUS estimators

Let $\rho_{k(v)} = \Pr(D_k = 1|V = v, T, \boldsymbol{A})$, for $k = 1, 2, 3$ and $v = 0, 1$. It is easy to see, for instance, that

$$
\begin{aligned}
\rho_{1(v)} &= \frac{\Pr(V = v, D_1 = 1|D_2 = 0, T, \boldsymbol{A})}{\Pr(V = v|T, \boldsymbol{A})} \\
&= \frac{\Pr(V = v|D_1 = 1, D_2 = 0, T, \boldsymbol{A})\Pr(D_1 = 1|T, \boldsymbol{A})}{\Pr(V = v|T, \boldsymbol{A})}.
\end{aligned}
$$

Hence, we can get, in particular,

$$\rho_{1(0)} = \frac{(1 - \pi_{10})\rho_1}{(1 - \pi_{10})\rho_1 + (1 - \pi_{01})\rho_2 + (1 - \pi_{00})\rho_3},$$

$$\rho_{2(0)} = \frac{(1 - \pi_{01})\rho_2}{(1 - \pi_{10})\rho_1 + (1 - \pi_{01})\rho_2 + (1 - \pi_{00})\rho_3},$$

$$\rho_{3(0)} = \frac{(1 - \pi_{00})\rho_3}{(1 - \pi_{10})\rho_1 + (1 - \pi_{01})\rho_2 + (1 - \pi_{00})\rho_3}.$$

Clearly, we also may consider quantities as

$$\rho_{1(1)} = \frac{\pi_{10}\rho_1}{\pi_{10}\rho_1 + \pi_{01}\rho_2 + \pi_{00}\rho_3}.$$

Then, we observe that

$$\begin{aligned}
\mathbb{E}(D_{1i} D_{2\ell} D_{3r} \mathrm{I}_{i\ell r}) &= \mathbb{E}_{T,A} \{\mathrm{I}_{i\ell r} \mathbb{E}(D_{1i} D_{2\ell} D_{3r} | T_i, A_i, T_\ell, A_\ell, T_r, A_r)\} \\
&= \mathbb{E}_{T,A} \{\mathrm{I}_{i\ell r} \mathbb{E}(D_{1i} | T_i, A_i) \mathbb{E}(D_{2\ell} | T_\ell, A_\ell) \mathbb{E}(D_{3r} | T_r, A_r)\} \\
&= \mathbb{E}_{T,A} (\rho_{1i} \rho_{2\ell} \rho_{3r} \mathrm{I}_{i\ell r}).
\end{aligned}$$

Similarly, we have

$$\mathbb{E}(D_{1i} D_{2\ell} D_{3r}) = \mathbb{E}_{T,A} (\rho_{1i} \rho_{2\ell} \rho_{3r}),$$

so that (1) can be rewritten as

$$\mu = \frac{\mathbb{E}_{T,A} (\rho_{1i} \rho_{2\ell} \rho_{3r} \mathrm{I}_{i\ell r})}{\mathbb{E}_{T,A} (\rho_{1i} \rho_{2\ell} \rho_{3r})}. \tag{13}$$

Equation (13) suggests how to build estimators of VUS when some disease labels are missing in the sample: we can use suitable estimates $\hat{\rho}_{ki}$ to replace the D_{ki}'s in (2). Therefore, a FI estimator of VUS is simply

$$\hat{\mu}_{\mathrm{FI}} = \frac{\sum_{i=1}^n \sum_{\ell=1,\ell\neq i}^n \sum_{\substack{r=1 \\ r\neq\ell,r\neq i}}^n \mathrm{I}_{i\ell r} \hat{\rho}_{1i} \hat{\rho}_{2\ell} \hat{\rho}_{3r}}{\sum_{i=1}^n \sum_{\ell=1,\ell\neq i}^n \sum_{\substack{r=1 \\ r\neq\ell,r\neq i}}^n \hat{\rho}_{1i} \hat{\rho}_{2\ell} \hat{\rho}_{3r}}, \tag{14}$$

where $\hat{\rho}_{ki}$ ($k = 1, 2, 3$ and $i = 1, \ldots, n$) are the estimated disease probabilities obtained from the disease model (4).

Since $\mathbb{E}[V_i \rho_{k(1)i} + (1 - V_i)\rho_{k(0)i} | T, A] = \rho_{ki}$, an alternative FI estimator of VUS could be obtained by replacing D_{ki}'s in (2) with the estimates $\tilde{D}_{ki,\mathrm{FI}} = V_i \hat{\rho}_{k(1)i} + (1 - V_i)\hat{\rho}_{k(0)i}$. Unlike FI approach, MSI estimator only replace the disease status D_{ki} by the estimate $\hat{\rho}_{k(0)i}$ for unverified subjects. Define $D_{ki,\mathrm{MSI}} = V_i D_{ki} + (1 - V_i)\rho_{k(0)i}$ and let $\tilde{D}_{ki,\mathrm{MSI}}$ be the estimated version with $\rho_{k(0)i}$ replaced by $\hat{\rho}_{k(0)i}$, and

$$\hat{\rho}_{1(0)i} = \frac{(1 - \hat{\pi}_{10i})\hat{\rho}_{1i}}{(1 - \hat{\pi}_{10i})\hat{\rho}_{ki} + (1 - \hat{\pi}_{01i})\hat{\rho}_{2i} + (1 - \hat{\pi}_{00i})\hat{\rho}_{3i}},$$

$$\hat{\rho}_{2(0)i} = \frac{(1 - \hat{\pi}_{01i})\hat{\rho}_{2i}}{(1 - \hat{\pi}_{10i})\hat{\rho}_{1i} + (1 - \hat{\pi}_{01i})\hat{\rho}_{2i} + (1 - \hat{\pi}_{00i})\hat{\rho}_{3i}},$$

$$\hat{\rho}_{3(0)i} = \frac{(1 - \hat{\pi}_{00i})\hat{\rho}_{3i}}{(1 - \hat{\pi}_{10i})\hat{\rho}_{1i} + (1 - \hat{\pi}_{01i})\hat{\rho}_{2i} + (1 - \hat{\pi}_{00i})\hat{\rho}_{3i}}.$$

Here, $\hat{\pi}_{10i} = \widehat{\Pr}(V_i = 1 | D_{1i} = 1, D_{2i} = 0, T_i, A_i)$, $\hat{\pi}_{01i} = \widehat{\Pr}(V_i = 1 | D_{1i} = 0, D_{2i} = 1, T_i, A_i)$ and $\hat{\pi}_{00i} = \widehat{\Pr}(V_i = 1 | D_{1i} = 0, D_{2i} = 0, T_i, A_i)$. Such estimates are derived from the verification model (3). Then, the MSI estimator of VUS is

$$\hat{\mu}_{\text{MSI}} = \frac{\sum_{i=1}^{n} \sum_{\ell=1, \ell \neq i}^{n} \sum_{\substack{r=1 \\ r \neq \ell, r \neq i}}^{n} \mathrm{I}_{i\ell r} \tilde{D}_{1i,\text{MSI}} \tilde{D}_{2\ell,\text{MSI}} \tilde{D}_{3r,\text{MSI}}}{\sum_{i=1}^{n} \sum_{\ell=1, \ell \neq i}^{n} \sum_{\substack{r=1 \\ r \neq \ell, r \neq i}}^{n} \tilde{D}_{1i,\text{MSI}} \tilde{D}_{2\ell,\text{MSI}} \tilde{D}_{3r,\text{MSI}}}. \tag{15}$$

In the IPW approach, instead, each observation in the subset of verified units is weighted by the inverse of the probability that the unit was selected for verification. Thus, the IPW estimator of VUS is

$$\hat{\mu}_{\text{IPW}} = \frac{\sum_{i=1}^{n} \sum_{\ell=1, \ell \neq i}^{n} \sum_{\substack{r=1 \\ r \neq \ell, r \neq i}}^{n} \mathrm{I}_{i\ell r} V_i V_\ell V_r D_{1i} D_{2\ell} D_{3r} \hat{\pi}_i^{-1} \hat{\pi}_\ell^{-1} \hat{\pi}_r^{-1}}{\sum_{i=1}^{n} \sum_{\ell=1, \ell \neq i}^{n} \sum_{\substack{r=1 \\ r \neq \ell, r \neq i}}^{n} V_i V_\ell V_r D_{1i} D_{2\ell} D_{3r} \hat{\pi}_i^{-1} \hat{\pi}_\ell^{-1} \hat{\pi}_r^{-1}}. \tag{16}$$

Clearly, the estimates $\hat{\pi}_i$ also arise from the selection model (3).

The last estimator is the pseudo doubly robust (PDR) estimator. We define

$$D_{ki,\text{PDR}} = \frac{V_i D_{ki}}{\pi_i} - \frac{\rho_{k(0)i}(V_i - \pi_i)}{\pi_i}.$$

An estimated version, $\tilde{D}_{ki,\text{PDR}}$, is obtained by entering the estimates $\hat{\pi}_i$ and $\hat{\rho}_{k(0)i}$ in the expression above. Then, the PDR estimator of VUS is

$$\hat{\mu}_{\text{PDR}} = \frac{\sum_{i=1}^{n} \sum_{\ell=1, \ell \neq i}^{n} \sum_{\substack{r=1 \\ r \neq \ell, r \neq i}}^{n} \mathrm{I}_{i\ell r} \tilde{D}_{1i,\text{PDR}} \tilde{D}_{2\ell,\text{PDR}} \tilde{D}_{3r,\text{PDR}}}{\sum_{i=1}^{n} \sum_{\ell=1, \ell \neq i}^{n} \sum_{\substack{r=1 \\ r \neq \ell, r \neq i}}^{n} \tilde{D}_{1i,\text{PDR}} \tilde{D}_{2\ell,\text{PDR}} \tilde{D}_{3r,\text{PDR}}}. \tag{17}$$

The PDR estimator has the same nature as the SPE estimator discussed in To Duc et al. (2016) under MAR assumption. However, under NI missing data mechanism it no longer has the doubly robust property. In fact, correct specification of both the verification model and the disease model is required for the PDR estimator to be consistent.

Note that all VUS estimators basically require maximum likelihood estimates of the parameters $\boldsymbol{\lambda}$, $\boldsymbol{\tau}_\pi$ and $\boldsymbol{\tau}_\rho$ of the working models (3) and (4).

