Received: 21 January 2022 | Revised: 11 September 2022 | Accepted: 8 October 2022

DOI: 10.1002/bimj.202200021

Biometrical Journal

RESEARCH ARTICLE

Bayesian and influence function-based empirical likelihoods for inference of sensitivity to the early diseased stage in diagnostic tests

Yan Hai | **Shuangfei Shi** | **Gengsheng Qin** | **for the Alzheimer's Disease Neuroimaging**

Department of Mathematics and Statistics, Georgia State University, Atlanta, Georgia, USA

Correspondence
Gengsheng Qin for the Alzheimer's Disease Neuroimaging, Department of Mathematics and Statistics, Georgia State University, Atlanta GA 30303, USA.
Email: gqin@gsu.edu

Funding information
Alzheimer's Disease Neuroimaging Initiative; National Institutes of Health, Grant/Award Number: U01 AG024904; Department of Defense, Grant/Award Number: W81XWH-12-2-0012; National Institute on Aging; National Institute of Biomedical Imaging and Bioengineering; AbbVie; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd; Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; Transition

Abstract

In practice, a disease process might involve three ordinal diagnostic stages: the normal healthy stage, the early stage of the disease, and the stage of full development of the disease. Early detection is critical for some diseases since it often means an optimal time window for therapeutic treatments of the diseases. In this study, we propose a new influence function-based empirical likelihood method and Bayesian empirical likelihood methods to construct confidence/credible intervals for the sensitivity of a test to patients in the early diseased stage given a specificity and a sensitivity of the test to patients in the fully diseased stage. Numerical studies are performed to compare the finite sample performances of the proposed approaches with existing methods. The proposed methods are shown to outperform existing methods in terms of coverage probability. A real dataset from the Alzheimer's Disease Neuroimaging Initiative (ANDI) is used to illustrate the proposed methods.

KEYWORDS
Bayesian inference, confidence interval, empirical likelihood, influence function, sensitivity of the early stage

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Therapeutics; Canadian Institutes of Health Research

This article has earned an open data badge "**Reproducible Research**" for making publicly available the code necessary to reproduce the reported results. The results reported in this article were reproduced partially due to computational complexity.

1 | INTRODUCTION

Medical diagnostic tests are usually used to classify subjects into two stages, the nondiseased stage and the diseased stage. The accuracy of a diagnostic test can be evaluated based on its sensitivity and specificity, that is, the probabilities of correct diagnoses for diseased subjects and for nondiseased subjects, respectively. Both sensitivity and specificity are not fixed, and they are functions of a cutoff value of test results when the diagnostic test result is continuous. The receiver operating characteristic (ROC) curve, a plot of sensitivity versus (1 − specificity), is an important tool in measuring the accuracy of a diagnostic test in a two-class classification problem. It shows the trade-off between sensitivity and specificity as the cutoff point varies through all possible values of the diagnostic test. However, a disease process might be more complicated and involves three diagnostic stages: the nondiseased (healthy) stage, the early diseased stage, and the fully diseased stage. In practice, the sensitivity to the early diseased stage is more significant since early detection means optimal therapeutic treatments for some diseases. For example, mild cognitive impairment (MCI) is a transitional stage between the normal levels of cognition and dementia in Parkinson's disease (PD) (Petersen, 2004). And it is important to diagnose patients with MCI because they are more likely to progress to dementia (Aarsland & Kurz, 2010).

To be more specific, let Y_1, Y_2, and Y_3 be the continuous test results of a diagnostic test from the nondiseased, the early diseased, and the fully diseased groups, respectively, and F_1, F_2, and F_3 be the corresponding cumulative distribution functions of the test results. Assume that the higher values of the test results indicate greater severity of the disease. Given a pair of cutoff values c_1 and c_2 ($c_1 < c_2$), a subject is identified as nondiseased if the test result is smaller than c_1, as fully diseased if the test result is larger than c_2, and as early diseased if the test result is between c_1 and c_2. The specificity P_1 of the test is the correct classification rate of subjects in the nondiseased stage. The sensitivity P_2 of the test to the early diseased stage and the sensitivity P_3 of the test to the fully diseased stage are the correct classification rates of subjects in the early and fully diseased stages, respectively. P_1, P_2, and P_3 are defined as

$$\begin{aligned} P_1 &= F_1(c_1), \\ P_2 &= F_2(c_2) - F_2(c_1) = F_2[F_3^{-1}(1 - P_3)] - F_2[F_1^{-1}(P_1)], \\ P_3 &= 1 - F_3(c_2). \end{aligned} \tag{1}$$