3.2 Asymptotic behavior

Let $\boldsymbol{\xi} = (\boldsymbol{\lambda}^\top, \boldsymbol{\tau}_\pi^\top, \boldsymbol{\tau}_\rho^\top)^\top$ be the nuisance parameter. Observe that the proposed VUS estimators can be found as solutions of appropriate estimating equations (solved along with the score equations). The estimating functions for FI, MSI, IPW and PDR estimators have generic term (corresponding to a generic triplet of sample units), respectively,

$$G_{i\ell r,\mathrm{FI}}(\mu, \boldsymbol{\xi}) = \rho_{1i}(\boldsymbol{\tau}_\rho)\rho_{2\ell}(\boldsymbol{\tau}_\rho)\rho_{3r}(\boldsymbol{\tau}_\rho)\,(I_{i\ell r} - \mu),$$

$$G_{i\ell r,\mathrm{MSI}}(\mu, \boldsymbol{\xi}) = D_{1i,\mathrm{MSI}}(\boldsymbol{\xi})D_{2\ell,\mathrm{MSI}}(\boldsymbol{\xi})D_{3r,\mathrm{MSI}}(\boldsymbol{\xi})\,(I_{i\ell r} - \mu),$$

$$G_{i\ell r,\mathrm{IPW}}(\mu, \boldsymbol{\xi}) = \frac{V_i V_\ell V_r D_{1i} D_{2\ell} D_{3r}}{\pi_i(\boldsymbol{\xi})\pi_\ell(\boldsymbol{\xi})\pi_k(\boldsymbol{\xi})}\,(I_{i\ell r} - \mu),$$

$$G_{i\ell r,\mathrm{PDR}}(\mu, \boldsymbol{\xi}) = D_{1i,\mathrm{PDR}}(\boldsymbol{\xi})D_{2\ell,\mathrm{PDR}}(\boldsymbol{\xi})D_{3r,\mathrm{PDR}}(\boldsymbol{\xi})\,(I_{i\ell r} - \mu).$$

In the following, we will use the general notation $G_{i\ell r,*}(\mu, \boldsymbol{\xi})$, where the star stands for FI, MSI, IPW and PDR.

Recall that the nuisance parameter $\boldsymbol{\xi}$ is estimated by maximizing the log-likelihood function (7). Let $\mathcal{S}_i(\boldsymbol{\xi})$ be the i-th subject's contribution to the score function, $\mathcal{I}(\boldsymbol{\xi}) = -\mathbb{E}\left(\frac{\partial}{\partial \boldsymbol{\xi}^\top}\mathcal{S}_i(\boldsymbol{\xi})\right)$ the Fisher information matrix for $\boldsymbol{\xi}$ and $\hat{\boldsymbol{\xi}}$ the maximum likelihood estimator. Let μ_0 be the true VUS value, and $\boldsymbol{\xi}_0 = (\boldsymbol{\lambda}_0^\top, \boldsymbol{\tau}_{0\pi}^\top, \boldsymbol{\tau}_{0\rho}^\top)^\top$ the true value of $\boldsymbol{\xi}$. To give general theoretical results, we assume that:

(C1) the U-process

$$U_{n,*}(\mu, \boldsymbol{\xi}) = \sqrt{n}\,\{G_*(\mu, \boldsymbol{\xi}) - e_*(\mu, \boldsymbol{\xi})\}$$

is stochastically equicontinuous, where

$$G_*(\mu, \boldsymbol{\xi}) = \frac{1}{6n(n-1)(n-2)}\sum_{i=1}^{n}\sum_{\substack{\ell=1,\ell\neq i}}^{n}\sum_{\substack{r=1\\r\neq\ell,r\neq i}}^{n}\left\{G_{i\ell r,*}(\mu, \boldsymbol{\xi})\right.$$
$$+ G_{ir\ell,*}(\mu, \boldsymbol{\xi}) + G_{\ell ir,*}(\mu, \boldsymbol{\xi}) + G_{\ell ri,*}(\mu, \boldsymbol{\xi})$$
$$\left. + G_{ri\ell,*}(\mu, \boldsymbol{\xi}) + G_{r\ell i,*}(\mu, \boldsymbol{\xi})\right\}$$

and

$$e_*(\mu, \boldsymbol{\xi}) = \frac{1}{6}\mathbb{E}\left\{G_{i\ell r,*}(\mu, \boldsymbol{\xi}) + G_{ir\ell,*}(\mu, \boldsymbol{\xi}) + G_{\ell ir,*}(\mu, \boldsymbol{\xi}) + G_{\ell ri,*}(\mu, \boldsymbol{\xi})\right.$$
$$\left. + G_{ri\ell,*}(\mu, \boldsymbol{\xi}) + G_{r\ell i,*}(\mu, \boldsymbol{\xi})\right\};$$

(C2) $e_*(\mu, \boldsymbol{\xi})$ is differentiable in $(\mu, \boldsymbol{\xi})$, and $\left.\dfrac{\partial e_*(\mu, \boldsymbol{\xi}_0)}{\partial \mu}\right|_{\mu=\mu_0} \neq 0;$

(C3) $G_*(\mu, \boldsymbol{\xi})$ and $\frac{\partial}{\partial \boldsymbol{\xi}^\top} G_*(\mu, \boldsymbol{\xi})$ converges uniformly (in probability) to $e_*(\mu, \boldsymbol{\xi})$ and $\frac{\partial}{\partial \boldsymbol{\xi}^\top} e_*(\mu, \boldsymbol{\xi})$, respectively.

We now state the two main results about consistency and asymptotic normality of the proposed estimators, whose proves are given in "Appendix 1".

Theorem 1 (Consistency) *Suppose that conditions (C1)–(C3) hold. Then, under the verification model* (3) *and the disease model* (4), $\hat{\mu}_* \xrightarrow{p} \mu_0$.

Recall that here the star indicates FI, MSI, IPW, and PDR.

Theorem 2 (Asymptotic normality) *Under conditions (C1)–(C3), if the verification model* (3) *and the disease model* (4) *hold, then*

$$\sqrt{n} \left(\hat{\mu}_* - \mu_0 \right) \xrightarrow{d} \mathcal{N}(0, \Lambda_*),$$

where Λ_* *is given in* (20).

It is worth noting that conditions (C1)–(C3) hold in our working model, which is based on (3), with $h(T, A; \boldsymbol{\tau}_\pi) = \tau_{\pi_1} + \tau_{\pi_2} T + \boldsymbol{A}^\top \boldsymbol{\tau}_{\pi_3}$, and (4), with $f(T, A; \boldsymbol{\tau}_{\rho_k}) = \tau_{\rho_{1k}} + \tau_{\rho_{2k}} T + \boldsymbol{A}^\top \boldsymbol{\tau}_{\rho_{3k}}$. In "Appendix 1" we discuss how to obtain a consistent estimator of Λ_*.

4 Simulation study

In this section, we provide empirical evidence, through simulation experiments, on the behavior of the proposed VUS estimators in finite samples. The number of replications in each simulation experiment is set to be 1000.

In the study, we consider two scenarios which correspond to quite different values of the true VUS. For both scenarios, we fix three sample sizes: 250, 500 and 1500.

In the first scenario, for each unit, we generate the test result T_i and a covariate A_i from a bivariate normal distribution,

$$(T_i, A_i) \sim \mathcal{N}_2 \left(\begin{pmatrix} 3.7 \\ 1.85 \end{pmatrix}, \begin{pmatrix} 3.71 & 1.36 \\ 1.36 & 3.13 \end{pmatrix} \right).$$

The disease status $\mathcal{D}_i$ is generated according to model (4) with $f(T, A; \boldsymbol{\tau}_{\rho_1}) = 15 - 3.3T - 0.7A$ and $f(T, A; \boldsymbol{\tau}_{\rho_2}) = 9.5 - 1.7T - 0.3A$. Then, the verification label V_i is obtained according to model (3) with $h(T, A; \boldsymbol{\tau}_\pi) = 2 + 0.5T - 1.2A$ and $\lambda_1 = -2, \lambda_2 = -1$. Under such data generating process, $\theta_1 = 0.4$, $\theta_2 = 0.35$, $\theta_3 = 0.25$, and the verification rate is roughly 0.57. The true VUS value is 0.791. In the second scenario, we generate the test result and the covariate from independent normal distributions. Specifically, $T_i \sim \mathcal{N}(0.65, 1)$ and $A_i \sim \mathcal{N}(-0.3, 0.64)$. The disease status $\mathcal{D}_i$ is generated according to model (4) with $f(T, A; \boldsymbol{\tau}_{\rho_1}) = 4.6 - 3.3T - 6.4A$ and $f(T, A; \boldsymbol{\tau}_{\rho_2}) = 4 - 1.7T - 3.2A$. Then, V_i is obtained according to model (3) with $h(T, A; \boldsymbol{\tau}_\pi) = 1 + 1.2T - 1.5A$ and $\lambda_1 = -2.5, \lambda_2 = -1$. Under this setting,

$\theta_1 = 0.55$, $\theta_2 = 0.32$, $\theta_3 = 0.13$, and the verification rate is roughly 0.58. The true VUS value is 0.387.

Table 1 contains Monte Carlo means, Monte Carlo standard deviations and estimated standard deviations for the proposed VUS estimators (FI, MSI, IPW, PDR) in the two considered scenarios, at the chosen sample sizes. The table also reports the empirical coverages of the 95% confidence intervals for the VUS, obtained through the normal approximation approach applied to each estimator. To make a comparison, Table 1 also gives the results for the semiparametric efficient estimator (SPE) discussed in To Duc et al. (2016), whose realizations are obtained, in all experiments, under the MAR assumption, i.e., by setting $\lambda_1 = \lambda_2 = 0$ in model (3). The comparison allows us to evaluate the possible impact of an incorrect hypothesis MAR on the most robust estimator among those, FI, MSI, IPW and SPE, which are built to work under ignorable missing data mechanism (see To Duc et al. 2016).

Overall, simulation results are consistent with our theoretical findings and show the usefulness of the proposed estimators, which also arises from the comparison with the SPE estimator used improperly. The results also show a good behavior of the estimated standard deviations, which are generally close to the corresponding Monte Carlo values. In general, FI and MSI estimators seem to be more efficient than IPW and PDR estimators. However, for all estimators, acceptable bias levels and sufficiently accurate associated confidence intervals seem to require a large sample size (at least 500, and, prudently, even higher).

This issue of poor accuracy has already been noted by several authors, including Liu and Zhou (2010), in the context of two-class classification problems. In our experience, the trouble appears to arise because of a bad behavior of the maximum likelihood estimates in the verification and disease models. If the sample size is not large enough, the data do not contain enough information to effectively estimate the parameters $\boldsymbol{\lambda}$, $\boldsymbol{\tau}_\pi$, $\boldsymbol{\tau}_{\rho_1}$ and $\boldsymbol{\tau}_{\rho_2}$. It seems particularly difficult to get good estimates of nonignorable parameters.

Table 2, giving the Monte Carlo means for the maximum likelihood estimators of the elements of $\boldsymbol{\lambda}$, $\boldsymbol{\tau}_\pi$, $\boldsymbol{\tau}_{\rho_1}$ and $\boldsymbol{\tau}_{\rho_2}$, for the three considered sample sizes, allows us to look at the bias of the estimators. More importantly, Figs. 1 and 2 (which refer to scenario I and II, respectively) graphically depict values of the estimates of λ_1, λ_2 and τ_{π_1} obtained in the thousand replications, for each sample size. The plots clearly show the instability of the maximum likelihood estimates at lower sample sizes, with many values dramatically different from the corresponding target values, heavily impacting on the bias of the estimators. With larger sample size, this phenomenon almost completely vanishes, the maximum likelihood estimators behave pretty well, with a positive impact on the behavior of the VUS estimators.

As requested by a Reviewer, in "Appendix 3" we present some other results, deriving from simulation experiments that cover certain specific situations.

5 An illustration

To illustrate the application of our proposed methods, we used data from Alzheimer's Disease Neuroimaging Initiative (ADNI, adni.loni.usc.edu). ADNI was launched in

Table 1 Monte Carlo means (MCmean), relative bias (Bias), Monte Carlo standard deviations (MCsd) and estimated standard deviations (Esd) for the proposed VUS estimators, and the SPE estimator under MAR assumption

Sample size	Estimators	MCmean	Bias (%)	MCsd	Esd	CP (%)
Scenario I: VUS = 0.791						
$n = 250$	FI	0.772	−2.4	0.056	0.050	89.9
	MSI	0.770	−2.7	0.057	0.051	90.6
	IPW	0.770	−2.6	0.070	0.061	88.1
	PDR	0.766	−3.2	0.085	0.075	90.8
	SPE (MAR)	0.771	−2.5	0.073	0.138	93.2
$n = 500$	FI	0.783	−1.0	0.035	0.032	93.3
	MSI	0.782	−1.1	0.036	0.033	93.4
	IPW	0.782	−1.2	0.047	0.042	92.2
	PDR	0.782	−1.2	0.053	0.058	94.0
	SPE (MAR)	0.771	−2.6	0.047	0.040	93.0
$n = 1500$	FI	0.790	−0.2	0.016	0.016	95.0
	MSI	0.789	−0.2	0.016	0.016	95.2
	IPW	0.788	−0.3	0.025	0.024	94.4
	PDR	0.789	−0.3	0.025	0.024	95.2
	SPE (MAR)	0.771	−2.5	0.027	0.025	89.4
Scenario II: VUS = 0.387						
$n = 250$	FI	0.368	−5.0	0.064	0.057	87.4
	MSI	0.367	−5.2	0.065	0.059	87.9
	IPW	0.377	−2.6	0.084	0.074	87.6
	PDR	0.369	−4.6	0.086	0.075	89.5
	SPE (MAR)	0.346	−10.6	0.063	0.058	84.5
$n = 500$	FI	0.379	−2.0	0.045	0.041	90.9
	MSI	0.379	−2.1	0.046	0.042	91.3
	IPW	0.380	−1.8	0.060	0.056	91.2
	PDR	0.381	−1.6	0.060	0.053	92.0
	SPE (MAR)	0.345	−10.8	0.044	0.042	76.5
$n = 1500$	FI	0.388	0.2	0.023	0.022	94.2
	MSI	0.388	0.2	0.023	0.023	94.3
	IPW	0.388	0.3	0.034	0.032	94.9
	PDR	0.389	0.4	0.033	0.029	93.2
	SPE (MAR)	0.346	−10.7	0.026	0.025	76.5

CP denotes Monte Carlo coverages for the 95% confidence intervals, obtained through the normal approximation approach applied to each estimator

2003 as a public–private partnership with the primary goal of finding suitable diagnostic tests or biomarkers for early detection and tracking of the Alzheimer's disease (see www.adni-info.org for up-to-date information).

Study subjects are classified in one of three classes, i.e., cognitively normal (CN), mild cognitive impairment (MCI), and Alzheimer's disease (AD) on the basis of neu-

Table 2 Monte Carlo means (MCmean) for the maximum likelihood estimators of the elements of nuisance parameters $\boldsymbol{\lambda}$, $\boldsymbol{\tau}_{\pi}$, $\boldsymbol{\tau}_{\rho_1}$ and $\boldsymbol{\tau}_{\rho_2}$

	Scenario I				Scenario II			
	True	MCmean			True	MCmean		
n		250	500	1500		250	500	1500
λ_1	−2.00	−1.01	−1.76	−1.95	−2.50	−2.09	−2.30	−2.50
λ_2	−1.00	−0.45	−0.87	−0.98	−1.00	−0.99	−0.96	−0.97
τ_{π_1}	2.00	1.25	1.80	1.95	1.00	1.17	1.00	1.00
τ_{π_2}	0.50	0.65	0.55	0.51	1.20	1.39	1.28	1.22
τ_{π_3}	−1.20	−1.24	−1.22	−1.21	−1.50	−1.25	−1.40	−1.51
$\tau_{\rho_{11}}$	15.00	15.53	15.28	15.10	4.60	4.44	4.58	4.66
$\tau_{\rho_{21}}$	−3.30	−3.41	−3.36	−3.32	−3.30	−3.29	−3.33	−3.34
$\tau_{\rho_{31}}$	−0.70	−0.89	−0.78	−0.72	−6.40	−6.94	−6.70	−6.48
$\tau_{\rho_{12}}$	9.50	10.03	9.71	9.57	4.00	4.12	4.11	4.05
$\tau_{\rho_{22}}$	−1.70	−1.79	−1.73	−1.71	−1.70	−1.77	−1.76	−1.73
$\tau_{\rho_{32}}$	−0.30	−0.40	−0.34	−0.31	−3.20	−3.62	−3.42	−3.25

ropsychological tests. Various clinical, imaging, genetic and biochemical markers are also available. Among them, we consider cerebrospinal fluid (CSF) tau protein as the primary marker, and amyloid beta 1–42 ($A\beta$1–42) as a covariate. The full dataset in our illustration refers to 1209 subjects (CN: 363, MCI: 618, AD: 228 at baseline visit). Exploratory analysis suggests that high values of CFS tau protein are associated with severe disease status.

Here, we want to evaluate the accuracy of CSF tau protein as a marker for the Alzheimer's disease assuming nonignorable missingness in the disease status of some patients. To this aim, we induce missingness by randomly selecting patients from the three classes in different proportions. Then, about 66% and 16% patients in AD and MCI classes, respectively, are given a missing disease status. No missing status is set in the CN class. Overall, in the resulting dataset, for 76.4% of the subjects the true disease status is present. For convenience, values of the considered variables, i.e., CSF tau protein and amyloid beta 1–42, are standardized. In what follows, the standardized variables will be denoted as T and A, respectively.

We focus on VUS estimation of marker T by applying the FI, MSI, IPW and PDR estimators proposed in Sect. 3. For the sake of comparison, bias-corrected VUS estimators under MAR assumption (To Duc et al. 2016), are also employed. Estimates under the MAR assumption are obtained through the R package `bcROCsurface` (To Duc 2017), in which the verification model is fitted by using logistic regression with T and A as covariates, and the disease model is fitted by means of multinomial logistic regression with T and A as regressors. Moreover, we also compute the VUS estimate based on the full dataset (Full estimate).

The maximum likelihood estimate of the non-ignorable parameter $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)^\top$ in the verification model (3) is $(18.766, 3.344)^\top$. The log-likelihood ratio test for the
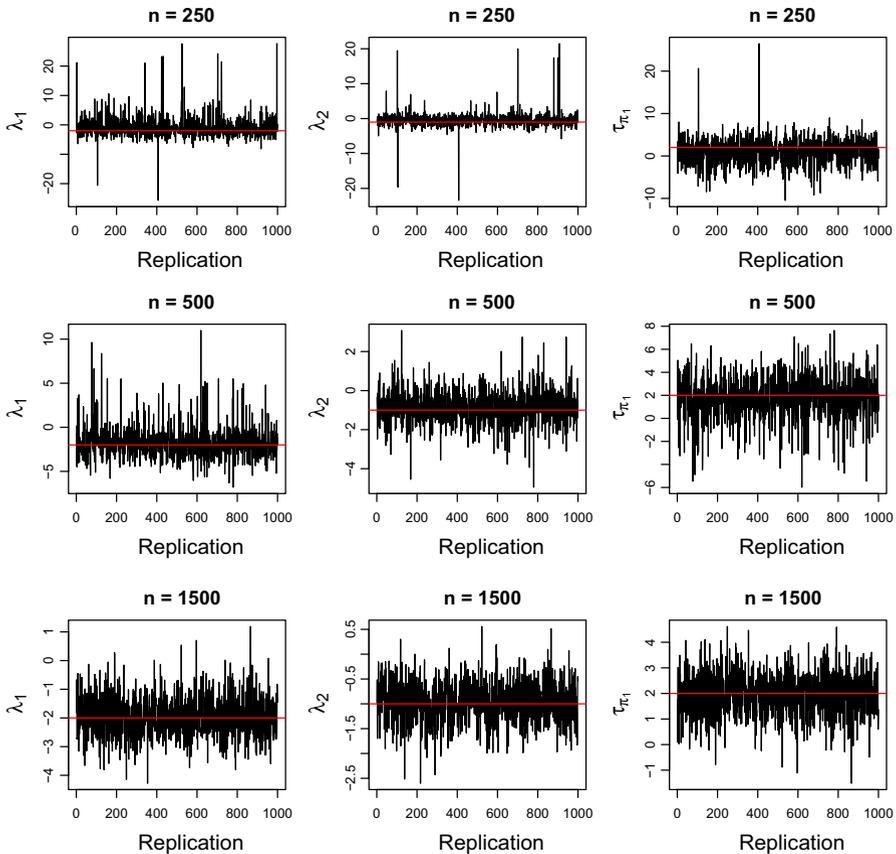
Fig. 1 Values of the estimates of λ_1, λ_2, and τ_{π_1}, obtained in thousand Monte Carlo replications (Scenario I), for each sample size. The horizontal lines indicate the true parameter values

hypothesis $H_0 : \boldsymbol{\lambda} = \mathbf{0}$ versus $H_A : \boldsymbol{\lambda} \neq \mathbf{0}$ achieves a p-value of 0.0066, indicating that the non-ignorable effect is significant.