Given P_1 and P_3, c_1, and c_2 can be determined if F_1 and F_3 are known distribution functions. Given the specificity P_1 and the sensitivity P_3 to the fully diseased stage, P_2, the sensitivity to the early diseased stage can be formulated as a function of P_1 and P_3, that is, $P_2 = P_2(P_1, P_3)$. The surface in the three-dimensional space (P_1, P_3, P_2) with $0 \le P_1, P_3 \le 1$ is called the ROC surface of the test. Scurfield (1996) proposed a ROC surface analysis. A few years later, Mossman (1999) independently proposed a similar analysis which was implemented using mathematica by Heckerling (2001). A nonparametric estimation of ROC surface was proposed by Nakas and Yiannoutsos (2004) and reshaped later by Xiong et al. (2006) and Li and Zhou (2009). For a disease such as PD, early detection of the disease is usually difficult and crucial because it usually means that patients can receive earlier treatments. Therefore, the probability associated with the early detection of the disease is a very important accuracy measure for the diagnostic test of the disease with three stages. Dong et al. (2011) first provided parametric and nonparametric confidence intervals for P_2, the sensitivity to the early diseased stage. However, their approaches fail if the normal assumption for test results cannot be satisfied. Dong and Tian (2015) proposed two empirical likelihood-based (EL) confidence intervals (ELP and ELB) for P_2 which can overcome the normal assumption, but their EL ratio statistic for P_2 follows a scaled chi-square distribution asymptotically. Thus, an extra step is required to estimate the unknown scale constant for inference on P_2 by using density estimation or the bootstrap method. Yu et al. (2012) proposed EL methods based on influence functions of parameters of interest. Hai et al. (2020)

TABLE 1 List of acronyms

Acronym	Explanation
EL	empirical likelihood
ANDI	Alzheimers Disease Neuroimaging Initiative
ROC	receiver operating characteristic
AUC	area under the ROC curve
ELP	plug-in empirical likelihood
ELB	empirical likelihood based on bootstrap
IF	influence function-based empirical likelihood
BEL1	Bayesian empirical likelihood 1
BEL2	Bayesian empirical likelihood 2
BIF1	Bayesian influence function-based empirical likelihood 1
BIF2	Bayesian influence function-based empirical likelihood 2
BpEL1	Bayesian pseudo empirical likelihood 1
BpEL2	Bayesian pseudo empirical likelihood 2
BpIF1	Bayesian pseudo influence function empirical likelihood 1
BpIF2	Bayesian pseudo influence function empirical likelihood 2

proposed an influence function-based EL method for inference of sensitivity of two-stage diagnostic tests. Motivated by Yu et al. (2012) and Hai et al. (2020), we propose an influence function and EL-based confidence interval for P_2 at a given value of the pair (P_1, P_3), that is, the sensitivity of the test to the early diseased stage at a specificity and a sensitivity of the test to the fully diseased stage. The idea is to replace the estimating function in the EL with an influence function of the sensitivity to the early stage. The corresponding empirical log-likelihood ratio statistic is shown to converge to a standard chi-square variable, which makes inference on P_2 more convenient. Despite wide applications in many areas, EL has only recently been used in Bayesian analysis. Lazar (2003) observed that properties of EL are, in many respects, similar to those of parametric likelihoods and explained why EL can be used in the Bayesian framework like parametric likelihoods. Motivated by Lazar's work, Bayesian empirical likelihood (BEL) approaches are also developed to construct credible intervals for P_2 in this article (see Table 1).

The article is organized as follows. In Section 2, we review Dong and Tian (2015)'s plug-in EL methods for interval estimation of sensitivity to the early stage. In Section 3, we introduce a new EL ratio statistic for sensitivity to the early stage based on influence function. In Section 4, we propose BEL methods based on influence function. In Section 5, we conduct simulation studies to compare the finite sample performance of the proposed methods with existing methods. In Section 6, a real dataset from the Alzheimer's Disease Neuroimaging Initiative (ANDI) is used to illustrate the proposed methods. In Section 7, we conduct a discussion on estimation of sensitivity to the early stage. The R codes and dataset used in this article are provided as Supplementary Information.

2 | PLUG-IN EMPIRICAL LIKELIHOOD METHOD

Without any underlying distribution assumptions, Dong and Tian (2015) proposed two empirical likelihood-based confidence intervals for P_2. They defined an indicator function Φ:

$$\Phi(X,Y,Z) = \begin{cases} 1, & X < Y < Z, \\ \frac{1}{2}, & X = Y < Z \text{ or } X < Y = Z, \\ \frac{1}{6}, & X = Y = Z, \\ 0, & \text{otherwise}, \end{cases} \tag{2}$$

and a random variable U at given P_1 and P_3:

$$U(Y) = \Phi[F_1^{-1}(P_1), Y, F_3^{-1}(1-P_3)], \tag{3}$$

where Y is a test result of a subject from the early diseased group.

For a given test result Y_2 of a test from the early diseased group, the value of $U(Y_2)$ can be interpreted as the placement value of Y_2 in the nondiseased and fully diseased populations. Let $\{Y_{1,j} : j = 1, 2, \dots, n_1\}$, $\{Y_{2,j} : j = 1, 2, \dots, n_2\}$, and $\{Y_{3,j} : j = 1, 2, \dots, n_3\}$ denote the n_1, n_2, and n_3 test results from the nondiseased, early stage, and fully diseased groups, respectively. From the following relationship between $U(Y_2)$ and P_2:

$$\begin{aligned} E(U(Y_2)) &= E\{\Phi[F_1^{-1}(P_1), Y_2, F_3^{-1}(1-P_3)]\} \\ &= P[F_1^{-1}(P_1) < Y_2 < F_3^{-1}(1-P_3)] \\ &= P[F_1^{-1}(P_1) < Y_2 \le F_3^{-1}(1-P_3)] \\ &= P_2, \end{aligned} \tag{4}$$

a plug-in empirical likelihood for P_2 can be defined as

$$L(P_2) = \sup_{\boldsymbol{p}} \left\{ \prod_{j=1}^{n_2} p_j : \sum_{j=1}^{n_2} p_j = 1, \sum_{j}^{n_2} p_j(\hat{U}_j - P_2) = 0 \right\}, \tag{5}$$

where $\hat{U}_j = \Phi[\hat{F}_1^{-1}(P_1), Y_{2,j}, \hat{F}_3^{-1}(1-P_3)]$, $j = 1, 2, \dots, n_2$, are plug-in estimates for $U(Y_{2,j})$s, $\hat{F}_1$ and $\hat{F}_3$ are the empirical distributions of F_1 and F_3, respectively. Using the Lagrange multiplier method, we can easily obtain the expression of p_j:

$$\tilde{p}_j = \frac{1}{n_2}\{1 + \tilde{\lambda}(P_2)(\hat{U}_j - P_2)\}^{-1} \tag{6}$$

where $\tilde{\lambda}(P_2)$ is the solution to

$$\frac{1}{n_2}\sum_{j=1}^{n_2} \frac{\hat{U}_j - P_2}{1 + \tilde{\lambda}(P_2)(\hat{U}_j - P_2)} = 0. \tag{7}$$

The corresponding plug-in empirical log-likelihood ratio for P_2 is

$$l(P_2) = 2\sum_{j=1}^{n_2} \log\{1 + \tilde{\lambda}(P_2)(\hat{U}_j - P_2)\}. \tag{8}$$

Dong and Tian (2015) have shown that the asymptotic distribution of $l(P_2)$ is a scaled (with unknown scale constant r_{P_1,P_2,P_3}) chi-square distribution with one degree of freedom. Thus, a $100(1-\alpha)\%$ level plug-in empirical likelihood-based confidence interval (ELP) for P_2 can be constructed as follows:

$$CI_1(P_2) = \{P_2 : \hat{r}_{P_1,P_2,P_3} l(P_2) \le \chi_1^2(1-\alpha)\}, \tag{9}$$

where $\chi_1^2(1-\alpha)$ is the $(1-\alpha)$th quantile of χ_1^2, and $\hat{r}_{P_1,P_2,P_3}$ is an estimate for the unknown scale constant:

$$\hat{r}_{P_1,P_2,P_3} = \frac{\hat{\hat{P}}_2(1-\hat{\hat{P}}_2)}{n_2\hat{\sigma}^2_{\hat{\hat{P}}_2}}, \tag{10}$$

where

$$\hat{\hat{P}}_2 = \frac{\sum_{j=1}^{n_2} I[\hat{F}_1^{-1}(P_1) < Y_{2,j} \le \hat{F}_3^{-1}(1-P_3)]}{n_2}, \tag{11}$$

$$\hat{\sigma}^2_{\hat{\hat{P}}_2} = \frac{\hat{\hat{P}}_2(1-\hat{\hat{P}}_2)}{n_2} + \frac{P_1(1-P_1)}{n_1}\frac{\hat{f}_2^2[\hat{F}_1^{-1}(P_1)]}{\hat{f}_1^2[\hat{F}_1^{-1}(P_1)]} + \frac{P_3(1-P_3)}{n_3}\frac{\hat{f}_2^2[\hat{F}_3^{-1}(1-P_3)]}{\hat{f}_3^2[\hat{F}_3^{-1}(1-P_3)]}, \tag{12}$$

and $\hat{f}_i$ is a kernel density estimate for the probability density function f_i of Y_i, $i = 1, 2, 3$.