Table 3 shows the Full and the bias-corrected NI and MAR estimates of VUS. The table also gives the estimated standard deviations, and approximated 95% confidence intervals. For the Full estimate, the bootstrap standard deviation is reported (250 bootstrap replications). Standard deviations of NI and MAR estimates are obtained by using asymptotic theory. As expected, taking the Full estimate as a benchmark, MAR estimators seem to overestimate the VUS, whereas the NI estimators appear to perform better, in particular FI and MSI. Consistently with simulation results in Table 1, the NI bias-corrected FI and MSI estimators appear also to be more efficient than the IPW and PDR estimators. Taking into account that the logistic model (3) used for the verification process does not reflect the induced verification mechanism, this seems to suggest that the NI bias-corrected FI and MSI estimators are less sensitive to possible misspecifications of the verification process.
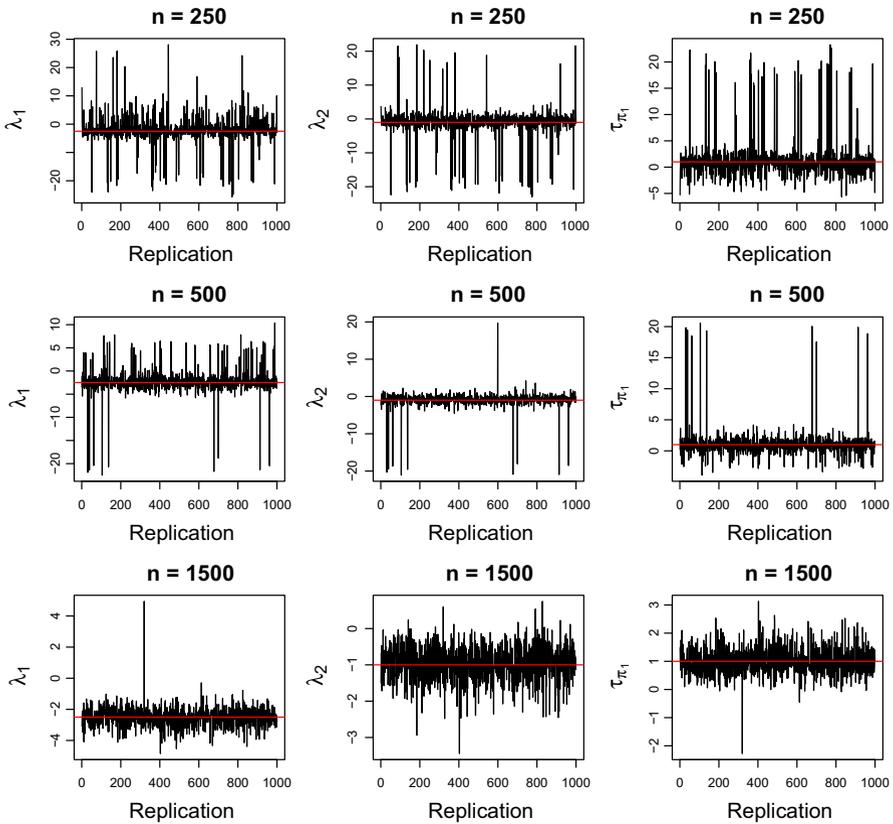
Fig. 2 Values of the estimates of λ_1, λ_2, and τ_{π_1}, obtained in thousand Monte Carlo replications (Scenario II), for each sample size. The horizontal lines indicate the true parameter values

Table 3 ADNI dataset. Estimates of VUS for the standardized CSF tau protein, associated estimated standard deviations (SD) and 95% confidence intervals. SD for Full estimated is obtained via bootstrap resampling

		VUS Estimate	SD	95% confidence interval
Full		0.338	0.019	(0.301, 0.375)
NI	FI	0.335	0.016	(0.303, 0.368)
	MSI	0.333	0.018	(0.298, 0.368)
	IPW	0.352	0.024	(0.305, 0.398)
	PDR	0.344	0.021	(0.302, 0.384)
MAR	FI	0.369	0.031	(0.309, 0.429)
	MSI	0.366	0.031	(0.304, 0.427)
	IPW	0.373	0.034	(0.307, 0.440)
	SPE	0.362	0.033	(0.297, 0.426)

6 Conclusion

In this paper, we have proposed four bias–corrected estimators of VUS under NI missing data mechanism. The estimators are obtained by a likelihood–based approach, which uses the verification model (3) together with the disease model (4). The identifiability of the joint model is proved, and hence, the nuisance parameters can be estimated by maximizing the log-likelihood function or solving the score equations. Consistency and asymptotic normality of the proposed FI, MSI, IPW and PDR estimators are established, and variance estimation is discussed.

The proposed VUS estimators are pretty easy to implement and require the use of some numerical routine to maximize the log-likelihood function (or to solve the score equations). Our simulation results show their usefulness, whilst confirming the evidence emerging in the two-class case, according to which a reasonable large sample size is necessary to make sufficiently accurate inference. In practice, among FI, MSI, IPW and PDR estimators, we would reccommend FI and MSI estimators thanks to their greater efficiency.

The poor accuracy problem seems to be related to an intrinsic difficulty of the maximum likelihood method in providing accurate estimates of the parameters of the disease and verification models, in particular of the nonignorable parameters. Overcoming this drawback is a stimulating challenge and deserves further investigation.

Appendix 1

Proves

Proof of Theorem 1 We can show that $\mathbb{E}\{G_{i\ell r,*}(\mu_0, \boldsymbol{\xi}_0)\} = 0$ (see the "Appendix 2"). Then $e_*(\mu_0, \boldsymbol{\xi}_0) = 0$, and, by condition (C2) and an application of implicit function theorem, there exists a neighborhood of $\boldsymbol{\xi}_0$ in which a continuously differentiable function, $m(\boldsymbol{\xi})$, is uniquely defined such that $m(\boldsymbol{\xi}_0) = \mu_0$ and $e_*(m(\boldsymbol{\xi}), \boldsymbol{\xi}) = 0$. Since the maximum likelihood estimator $\hat{\boldsymbol{\xi}}$ is consistent, i.e., $\hat{\boldsymbol{\xi}} \xrightarrow{p} \boldsymbol{\xi}_0$, we have that

$\tilde{\mu}_* = m(\hat{\boldsymbol{\xi}}) \xrightarrow{P} \mu_0$. On the other hand, $G_*(\hat{\mu}_*, \hat{\boldsymbol{\xi}}) = 0$ and condition (C3) implies that $e_*(\hat{\mu}_*, \hat{\boldsymbol{\xi}}) \xrightarrow{P} 0$. Thus, $\hat{\mu}_* \xrightarrow{P} \tilde{\mu}_*$. □

Proof of Theorem 2 We have

$$0 = \sqrt{n} G_*(\hat{\mu}_*, \hat{\boldsymbol{\xi}})$$
$$0 = \sqrt{n} G_*(\hat{\mu}_*, \hat{\boldsymbol{\xi}}) + \sqrt{n} e_*(\hat{\mu}_*, \hat{\boldsymbol{\xi}}) - \sqrt{n} e_*(\hat{\mu}_*, \hat{\boldsymbol{\xi}}).$$

Since $e_*(\mu_0, \boldsymbol{\xi}_0) = 0$, we get

$$\begin{aligned}
0 &= \sqrt{n} G_*(\hat{\mu}_*, \hat{\boldsymbol{\xi}}) + \sqrt{n} e_*(\hat{\mu}_*, \hat{\boldsymbol{\xi}}) - \sqrt{n} e_*(\hat{\mu}_*, \hat{\boldsymbol{\xi}}) + \sqrt{n} e_*(\mu_0, \boldsymbol{\xi}_0) - \sqrt{n} e_*(\mu_0, \boldsymbol{\xi}_0) \\
&= \sqrt{n} \left\{ G_*(\hat{\mu}_*, \hat{\boldsymbol{\xi}}) - e_*(\hat{\mu}_*, \hat{\boldsymbol{\xi}}) \right\} + \sqrt{n} \left\{ e_*(\hat{\mu}_*, \hat{\boldsymbol{\xi}}) - e_*(\mu_0, \boldsymbol{\xi}_0) \right\} + \sqrt{n} e_*(\mu_0, \boldsymbol{\xi}_0) \\
&\quad - \sqrt{n} G_*(\mu_0, \boldsymbol{\xi}_0) + \sqrt{n} G_*(\mu_0, \boldsymbol{\xi}_0) \\
&= \left[\sqrt{n} \left\{ G_*(\hat{\mu}_*, \hat{\boldsymbol{\xi}}) - e_*(\hat{\mu}_*, \hat{\boldsymbol{\xi}}) \right\} - \sqrt{n} \left\{ G_*(\mu_0, \boldsymbol{\xi}_0) - e_*(\mu_0, \boldsymbol{\xi}_0) \right\} \right] \\
&\quad + \sqrt{n} \left\{ e_*(\hat{\mu}_*, \hat{\boldsymbol{\xi}}) - e_*(\mu_0, \boldsymbol{\xi}_0) \right\} + \sqrt{n} G_*(\mu_0, \boldsymbol{\xi}_0).
\end{aligned}$$

Condition (C1) implies that the first term in right hand side of the last identity is $o_p(1)$. Using the Taylor expansion, we have

$$\begin{aligned}
0 &= o_p(1) + \sqrt{n} \left\{ e_*(\hat{\mu}_*, \hat{\boldsymbol{\xi}}) - e_*(\mu_0, \boldsymbol{\xi}_0) \right\} + \sqrt{n} G_*(\mu_0, \boldsymbol{\xi}_0) \\
&= o_p(1) + \sqrt{n}(\hat{\mu}_* - \mu_0) \frac{\partial e_*(\mu, \boldsymbol{\xi}_0)}{\partial \mu} \bigg|_{\mu = \mu_0} \\
&\quad + \sqrt{n}(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0) \frac{\partial e_*(\mu_0, \boldsymbol{\xi})}{\partial \boldsymbol{\xi}^\top} \bigg|_{\boldsymbol{\xi} = \boldsymbol{\xi}_0} + \sqrt{n} G_*(\mu_0, \boldsymbol{\xi}_0).
\end{aligned} \tag{18}$$