To estimate the density function f_i, Dong and Tian (2015) used the "oversmoothed bandwidth selector" by Wand and Jones (1995) to select the bandwidth with a Gaussian kernel function. The performance of this method highly depends on the kernel density estimates whose bandwidths are chosen without a well-recognized standard. Therefore, they proposed the following bootstrap procedure to obtain a bootstrap estimate $\hat{\sigma}^{2*}_{\hat{P}_2}$ for the variance instead of $\hat{\sigma}^{2}_{\hat{P}_2}$ in Equation (10):

Step 1: Draw bootstrap resamples of sizes n_1, n_2, and n_3 with replacement from the nondiseased sample Y_{1j}s, the early diseased sample Y_{2j}s, and the fully diseased sample Y_{3j}s, respectively. Denote the bootstrap samples as $\{Y^b_{ij}\}$, $i = 1, 2, 3$, $j = 1, 2, \ldots, n_i$.

Step 2: Calculate the bootstrap version $\hat{P}^b_2$ of $\hat{P}_2$ according to Equation (11).

Step 3: Repeat the first two steps B times to obtain the bootstrap variance estimate of $\hat{P}_2$ which is defined as

$$\hat{\sigma}^{2*}_{\hat{P}_2} = \frac{1}{B-1}\sum_{b=1}^{B}(\hat{P}^b_2 - \bar{\hat{P}}^*_2)^2, \tag{13}$$

where $\bar{\hat{P}}^*_2 = \frac{1}{B}\sum_{b=1}^{B}\hat{P}^b_2$.

This leads to the second $100(1-\alpha)\%$ level empirical likelihood confidence interval (ELB) for P_2:

$$CI_2(P_2) = \{P_2 : r^*_{P_1,P_2,P_3} l(P_2) \le \chi^2_1(1-\alpha)\}, \tag{14}$$

where

$$r^*_{P_1,P_2,P_3} = \frac{\bar{\hat{P}}^*_2(1-\bar{\hat{P}}^*_2)}{n_2\hat{\sigma}^{2*}_{\hat{P}_2}}. \tag{15}$$

3 | INFLUENCE FUNCTION-BASED EMPIRICAL LIKELIHOOD METHOD

The application of the existing plug-in empirical likelihood-based ELP and ELB intervals for P_2 is computationally expensive due to the need for estimation of an unknown scale constant by density estimation and bootstrap process. Moreover, the finite sample performance of ELP and ELB intervals depends on estimation accuracy of the estimators for the scale constant. Hai et al. (2020) initially provided influence function-based EL methods for sensitivity of two-stage diagnostic tests. However, the extension of the results for the two stage in Hai et al. (2020) to three-stage tests is not trivial. Herein, we disclose a new influence function-based EL method, allowing construction of confidence intervals for sensitivity to the early diseased stage to remove the estimation of the unknown scale constant in Dong and Tian (2015).

We combine the samples $\{Y_{i,j} : i = 1, 2, 3; j = 1, 2, \ldots, n_i\}$ as

$$Z_l = \begin{cases} Y_{1,l}, & l = 1, \ldots, n_1, \\ Y_{2,l-n_1}, & l = 1+n_1, \ldots, n_1+n_2, \\ Y_{3,l-n_1-n_2}, & l = 1+n_1+n_2, \ldots, N, \end{cases} \tag{16}$$

where $N = n_1 + n_2 + n_3$.

Let $c_1 = F_1^{-1}(P_1)$, and $c_2 = F_3^{-1}(1-P_3)$. Define $\hat{F}_2$ as the empirical distribution of F_2, $\hat{c}_1 = \hat{F}_1^{-1}(P_1)$ (i.e., the P_1th sample quantile of $Y_{1,j}$s), and $\hat{c}_2 = \hat{F}_3^{-1}(1-P_3)$ (i.e., the $(1-P_3)$th sample quantile of $Y_{3,j}$s). Then the sensitivity P_2 of the test to the early diseased stage can be consistently estimated by

$$\tilde{P}_2 = \hat{F}_2(\hat{c}_2) - \hat{F}_2(\hat{c}_1) = \hat{F}_2[\hat{F}_3^{-1}(1-P_3)] - \hat{F}_2[\hat{F}_1^{-1}(P_1)] \equiv \hat{P}_2(P_1, P_3). \tag{17}$$

We have the following decomposition:

$$\begin{aligned} \tilde{P}_2 - P_2 &= [\hat{F}_2(\hat{c}_2) - \hat{F}_2(\hat{c}_1)] - [F_2(c_2) - F_2(c_1)] \\ &= [\hat{F}_2(\hat{c}_2) - \hat{F}_2(c_2)] - [\hat{F}_2(\hat{c}_1) - \hat{F}_2(c_1)] + \{[\hat{F}_2(c_2) - \hat{F}_2(c_1)] - [F_2(c_2) - F_2(c_1)]\} \\ &\equiv I_1 - I_2 + I_3. \end{aligned} \tag{18}$$

The third term of Equation (18) can be written as

$$\begin{aligned} I_3 &= [\hat{F}_2(c_2) - \hat{F}_2(c_1)] - [F_2(c_2) - F_2(c_1)] \\ &= \frac{1}{n_2}\sum_{j=1}^{n_2}[I(c_1 < Y_{2,j} \le c_2) - P_2] \\ &= \frac{1}{N}\sum_{l=n_1+1}^{n_1+n_2}\frac{N}{n_2}[I(c_1 < Z_l \le c_2) - P_2]. \end{aligned} \tag{19}$$

From the Bahadur representation for the sample quantiles $\hat{c}_1$ and $\hat{c}_2$ (Ghosh, 1971),

$$\begin{aligned} \hat{c}_1 - c_1 &= \frac{P_1 - \frac{1}{n_1}\sum_{j=1}^{n_1} I(Y_{1,j} \le c_1)}{f_1(c_1)} + o_p({n_1}^{-\frac{1}{2}}), \\ \hat{c}_2 - c_2 &= \frac{\frac{1}{n_3}\sum_{j=1}^{n_3} I(Y_{3,j} > c_2) - P_3}{f_3(c_2)} + o_p({n_3}^{-\frac{1}{2}}), \end{aligned} \tag{20}$$

it follows that

$$\begin{aligned} I_2 &= \hat{F}_2(\hat{c}_1) - \hat{F}_2(c_1) = \int [I(y \le \hat{c}_1) - I(y \le c_1)]d\hat{F}_2(y) \\ &= \int [I(y \le \hat{c}_1) - I(y \le c_1)]dF_2(y) + o_p(n_2^{-1/2}) \\ &= f_2(c_1)(\hat{c}_1 - c_1) + o_p(n_2^{-1/2} + n_1^{-1/2}) \\ &= -\frac{1}{n_1}\frac{f_2(c_1)}{f_1(c_1)}\sum_{j=1}^{n_1}[I(Y_{1,j} \le c_1) - P_1] + o_p(n_2^{-1/2} + n_1^{-1/2}) \\ &= -\frac{1}{N}\sum_{l=1}^{n_1}\frac{N}{n_1}\frac{f_2(c_1)}{f_1(c_1)}[I(Z_l \le c_1) - P_1] + o_p(N^{-1/2}), \end{aligned} \tag{21}$$

$$\begin{aligned} I_1 &= \hat{F}_2(\hat{c}_2) - \hat{F}_2(c_2) = \int [I(y \le \hat{c}_2) - I(y \le c_2)]d\hat{F}_2(y) \\ &= \int [I(y \le \hat{c}_2) - I(y \le c_2)]dF_2(y) + o_p(n_2^{-1/2}) \\ &= f_2(c_2)(\hat{c}_2 - c_2) + o_p(n_2^{-1/2} + n_3^{-1/2}) \\ &= \frac{1}{n_3}\frac{f_2(c_2)}{f_3(c_2)}\sum_{j=1}^{n_3}[I(Y_{3,j} > c_2) - P_3] + o_p(n_2^{-1/2} + n_3^{-1/2}) \\ &= \frac{1}{N}\sum_{l=n_1+n_2+1}^{N}\frac{N}{n_3}\frac{f_2(c_2)}{f_3(c_2)}[I(Z_l > c_2) - P_3] + o_p(N^{-1/2}). \end{aligned} \tag{22}$$