It is straightforward to show that

$$\frac{\partial e_*(\mu, \boldsymbol{\xi}_0)}{\partial \mu} \bigg|_{\mu = \mu_0} = -\Pr(D_1 = 1)\Pr(D_2 = 1)\Pr(D_3 = 1) = -\theta_1 \theta_2 \theta_3.$$

By standard results on the limit distribution of U-statistics (van der Vaart 2000, Theorem 12.3, Chap. 12),

$$\begin{aligned}
\sqrt{n} U_{n,*}(\mu_0, \boldsymbol{\xi}_0) &= \sqrt{n} \left\{ G_*(\mu_0, \boldsymbol{\xi}_0) - e_*(\mu_0, \boldsymbol{\xi}_0) \right\} \\
&= \sqrt{n} G_*(\mu_0, \boldsymbol{\xi}_0) \xrightarrow{p} \sqrt{n} \tilde{G}_*(\mu_0, \boldsymbol{\xi}_0),
\end{aligned}$$

where $\sqrt{n} \tilde{G}_*(\mu, \boldsymbol{\xi})$ is the projection of $U_{n,*}$ onto the set of all statistics of the form

$$\sqrt{n}\tilde{G}_{n,*}(\mu, \boldsymbol{\xi}) = \frac{1}{2\sqrt{n}} \sum_{i=1}^{n} \mathbb{E}\Big\{ G_{i\ell r,*}(\mu, \boldsymbol{\xi}) + G_{ir\ell,*}(\mu, \boldsymbol{\xi}) + G_{\ell ir,*}(\mu, \boldsymbol{\xi})$$

$$+ G_{\ell ri,*}(\mu, \boldsymbol{\xi}) + G_{ri\ell,*}(\mu, \boldsymbol{\xi}) + G_{r\ell i,*}(\mu, \boldsymbol{\xi}) \big| O_i \Big\}$$

for $\ell \neq i$ and $r \neq \ell, r \neq i$. For the maximum likelihood estimator $\hat{\boldsymbol{\xi}}$, we can write

$$\sqrt{n}\left(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0\right) = \frac{1}{\sqrt{n}} \left[-\frac{\partial \mathbb{E}\left\{ \mathcal{S}_i(\boldsymbol{\xi}) \right\}}{\partial \boldsymbol{\xi}^\top} \Bigg|_{\boldsymbol{\xi}=\boldsymbol{\xi}_0} \right]^{-1} \sum_{i=1}^{n} \mathcal{S}_i(\boldsymbol{\xi}_0) + o_p(1)$$

$$= \frac{1}{\sqrt{n}} \mathcal{I}(\boldsymbol{\xi})^{-1} \sum_{i=1}^{n} \mathcal{S}_i(\boldsymbol{\xi}_0) + o_p(1).$$

Hence, from (18),

$$\theta_1\theta_2\theta_3\sqrt{n}(\hat{\mu}_* - \mu_0)$$

$$= o_p(1) + \frac{1}{\sqrt{n}} \frac{\partial e_*(\mu_0, \boldsymbol{\xi})}{\partial \boldsymbol{\xi}^\top} \Bigg|_{\boldsymbol{\xi}=\boldsymbol{\xi}_0} \mathcal{I}(\boldsymbol{\xi})^{-1} \sum_{i=1}^{n} \mathcal{S}_i(\boldsymbol{\xi}_0)$$

$$+ \frac{1}{2\sqrt{n}} \sum_{i=1}^{n} \mathbb{E}\Big\{ G_{i\ell r,*}(\mu_0, \boldsymbol{\xi}_0) + G_{ir\ell,*}(\mu_0, \boldsymbol{\xi}_0) + G_{\ell ir,*}(\mu_0, \boldsymbol{\xi}_0)$$

$$+ G_{\ell ri,*}(\mu_0, \boldsymbol{\xi}_0) + G_{ri\ell,*}(\mu_0, \boldsymbol{\xi}_0) + G_{r\ell i,*}(\mu_0, \boldsymbol{\xi}_0) \big| O_i \Big\}$$

$$= o_p(1) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Bigg[\frac{\partial e_*(\mu_0, \boldsymbol{\xi})}{\partial \boldsymbol{\xi}^\top} \Bigg|_{\boldsymbol{\xi}=\boldsymbol{\xi}_0} \mathcal{I}(\boldsymbol{\xi})^{-1} \mathcal{S}_i(\boldsymbol{\xi}_0)$$

$$+ \frac{1}{2} \mathbb{E}\Big\{ G_{i\ell r,*}(\mu_0, \boldsymbol{\xi}_0) + G_{ir\ell,*}(\mu_0, \boldsymbol{\xi}_0) + G_{\ell ir,*}(\mu_0, \boldsymbol{\xi}_0)$$

$$+ G_{\ell ri,*}(\mu_0, \boldsymbol{\xi}_0) + G_{ri\ell,*}(\mu_0, \boldsymbol{\xi}_0) + G_{r\ell i,*}(\mu_0, \boldsymbol{\xi}_0) \big| O_i \Big\} \Bigg]$$

$$= o_p(1) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Q_{i,*}(\mu_0, \boldsymbol{\xi}_0) = o_p(1) + \frac{1}{\sqrt{n}} Q_*(\mu_0, \boldsymbol{\xi}_0). \tag{19}$$

Note that the observed data O_i are i.i.d, then $Q_{i,*}(\mu_0, \boldsymbol{\xi}_0)$ are also i.i.d. In addition, we easily show that

$$0 = \mathbb{E}\Bigg[\mathbb{E}\Big\{ G_{i\ell r,*}(\mu_0, \boldsymbol{\xi}_0) + G_{ir\ell,*}(\mu_0, \boldsymbol{\xi}_0) + G_{\ell ir,*}(\mu_0, \boldsymbol{\xi}_0) + G_{\ell ri,*}(\mu_0, \boldsymbol{\xi}_0)$$

$$+ G_{ri\ell,*}(\mu_0, \boldsymbol{\xi}_0) + G_{r\ell i,*}(\mu_0, \boldsymbol{\xi}_0) \big| O_i \Big\} \Bigg].$$

Therefore, $\mathbb{E}\{Q_{i,*}(\mu_0, \boldsymbol{\xi}_0)\} = 0$, and $\frac{1}{\sqrt{n}} Q_*(\mu_0, \boldsymbol{\xi}_0) \xrightarrow{d} \mathcal{N}(0, \mathbb{V}\mathrm{ar}\{Q_{i,*}(\mu_0, \boldsymbol{\xi}_0)\})$ by the Central Limit Theorem. It follows that

$$\sqrt{n}\left(\hat{\mu}_* - \mu_0\right) \xrightarrow{d} \mathcal{N}\left(0, \Lambda_*\right),$$

where

$$\Lambda_* = \frac{\mathbb{V}\mathrm{ar}\{Q_{i,*}(\mu_0, \boldsymbol{\xi}_0)\}}{\theta_1^2 \theta_2^2 \theta_3^2}. \tag{20}$$

$\square$

Variance estimation

Under condition (C3), a consistent estimator of Λ_* can be obtained as

$$\hat{\Lambda}_* = \frac{\mathbb{V}\mathrm{ar}\left\{\hat{Q}_{i,*}(\hat{\mu}_*, \hat{\boldsymbol{\xi}})\right\}}{\hat{\theta}_{1,*}^2 \hat{\theta}_{2,*}^2 \hat{\theta}_{3,*}^2} = \frac{\frac{1}{n-1} \sum_{i=1}^{n} \hat{Q}_{i,*}^2(\hat{\mu}_*, \hat{\boldsymbol{\xi}})}{\hat{\theta}_{1,*}^2 \hat{\theta}_{2,*}^2 \hat{\theta}_{3,*}^2}, \tag{21}$$

where $\hat{\theta}_{k,*}$ are the estimates of the disease probabilities, θ_k for $k = 1, 2, 3$. Specifically, $\hat{\theta}_{k,\mathrm{FI}} = \frac{1}{n} \sum_{i=1}^{n} \hat{\rho}_{ki}$, $\hat{\theta}_{k,\mathrm{MSI}} = \frac{1}{n} \sum_{i=1}^{n} \tilde{D}_{ki,\mathrm{MSI}}$, $\hat{\theta}_{k,\mathrm{PDR}} = \frac{1}{n} \sum_{i=1}^{n} \tilde{D}_{ki,\mathrm{PDR}}$ and $\hat{\theta}_{k,\mathrm{IPW}} = \sum_{i=1}^{n} V_i D_{ki} \hat{\pi}_i^{-1} \Big/ \sum_{i=1}^{n} V_i \hat{\pi}_i^{-1}$. According to (19), we have that

$$\hat{Q}_{i,*}(\hat{\mu}_*, \hat{\boldsymbol{\xi}})$$

$$= \left\{ \frac{1}{(n-1)(n-2)} \sum_{\substack{i=1 \\ \ell=i}}^{n} \sum_{\substack{r=1 \\ r \neq \ell, r \neq i}}^{n} \frac{\partial G_{i\ell r,*}(\hat{\mu}_*, \boldsymbol{\xi})}{\partial \boldsymbol{\xi}^\top} \bigg|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}} \right\}$$

$$\times \left\{ -\sum_{i=1}^{n} \frac{\partial \mathcal{S}_i(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}^\top} \bigg|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}} \right\}^{-1} \mathcal{S}_i(\hat{\boldsymbol{\xi}})$$

$$+ \frac{1}{2(n-1)(n-2)} \sum_{\substack{\ell=1 \\ \ell \neq i}}^{n} \sum_{\substack{r=1 \\ r \neq i, r \neq \ell}}^{n} \left\{ G_{i\ell r,*}(\hat{\mu}_*, \hat{\boldsymbol{\xi}}) + G_{ir\ell,*}(\hat{\mu}_*, \hat{\boldsymbol{\xi}}) + G_{\ell ir,*}(\hat{\mu}_*, \hat{\boldsymbol{\xi}}) \right.$$

$$\left. + G_{\ell ri,*}(\hat{\mu}_*, \hat{\boldsymbol{\xi}}) + G_{ri\ell,*}(\hat{\mu}_*, \hat{\boldsymbol{\xi}}) + G_{r\ell i,*}(\hat{\mu}_*, \hat{\boldsymbol{\xi}}) \right\}.$$

In addition, for fixed i, we also have that

$$\sum_{\substack{\ell=1 \\ \ell \neq i}}^{n} \sum_{\substack{r=1 \\ r \neq i, r \neq \ell}}^{n} \left\{ G_{i\ell r,*}(\hat{\mu}_*, \hat{\boldsymbol{\xi}}) + G_{ikr,*}(\hat{\mu}_*, \hat{\boldsymbol{\xi}}) \right\} = 2 \sum_{\substack{\ell=1 \\ \ell \neq i}}^{n} \sum_{\substack{r=1 \\ r \neq i, r \neq \ell}}^{n} G_{i\ell r,*}(\hat{\mu}_*, \hat{\boldsymbol{\xi}}),$$

$$\sum_{\substack{\ell=1 \\ \ell \neq i}}^{n} \sum_{\substack{r=1 \\ r \neq i, r \neq \ell}}^{n} \left\{ G_{\ell i r, *}(\hat{\mu}_*, \hat{\boldsymbol{\xi}}) + G_{r i \ell, *}(\hat{\mu}_*, \hat{\boldsymbol{\xi}}) \right\} = 2 \sum_{\substack{\ell=1 \\ \ell \neq i}}^{n} \sum_{\substack{r=1 \\ r \neq i, r \neq \ell}}^{n} G_{\ell i r, *}(\hat{\mu}_*, \hat{\boldsymbol{\xi}}),$$

$$\sum_{\substack{\ell=1 \\ \ell \neq i}}^{n} \sum_{\substack{r=1 \\ r \neq i, r \neq \ell}}^{n} \left\{ G_{\ell r i, *}(\hat{\mu}_*, \hat{\boldsymbol{\xi}}) + G_{r \ell i, *}(\hat{\mu}_*, \hat{\boldsymbol{\xi}}) \right\} = 2 \sum_{\substack{\ell=1 \\ \ell \neq i}}^{n} \sum_{\substack{r=1 \\ r \neq i, r \neq \ell}}^{n} G_{r \ell i, *}(\hat{\mu}_*, \hat{\boldsymbol{\xi}}).$$

Therefore,

$$
\begin{aligned}
&\hat{Q}_{i, *}(\hat{\mu}_*, \hat{\boldsymbol{\xi}}) \\
&= \left\{ \frac{1}{(n-1)(n-2)} \sum_{i=1}^{n} \sum_{\substack{\ell=1 \\ \ell \neq i}}^{n} \sum_{\substack{r=1 \\ r \neq \ell, r \neq i}}^{n} \frac{\partial G_{i \ell r, *}(\hat{\mu}_*, \boldsymbol{\xi})}{\partial \boldsymbol{\xi}^{\top}} \bigg|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}} \right\} \\
&\quad \times \left\{ -\sum_{i=1}^{n} \frac{\partial \mathcal{S}_i(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}^{\top}} \bigg|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}} \right\}^{-1} \mathcal{S}_i(\hat{\boldsymbol{\xi}}) \\
&\quad + \frac{1}{(n-1)(n-2)} \sum_{\substack{\ell=1 \\ \ell \neq i}}^{n} \sum_{\substack{r=1 \\ r \neq i, r \neq \ell}}^{n} \left\{ G_{i \ell r, *}(\hat{\mu}_*, \hat{\boldsymbol{\xi}}) + G_{\ell i r, *}(\hat{\mu}_*, \hat{\boldsymbol{\xi}}) + G_{r \ell i, *}(\hat{\mu}_*, \hat{\boldsymbol{\xi}}) \right\}.
\end{aligned}
\tag{22}
$$

The quantity $\sum_{i=1}^{n} \frac{\partial \mathcal{S}_i(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}^{\top}} \big|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}}$ could be obtained as the Hessian matrix of the log-likelihood function at $\hat{\boldsymbol{\xi}}$. In order to compute $\frac{\partial G_{i \ell r, *}(\hat{\mu}_*, \boldsymbol{\xi})}{\partial \boldsymbol{\xi}^{\top}} \big|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}}$, we have to get the derivatives $\frac{\partial}{\partial \boldsymbol{\xi}^{\top}} \rho_{ki}(\boldsymbol{\tau}_{0\rho_k})$, $\frac{\partial}{\partial \boldsymbol{\xi}^{\top}} \rho_{k(0)i}(\boldsymbol{\xi})$, $\frac{\partial}{\partial \boldsymbol{\xi}^{\top}} \pi_i^{-1}(\boldsymbol{\lambda}, \boldsymbol{\tau}_{\pi})$, $\frac{\partial}{\partial \boldsymbol{\xi}^{\top}} \pi_{10i}(\boldsymbol{\lambda}, \boldsymbol{\tau}_{\pi})$, $\frac{\partial}{\partial \boldsymbol{\xi}^{\top}} \pi_{01i}(\boldsymbol{\lambda}, \boldsymbol{\tau}_{\pi})$ and $\frac{\partial}{\partial \boldsymbol{\xi}^{\top}} \pi_{00i}(\boldsymbol{\lambda}, \boldsymbol{\tau}_{\pi})$.

In Sect. 2.3, we obtain

$$
\begin{aligned}
\frac{\partial}{\partial \lambda_1} \pi_{10i}(\boldsymbol{\lambda}, \boldsymbol{\tau}_{\pi}) &= \pi_{10i}(1 - \pi_{10i}); & \frac{\partial}{\partial \lambda_2} \pi_{10i}(\boldsymbol{\lambda}, \boldsymbol{\tau}_{\pi}) &= 0; \\
\frac{\partial}{\partial \lambda_1} \pi_{01i}(\boldsymbol{\lambda}, \boldsymbol{\tau}_{\pi}) &= 0; & \frac{\partial}{\partial \lambda_2} \pi_{01i}(\boldsymbol{\lambda}, \boldsymbol{\tau}_{\pi}) &= \pi_{01i}(1 - \pi_{01i}); \\
\frac{\partial}{\partial \lambda_1} \pi_{00i}(\boldsymbol{\lambda}, \boldsymbol{\tau}_{\pi}) &= 0; & \frac{\partial}{\partial \lambda_2} \pi_{00i}(\boldsymbol{\lambda}, \boldsymbol{\tau}_{\pi}) &= 0.
\end{aligned}
$$

and

$$\frac{\partial}{\partial \boldsymbol{\tau}_{\pi}^{\top}} \pi_{d_1 d_2 i} = \boldsymbol{U}_i (1 - \pi_{d_1 d_2 i}) \pi_{d_1 d_2 i},$$

where (d_1, d_2) belongs to the set $\{(1, 0), (0, 1), (0, 0)\}$. Also, we have

$$\frac{\partial}{\partial \boldsymbol{\tau}_{\rho_1}^\top} \rho_{1i}(\boldsymbol{\tau}_\rho) = \boldsymbol{U}_i \rho_{1i}(1 - \rho_{1i}); \quad \frac{\partial}{\partial \boldsymbol{\tau}_{\rho_2}^\top} \rho_{1i}(\boldsymbol{\tau}_\rho) = -\boldsymbol{U}_i \rho_{1i} \rho_{2i};$$

$$\frac{\partial}{\partial \boldsymbol{\tau}_{\rho_2}^\top} \rho_{2i}(\boldsymbol{\tau}_\rho) = \boldsymbol{U}_i \rho_{2i}(1 - \rho_{2i}); \quad \frac{\partial}{\partial \boldsymbol{\tau}_{\rho_1}^\top} \rho_{2i}(\boldsymbol{\tau}_\rho) = -\boldsymbol{U}_i \rho_{1i} \rho_{2i}.$$

Moreover,

$$\frac{\partial}{\partial \lambda_s} \pi_i^{-1}(\boldsymbol{\lambda}, \boldsymbol{\tau}_\pi) = -D_{si} \frac{1 - \pi_i}{\pi_i}; \quad \frac{\partial}{\partial \boldsymbol{\tau}_\pi^\top} \pi_i^{-1}(\boldsymbol{\lambda}, \boldsymbol{\tau}_\pi) = -\boldsymbol{U}_i \frac{1 - \pi_i}{\pi_i},$$

with $s = 1, 2$. Then, recall that

$$\rho_{1(0)i} = \frac{(1 - \pi_{10i})\rho_{1i}}{(1 - \pi_{10i})\rho_{1i} + (1 - \pi_{01i})\rho_{2i} + (1 - \pi_{00i})\rho_{3i}},$$

$$\rho_{2(0)i} = \frac{(1 - \pi_{01i})\rho_{2i}}{(1 - \pi_{10i})\rho_{1i} + (1 - \pi_{01i})\rho_{2i} + (1 - \pi_{00i})\rho_{3i}},$$

$$\rho_{3(0)i} = \frac{(1 - \pi_{00i})\rho_{3i}}{(1 - \pi_{10i})\rho_{1i} + (1 - \pi_{01i})\rho_{2i} + (1 - \pi_{00i})\rho_{3i}}.$$

After some algebra, we get

$$\frac{\partial}{\partial \lambda_1} \rho_{1(0)i}(\boldsymbol{\xi}) = \frac{1}{z^2} \left[-\pi_{10i}(1 - \pi_{10i})\rho_{1i} \{(1 - \pi_{01i})\rho_{2i} + (1 - \pi_{00i})\rho_{3i}\} \right],$$

$$\frac{\partial}{\partial \lambda_2} \rho_{1(0)i}(\boldsymbol{\xi}) = \frac{1}{z^2} \rho_{1i} \rho_{2i} \pi_{01i}(1 - \pi_{01i})(1 - \pi_{10i}),$$

$$\frac{\partial}{\partial \boldsymbol{\tau}_\pi^\top} \rho_{1(0)i}(\boldsymbol{\xi}) = -\frac{\boldsymbol{U}_i}{z^2} \rho_{1i}(1 - \pi_{10i}) \left\{ \rho_{2i}(1 - \pi_{01i})(\pi_{10i} - \pi_{01i}) \right.$$

$$\left. + \rho_{3i}(1 - \pi_{00i})(\pi_{10i} - \pi_{00i}) \right\},$$

$$\frac{\partial}{\partial \boldsymbol{\tau}_{\rho_1}^\top} \rho_{1(0)i}(\boldsymbol{\xi}) = \frac{\boldsymbol{U}_i}{z^2} \rho_{1i}(1 - \pi_{10i}) \{\rho_{2i}(1 - \pi_{01i}) + \rho_{3i}(1 - \pi_{00i})\},$$

$$\frac{\partial}{\partial \boldsymbol{\tau}_{\rho_2}^\top} \rho_{1(0)i}(\boldsymbol{\xi}) = -\frac{\boldsymbol{U}_i}{z^2} \rho_{1i} \rho_{2i}(1 - \pi_{10i})(1 - \pi_{01i}).$$