Therefore,

$$\tilde{P}_2 - P_2 = \frac{1}{N}\sum_{l=1}^{N} W_l(P_1, P_2, P_3) + o_p(N^{-1/2}), \tag{23}$$

where

$$W_l(P_1,P_2,P_3)=\begin{cases}\dfrac{N}{n_1}\dfrac{f_2(c_1)}{f_1(c_1)}[I(Z_l\le c_1)-P_1], & l=1,\dots,n_1,\\ \dfrac{N}{n_2}[I(c_1<Z_l\le c_2)-P_2], & l=n_1+1,\dots,n_1+n_2,\\ \dfrac{N}{n_3}\dfrac{f_2(c_2)}{f_3(c_2)}[I(Z_l>c_2)-P_3], & l=n_1+n_2+1,\dots,N,\end{cases} \tag{24}$$

is called the influence function of $\tilde{P}_2$.

From Equation (23), we can get the following asymptotic distribution of the empirical estimator $\tilde{P}_2$ for P_2.

Proposition 1. *Assume that F_1, F_2, and F_3 are continuous distribution functions with density functions f_1, f_2, and f_3, respectively, $f_2'(x)$ is bounded in neighborhoods of $c_1=F_1^{-1}(P_1)$, and $c_2=F_3^{-1}(1-P_3)$, $f_1(c_1)$, and $f_3(c_2)$ are strictly positive, $\frac{f_2(x)}{f_1(x)}$ is bounded in a neighborhood of c_1, and $\frac{f_2(x)}{f_3(x)}$ is bounded in a neighborhood of c_2. If $\lim\frac{n_1}{n_2}=\rho_1$ $(0<\rho_1<\infty)$, $\lim\frac{n_3}{n_2}=\rho_2$ $(0<\rho_2<\infty)$, and $\lim\frac{n_1}{n_3}=\rho_3$ $(0<\rho_3<\infty)$, then*

$$\sqrt{N}(\tilde{P}_2-P_2)\xrightarrow{d}N(0,\sigma^2), \tag{25}$$

where $N=n_1+n_2+n_3$, and

$$\begin{aligned}\sigma^2=&(1+\rho_1^{-1}+\rho_3^{-1})P_1(1-P_1)\frac{f_2^2[F_1^{-1}(P_1)]}{f_1^2[F_1^{-1}(P_1)]}+(1+\rho_1+\rho_2)P_2(1-P_2)\\&+(1+\rho_2^{-1}+\rho_3)P_3(1-P_3)\frac{f_2^2[F_3^{-1}(1-P_3)]}{f_3^2[F_3^{-1}(1-P_3)]}.\end{aligned} \tag{26}$$

Dong and Tian (2015) derived a similar result to Proposition 1 using a heuristic approach. To derive the influence function of $\tilde{P}_2$, we provide a rigorous proof of Proposition 1 in the Appendix. Proposition 1 can be used to construct a normal approximation-based confidence interval for P_2 if we can get a good estimate for σ^2. But estimating σ^2 involves estimation of unknown densities and quantiles. To avoid the complex variance estimation, we propose the following influence function-based EL method for inference on P_2.

Based on the influence function in (24), an EL for the sensitivity P_2 to the early diseased stage at a given pair of (P_1,P_3) can be defined as follows:

$$L_{IF}(P_2)=\sup_{\boldsymbol{p}}\left\{\prod_{l=1}^{N}p_l:\sum_{l=1}^{N}p_l=1,\sum_{l=1}^{N}p_l\hat{W}_l(P_1,P_2,P_3)=0\right\}, \tag{27}$$

where $\boldsymbol{p}=(p_1,\dots,p_N)$ is a probability vector and $\hat{W}_l(P_2)$ is the estimated influence function of $\tilde{P}_2$ which is given as follows:

$$\hat{W}_l(P_1,P_2,P_3)=\begin{cases}\dfrac{N}{n_1}\dfrac{\hat{f}_2(\hat{c}_1)}{\hat{f}_1(\hat{c}_1)}[I(Z_l\le\hat{c}_1)-P_1], & l=1,\dots,n_1,\\ \dfrac{N}{n_2}[I(\hat{c}_1<Z_l\le\hat{c}_2)-P_2], & l=n_1+1,\dots,n_1+n_2,\\ \dfrac{N}{n_3}\dfrac{\hat{f}_2(\hat{c}_2)}{\hat{f}_3(\hat{c}_2)}[I(Z_l>\hat{c}_2)-P_3], & l=n_1+n_2+1,\dots,N,\end{cases} \tag{28}$$

where $\hat{f}_i$ is a density estimator for f_i, $i=1,2,3$. We use the oversmoothed bandwidth selector to select the bandwidth for the Gaussian kernel function for f_i as described in Dong and Tian (2015).