Finally, we set $z = (1 - \pi_{10i})\rho_{1i} + (1 - \pi_{01i})\rho_{2i} + (1 - \pi_{00i})\rho_{3i}$, and get

$$\frac{\partial}{\partial \lambda_1} \rho_{2(0)i}(\boldsymbol{\xi}) = \frac{1}{z^2} \rho_{1i} \rho_{2i} \pi_{10i}(1 - \pi_{10i})(1 - \pi_{01i}),$$

$$\frac{\partial}{\partial \lambda_2} \rho_{2(0)i}(\boldsymbol{\xi}) = \frac{1}{z^2} \left[-\pi_{01i}(1 - \pi_{01i})\rho_{2i} \{(1 - \pi_{10i})\rho_{1i} + (1 - \pi_{00i})\rho_{3i}\} \right],$$

$$\frac{\partial}{\partial \boldsymbol{\tau}_\pi^\top} \rho_{2(0)i}(\boldsymbol{\xi}) = -\frac{\boldsymbol{U}_i}{z^2} \rho_{2i}(1 - \pi_{01i}) \left\{ \rho_{1i}(1 - \pi_{10i})(\pi_{01i} - \pi_{10i}) \right.$$

$$\left. + \rho_{3i}(1 - \pi_{00i})(\pi_{01i} - \pi_{00i}) \right\},$$

$$\frac{\partial}{\partial \boldsymbol{\tau}_{\rho_1}^\top} \rho_{2(0)i}(\boldsymbol{\xi}) = -\frac{\boldsymbol{U}_i}{z^2} \rho_{1i} \rho_{2i} (1 - \pi_{10i})(1 - \pi_{01i}),$$

$$\frac{\partial}{\partial \boldsymbol{\tau}_{\rho_2}^\top} \rho_{2(0)i}(\boldsymbol{\xi}) = \frac{\boldsymbol{U}_i}{z^2} \rho_{2i} (1 - \pi_{01i}) \{\rho_{1i}(1 - \pi_{10i}) + \rho_{3i}(1 - \pi_{00i})\}.$$

The derivative $\dfrac{\partial}{\partial \boldsymbol{\xi}^\top} \rho_{3(0)i}(\boldsymbol{\xi})$ can be computed by using the fact that $\rho_{3(0)i} = 1 - \rho_{1(0)i} - \rho_{2(0)i}$.

Appendix 2

Here, we show that the estimating functions $G_{i\ell r,*}$ are unbiased under the working disease and verification models. Recall that $\boldsymbol{\xi} = (\boldsymbol{\lambda}^\top, \boldsymbol{\tau}_\pi^\top, \boldsymbol{\tau}_\rho^\top)^\top$.

– *FI estimator*
  We have

$$\begin{aligned}
\mathbb{E}\left\{G_{i\ell r,\mathrm{FI}}(\mu_0, \boldsymbol{\xi}_0)\right\} &= \mathbb{E}\left\{\rho_{1i}(\boldsymbol{\tau}_{0\rho})\rho_{2\ell}(\boldsymbol{\tau}_{0\rho})\rho_{3r}(\boldsymbol{\tau}_{0\rho})(I_{i\ell r} - \mu)\right\} \\
&= \mathbb{E}\left\{\rho_{1i}\rho_{2\ell}\rho_{3r}(I_{i\ell r} - \mu_0)\right\}.
\end{aligned}$$

  Hence, $\mathbb{E}\left\{G_{i\ell r,\mathrm{FI}}(\mu_0, \boldsymbol{\xi}_0)\right\} = 0$ from (13).
– *MSI estimator*
  Consider $\mathbb{E}\left\{D_{ki,\mathrm{MSI}}(\boldsymbol{\xi}_0)|T_i, A_i\right\}$. We have

$$\begin{aligned}
&\mathbb{E}\left\{D_{ki,\mathrm{MSI}}(\boldsymbol{\xi}_0)|T_i, A_i\right\} \\
&= \mathbb{E}\left\{V_i D_{ki} + (1 - V_i)\rho_{k(0)i}(\boldsymbol{\xi}_0)|T_i, A_i\right\} \\
&= \mathbb{E}\left[\mathbb{E}\left\{V_i D_{ki} + (1 - V_i)\rho_{k(0)i}(\boldsymbol{\xi}_0)|T_i, A_i, V_i\right\}|T_i, A_i\right] \\
&= \Pr(V_i = 1|T_i, A_i)\mathbb{E}(D_{ki}|V_i = 1, T_i, A_i) \\
&\quad + \Pr(V_i = 0|T_i, A_i)\mathbb{E}\left(\rho_{k(0)i}(\boldsymbol{\xi}_0)|V_i = 0, T_i, A_i\right) \\
&= \Pr(V_i = 1|T_i, A_i)\Pr(D_{ki} = 1|V_i = 1, T_i, A_i) \\
&\quad + \Pr(V_i = 0|T_i, A_i)\Pr(D_{ki} = 1|V_i = 0, T_i, A_i) \\
&= \Pr(D_{ki} = 1|T_i, A_i) = \rho_{ki}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\mathbb{E}\left\{G_{i\ell r,\mathrm{MSI}}(\mu_0, \boldsymbol{\xi}_0)\right\} \\
&= \mathbb{E}\left\{D_{1i,\mathrm{MSI}}(\boldsymbol{\xi}_0)D_{2\ell,\mathrm{MSI}}(\boldsymbol{\xi}_0)D_{3r,\mathrm{MSI}}(\boldsymbol{\xi}_0)(I_{i\ell r} - \mu_0)\right\} \\
&= \mathbb{E}\Big[(I_{i\ell r} - \mu_0)\mathbb{E}\left\{D_{1i,\mathrm{MSI}}(\boldsymbol{\xi}_0)|T_i, A_i\right\}\mathbb{E}\left\{D_{2\ell,\mathrm{MSI}}(\boldsymbol{\xi}_0)|T_\ell, A_\ell\right\} \\
&\quad \times \mathbb{E}\left\{D_{3r,\mathrm{MSI}}(\boldsymbol{\xi}_0)|T_r, A_r\right\}\Big] \\
&= \mathbb{E}\left\{\rho_{1i}\rho_{2\ell}\rho_{3r}(I_{i\ell r} - \mu_0)\right\}.
\end{aligned}$$

– *IPW estimator*
  In this case,

$$
\begin{aligned}
\mathbb{E}\left(\frac{V_i D_{ki}}{\pi_i(\boldsymbol{\xi}_0)}\middle| T_i, \boldsymbol{A}_i\right) &= \frac{\mathbb{E}(V_i D_{ki}|T_i, \boldsymbol{A}_i)}{\pi_i(\boldsymbol{\xi}_0)} \\
&= \frac{\mathbb{E}\left\{D_{ki}\mathbb{E}(V_i|D_{1i}, D_{2i}, T_i, \boldsymbol{A}_i)\middle| T_i, \boldsymbol{A}_i\right\}}{\pi_i(\boldsymbol{\xi}_0)} \\
&= \frac{\mathbb{E}(\pi_i D_{ki}|T_i, \boldsymbol{A}_i)}{\pi_i} = \rho_{ki}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&\mathbb{E}\left\{G_{i\ell r,\mathrm{IPW}}(\mu_0, \boldsymbol{\xi}_0)\right\} \\
&= \mathbb{E}\left\{\frac{V_i V_\ell V_r D_{1i} D_{2\ell} D_{3r}}{\pi_i(\boldsymbol{\xi}_0)\pi_\ell(\boldsymbol{\xi}_0)\pi_k(\boldsymbol{\xi}_0)}(I_{i\ell r} - \mu_0)\right\} \\
&= \mathbb{E}\left\{(I_{i\ell r} - \mu_0)\mathbb{E}\left(\frac{V_i D_{1i}}{\pi_i(\boldsymbol{\xi}_0)}\middle| T_i, \boldsymbol{A}_i\right)\mathbb{E}\left(\frac{V_\ell D_{2\ell}}{\pi_\ell(\boldsymbol{\xi}_0)}\middle| T_\ell, \boldsymbol{A}_\ell\right)\right. \\
&\quad \left. \times \mathbb{E}\left(\frac{V_r D_{3r}}{\pi_r(\boldsymbol{\xi}_0)}\middle| T_r, \boldsymbol{A}_r\right)\right\} \\
&= \mathbb{E}\left\{\rho_{1i}\rho_{2\ell}\rho_{3r}(I_{i\ell r} - \mu_0)\right\}.
\end{aligned}
$$

– *PDR estimator*

$$
\begin{aligned}
&\mathbb{E}\left\{D_{ki,\mathrm{PDR}}(\boldsymbol{\xi}_0)|T_i, \boldsymbol{A}_i\right\} \\
&= \mathbb{E}\left[\mathbb{E}\left\{\frac{V_i D_{ki}}{\pi_i(\boldsymbol{\xi}_0)} - \rho_{k(0)i}(\boldsymbol{\xi}_0)\left(\frac{V_i}{\pi_i(\boldsymbol{\xi}_0)} - 1\right)\middle| D_{1i}, D_{2i}, T_i, \boldsymbol{A}_i\right\}\middle| T_i, \boldsymbol{A}_i\right] \\
&= \mathbb{E}\left\{D_{ki}\mathbb{E}\left(\frac{V_i}{\pi_i(\boldsymbol{\xi}_0)}\middle| D_{1i}, D_{2i}, T_i, \boldsymbol{A}_i\right)\right. \\
&\quad \left. - \rho_{k(0)i}(\boldsymbol{\xi}_0)\mathbb{E}\left(\frac{V_i}{\pi_i(\boldsymbol{\xi}_0)} - 1\middle| D_{1i}, D_{2i}, T_i, \boldsymbol{A}_i\right)\middle| T_i, \boldsymbol{A}_i\right\} \\
&= \mathbb{E}(D_{ki}|T_i, \boldsymbol{A}_i) = \rho_{ki}.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
&\mathbb{E}\left\{G_{i\ell r,\mathrm{PDR}}(\mu_0, \boldsymbol{\xi}_0)\right\} \\
&= \mathbb{E}\left\{D_{1i,\mathrm{PDR}}(\boldsymbol{\xi}_0)D_{2\ell,\mathrm{PDR}}(\boldsymbol{\xi}_0)D_{3r,\mathrm{PDR}}(\boldsymbol{\xi}_0)(I_{i\ell r} - \mu_0)\right\} \\
&= \mathbb{E}\left[(I_{i\ell r} - \mu_0)\mathbb{E}\left\{D_{1i,\mathrm{PDR}}(\boldsymbol{\xi}_0)|T_i, \boldsymbol{A}_i\right\}\mathbb{E}\left\{D_{2\ell,\mathrm{PDR}}(\boldsymbol{\xi}_0)|T_\ell, \boldsymbol{A}_\ell\right\}\right. \\
&\quad \left. \times \mathbb{E}\left\{D_{3r,\mathrm{PDR}}(\boldsymbol{\xi}_0)|T_r, \boldsymbol{A}_r\right\}\right] \\
&= \mathbb{E}\left\{\rho_{1i}\rho_{2\ell}\rho_{3r}(I_{i\ell r} - \mu_0)\right\}.
\end{aligned}
$$