By the Lagrange multiplier, the maximization of Equation (27) is achieved at

$$p_l = \frac{1}{N}[1 + \lambda \hat{W}_l(P_1, P_2, P_3)]^{-1}, l = 1, \ldots, N, \tag{29}$$

where $\lambda \equiv \lambda(P_2)$ is the solution to

$$\frac{1}{N}\sum_{l=1}^{N} \frac{\hat{W}_l(P_1, P_2, P_3)}{1 + \lambda \hat{W}_l(P_1, P_2, P_3)} = 0. \tag{30}$$

The corresponding empirical log-likelihood ratio statistic is

$$l_{IF}(P_2) = -\sum_{l=1}^{N} \log\{1 + \lambda \hat{W}_l(P_1, P_2, P_3)\}. \tag{31}$$

The following theorem establishes the asymptotic distribution of $l_{IF}(P_2)$, and a proof is given in the Appendix.

Theorem 1. *Assume that F_1, F_2, and F_3 are continuous distribution functions with density functions f_1, f_2, and f_3, respectively, $f_2'(x)$ is bounded in neighborhoods of $c_1 = F_1^{-1}(P_1)$ and $c_2 = F_3^{-1}(1 - P_3)$, $f_1(c_1)$ and $f_3(c_2)$ are strictly positive, $\frac{f_2(x)}{f_1(x)}$ is bounded in a neighborhood of c_1, and $\frac{f_2(x)}{f_3(x)}$ is bounded in a neighborhood of c_2. If $\lim \frac{n_1}{n_2} = \rho_1$ $(0 < \rho_1 < \infty)$, $\lim \frac{n_3}{n_2} = \rho_2$ $(0 < \rho_2 < \infty)$, and $\lim \frac{n_1}{n_3} = \rho_3$ $(0 < \rho_3 < \infty)$, and P_2^0 is the true value of sensitivity P_2 to the early diseased stage at a fixed level P_1 of specificity and P_3 of sensitivity to the fully diseased stage, then the asymptotic distribution of $-2l_{IF}(P_2^0)$ is a standard chi-squared distribution with one degree of freedom as $n_1, n_2, n_3 \to \infty$.*

From Theorem 1, a $100(1 - \alpha)\%$ level influence function-based (IF) EL confidence interval for P_2 can be constructed as follows:

$$CI_{IF}(P_2) = \{P_2 : -2l_{IF}(P_2) \le \chi_1^2(1 - \alpha)\}. \tag{32}$$

This IF-based EL interval for P_2 can be easily obtained using the algorithm for the standard EL method (see Hall & La Scala, 1990) without estimation of the variance in Proposition 1 and the scale constant in the plug-in EL-based method of Dong and Tian (2015).

Remark. In practice, it is possible that $1 + \lambda \hat{W}_l(P_1, P_2, P_3) < 0$ for an l $(l = 1, \ldots, N)$ with a sample of real test results. Under this situation, the adjusted empirical likelihood and the numerical algorithm by Chen et al. (2008) can be used for the calculation of the empirical log-likelihood ratio statistic of the sensitivity P_2.

4 | BAYESIAN EMPIRICAL LIKELIHOOD METHOD

BEL can be used as the basis for Bayesian inference. As Lazar (2003) pointed out that EL has many asymptotic properties similar to those of parametric likelihoods, BEL methods are naturally used to quantify uncertainty and can have good small sample properties. With a similar motivation to BEL inference for sensitivity of two-stage diagnostic tests in Hai et al. (2020), we propose two types of BEL methods for better interval estimation of sensitivity of a three-stage diagnostic test to early stage disease in this section.

4.1 | Bayesian empirical likelihood based on sensitivity

We follow Lazar (2003)'s idea to combine empirical likelihood $L(P_2)$ with prior $\pi(P_2)$ on P_2 by the Bayesian theorem to obtain a posterior:

$$\pi(P_2|data) \propto L(P_2)\pi(P_2). \tag{33}$$

We consider reference priors, originally introduced by Bernardo (1979), and further developed in Berger and Bernardo (1992), on P_2 in this study. Reference priors only depend on the assumed model and the available data. In our problem, we do not have a parametric model. Therefore, we follow Clarke and Yuan (2010) to derive reference priors for EL. The following proposition gives the reference priors for the BEL method, where $[L(P_2)]^{\hat{P}_{P_1,P_2,P_3}}$ in the plug-in EL is used as the likelihood.