Table 4 Monte Carlo means (MCmean) for the maximum likelihood estimators of the elements of nuisance parameters λ, τ_π, τ_{ρ_1} and τ_{ρ_2}, when the true missing data mechanism is MAR

	True	MCmean		
n		250	500	1500
λ_1	0.00	0.41	0.23	0.03
λ_2	0.00	0.37	0.19	0.03
τ_{π_1}	-1.00	-1.41	-1.28	-1.03
τ_{π_2}	1.00	1.12	1.07	1.01
τ_{π_3}	-1.20	-1.26	-1.23	-1.21
$\tau_{\rho_{11}}$	15.00	15.64	15.22	15.06
$\tau_{\rho_{21}}$	-3.30	-3.43	-3.35	-3.31
$\tau_{\rho_{31}}$	-0.70	-0.81	-0.75	-0.71
$\tau_{\rho_{12}}$	9.50	9.92	9.65	9.54
$\tau_{\rho_{22}}$	-1.70	-1.77	-1.73	-1.70
$\tau_{\rho_{32}}$	-0.30	-0.38	-0.33	-0.31

Table 5 Monte Carlo means (MCmean), relative bias (Bias), Monte Carlo standard deviations (MCsd) and estimated standard deviations (Esd) for the proposed VUS estimators and the SPE estimator, when the true missing data mechanism is MAR

Sample size	Estimators	MCmean	Bias (%)	MCsd	Esd	CP (%)
$n = 250$	FI	0.773	-2.3	0.059	0.056	91.0
	MSI	0.771	-2.6	0.059	0.058	91.6
	IPW	0.774	-2.2	0.082	0.069	87.4
	PDR	0.766	-3.1	0.128	0.098	91.6
	SPE (MAR)	0.791	0.0	0.084	0.063	90.2
$n = 500$	FI	0.783	-1.0	0.037	0.037	94.6
	MSI	0.782	-1.1	0.037	0.037	95.0
	IPW	0.784	-0.9	0.057	0.048	90.3
	PDR	0.781	-1.3	0.069	0.060	94.5
	SPE (MAR)	0.792	0.2	0.050	0.040	92.9
$n = 1500$	FI	0.788	-0.3	0.020	0.019	93.8
	MSI	0.788	-0.4	0.020	0.019	94.4
	IPW	0.789	-0.3	0.034	0.028	92.9
	PDR	0.789	-0.3	0.035	0.029	95.0
	SPE (MAR)	0.791	0.0	0.030	0.026	93.0

CP denotes Monte Carlo coverages for the 95% confidence intervals, obtained through the normal approximation approach applied to each estimator

Appendix 3

Here, we present results of an additional simulation study, that covers the cases of: (i) missing at random (MAR) assumption for the missingness of the disease status; (ii) model misspecification in the estimation process.

In the study, the diagnostic test T, covariate A and the disease status $\mathcal{D}$ are generated as in scenario I of Sect. 4 of the paper. Moreover, the verification status V is:

Table 6 Monte Carlo means (MCmean) for the maximum likelihood estimators of the elements of nuisance parameters λ, τ_π, $\tau_{\rho 1}$ and $\tau_{\rho 2}$, when the estimated models are misspecified

	True	MCmean		
n		250	500	1500
λ_1	-2.00	-4.89	-4.55	-4.43
λ_2	-1.00	-2.08	-2.08	-2.03
$\tau_{\pi 1}$	2.00	4.87	4.54	4.43
$\tau_{\pi 2}$	0.50	0.90	0.85	0.82
$\tau_{\pi 3}$	-1.20	-1.37	-1.33	-1.31
$\tau_{\rho 11}$	15.00	15.79	15.14	14.90
$\tau_{\rho 21}$	-3.30	-3.43	-3.30	-3.25
$\tau_{\rho 31}$	-0.70	-0.46	-0.36	-0.36
$\tau_{\rho 12}$	9.50	9.88	9.53	9.35
$\tau_{\rho 22}$	-1.70	-1.75	-1.70	-1.67
$\tau_{\rho 32}$	-0.30	-0.36	-0.30	-0.29

Table 7 Monte Carlo means (MCmean), relative bias (Bias), Monte Carlo standard deviations (MCsd) and estimated standard deviations (Esd) for the proposed VUS estimators and the SPE estimator, when the estimated models are misspecified

Sample size	Estimators	MCmean	Bias (%)	MCsd	Esd	CP (%)
$n = 250$	FI	0.779	-1.5	0.043	0.041	92.6
	MSI	0.777	-1.8	0.043	0.042	93.4
	IPW	0.781	-1.3	0.059	0.043	88.4
	PDR	0.779	-1.6	0.108	0.046	92.4
	SPE (MAR)	0.780	-1.4	0.090	0.043	94.5
$n = 500$	FI	0.779	-1.5	0.030	0.029	93.5
	MSI	0.778	-1.6	0.030	0.030	93.3
	IPW	0.787	-0.5	0.042	0.032	89.8
	PDR	0.781	-1.2	0.050	0.033	94.8
	SPE (MAR)	0.782	-1.2	0.046	0.030	93.6
$n = 1500$	FI	0.781	-1.3	0.017	0.017	92.5
	MSI	0.780	-1.3	0.017	0.017	92.1
	IPW	0.786	-0.6	0.030	0.021	91.9
	PDR	0.781	-1.2	0.045	0.020	94.5
	SPE (MAR)	0.780	-1.4	0.026	0.018	92.8

CP denotes Monte Carlo coverages for the 95% confidence intervals, obtained through the normal approximation approach applied to each estimator

(i) generated as in scenario I with $h(T, A; \tau_\pi) = -1 + T - 1.2A$ and $\lambda_1 = \lambda_2 = 0$, i.e., under MAR assumption (verification rate roughly equal to 0.57);

(ii) generated as in scenario I, but models for the verification and disease processes used in the fitting procedure are misspecified, because the estimated verification model uses as predictors $T^{1/3}$ and $\log |A|$ instead of T and A, respectively, and the estimated disease model uses $A^{1/3}$ instead of A.

In both (i) and (ii), the true VUS is 0.791. We consider three different values of sample size, i.e., 250, 500 and 1500. The number of replications in each simulation experiment is set to 1000.

Simulation results are given in Tables 4 and 5, for the case (i), and in Tables 6 and 7, for the case (ii). As expected, in case (i) results show some bias of the proposed VUS estimators when compared to the SPE estimator which is properly used here. However, the bias decreases when the sample size increases. In case (ii), all estimators appear to be biased, even when the sample size is large. Moreover, although in the considered case the bias seems to stay on acceptable levels, we expect that, given the nature of the estimators, it could be even dramatically high with other kinds of misspecification.

References

Baker SG (1995) Evaluating multiple diagnostic tests with partial verification. Biometrics 51(1):330–337

Chi YY, Zhou XH (2008) Receiver operating characteristic surfaces in the presence of verification bias. J R Stat Soc Ser C (Appl Stat) 57(1):1–23

Fluss R, Reiser B, Faraggi D, Rotnitzky A (2009) Estimation of the ROC curve under verification bias. Biom J 51(3):475–490

Fluss R, Reiser B, Faraggi D (2012) Adjusting ROC curve for covariates in the presence of verification bias. J Stat Plan Inference 142(1):1–11

Kang L, Tian L (2013) Estimation of the volume under the ROC surface with three ordinal diagnostic categories. Comput Stat Data Anal 62:39–51

Li J, Zhou XH (2009) Nonparametric and semiparametric estimation of the three way receiver operating characteristic surface. J Stat Plan Inference 139(12):4133–4142

Little RJ, Rubin DB (2002) Statistical analysis with missing data. Wiley, New York

Liu D, Zhou XH (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

Nakas CT, Yiannoutsos CT (2004) Ordered multiple-class ROC analysis with continuous measurements. Stat Med 23(22):3437–3449

Rotnitzky A, Faraggi D, Schisterman E (2006) Doubly robust estimation of the area under the receiver-operating characteristic curve in the presence of verification bias. J Am Stat Assoc 101(475):1276–1288

Scurfield BK (1996) Multiple-event forced-choice tasks in the theory of signal detectability. J Math Psychol 40(3):253–269

To Duc K (2017) bcROCsurface: an R package for correcting verification bias in estimation of the ROC surface and its volume for continuous diagnostic tests. BMC Bioinform 18(1):503

To Duc K, Chiogna M, Adimari G (2016) Bias-corrected methods for estimating the receiver operating characteristic surface of continuous diagnostic tests. Electron J Stat 10(2):3063–3113

van der Vaart AW (2000) Asymptotic statistics. Cambridge University Press, Cambridge

Xiong C, van Belle G, Miller JP, Morris JC (2006) Measuring and estimating diagnostic accuracy when there are three ordinal diagnostic groups. Stat Med 25(7):1251–1273

Zhang Y, Alonzo TA (2018) Estimation of the volume under the receiver-operating characteristic surface adjusting for non-ignorable verification bias. Stat Methods Med Res 27(3):715–739

Zhou XH, Castelluccio P (2003) Nonparametric analysis for the ROC areas of two diagnostic tests in the presence of nonignorable verification bias. J Stat Plan Inference 115(1):193–213

Zhou XH, Castelluccio P (2004) Adjusting for non-ignorable verification bias in clinical studies for Alzheimer's disease. Stat Med 23(2):221–230

Zhou XH, Rodenberg CA (1998) Estimating an ROC curve in the presence of nonignorable verification bias. Commun Stat 27(3):273–285