Proposition 2. *The reference prior based on the relative entropy for the plug-in EL is*

$$\pi_{EL,1}(P_2) = \beta(\frac{3}{2}, \frac{3}{2}), \tag{34}$$

and the reference prior based on the Hellinger distance is

$$\pi_{EL,2}(P_2) = \beta(\frac{1}{2}, \frac{1}{2}), \tag{35}$$

where $\beta(a, b)$ is the beta distribution with parameters a and b.

The corresponding posterior is

$$\pi_{EL}(P_2|Y) \propto \prod_{j=1}^{n_2}[1 + \tilde{\lambda}(P_2)(\hat{U}_j - P_2)]^{-\hat{P}_{P_1,P_2,P_3}}\pi_{EL}(P_2), \tag{36}$$

where $\pi_{EL}(P_2) = \pi_{EL,1}(P_2)$, or $\pi_{EL,2}(P_2)$, and $\hat{U}_j = \Phi[\hat{F}_1^{-1}(P_1), Y_{2,j}, \hat{F}_3^{-1}(1 - P_3)]$, $j = 1, 2, \ldots, n_2$. Based on these posteriors, we can calculate equal-tail credible intervals for P_2. These two new methods are called as Bayesian empirical likelihood 1 (BEL1) and Bayesian empirical likelihood 2 (BEL2).

Similarly, to construct Bayesian credible intervals for P_2 based on the influence function, we propose the following reference priors on P_2 for the BEL based on the influence function $W_l(P_1, P_2, P_3)$ in Equation (24):

$$\pi_{IF,1}(P_2) \propto \left[\frac{1}{n_1}P_1(1 - P_1)\frac{f_2^2[F_1^{-1}(P_1)]}{f_1^2[F_1^{-1}(P_1)]} + \frac{1}{n_2}P_2(1 - P_2) + \frac{1}{n_3}P_3(1 - P_3)\frac{f_2^2[F_3^{-1}(1 - P_3)]}{f_3^2[F_3^{-1}(1 - P_3)]}\right]^{\frac{1}{2}}, \tag{37}$$

and

$$\pi_{IF,2}(P_2) \propto \left[\frac{1}{n_1}P_1(1 - P_1)\frac{f_2^2[F_1^{-1}(P_1)]}{f_1^2[F_1^{-1}(P_1)]} + \frac{1}{n_2}P_2(1 - P_2) + \frac{1}{n_3}P_3(1 - P_3)\frac{f_2^2[F_3^{-1}(1 - P_3)]}{f_3^2[F_3^{-1}(1 - P_3)]}\right]^{-\frac{1}{2}}. \tag{38}$$

Both priors are proper since $\pi_{IF,1}(P_2)$ is bounded by a constant and $\pi_{IF,2}(P_2)$ is bounded by a beta distribution. In practice, we use $\hat{W}_l(P_1, P_2, P_3)$ to estimate the influence function $W_l(P_1, P_2, P_3)$ and replace f_1, f_2, f_3, c_1, and c_2 with their estimates since they are generally unknown. The posterior based on this approach is then

$$\pi_{IF}(P_2|Z) \propto \prod_{l=1}^{N}[1 + \lambda(P_2)\hat{W}_l(P_1, P_2, P_3)]^{-1}\pi_{IF}(P_2), \tag{39}$$

where $\pi_{IF}(P_2) = \pi_{IF,1}(P_2)$, or $\pi_{IF,2}(P_2)$. Based on these posteriors, we can calculate equal-tail credible intervals for P_2. Therefore, we have two more new intervals for P_2: the first one is called the Bayesian influence function-based empirical likelihood 1 (BIF1) interval and the second one is called the Bayesian influence function-based empirical likelihood 2 (BIF2) interval.

4.2 | Bayesian pseudo empirical likelihood based on probability vector

Rao and Wu (2010) proposed Bayesian pseudo-EL (BpEL) methods for complex surveys. Inspired by their methods, in this section, we develop BEL based on probability vector $(p_1, \dots, p_l)$ instead of P_2. We treat $(p_1, \dots, p_l)$ as unknown parameters and the EL function is defined as

$$L_{EL}(p_1, \dots, p_l) = \prod_{i=1}^{l} p_i, \tag{40}$$

where $l = n_2$ for the plug-in EL, and $l = N$ for the influence function-based EL. Consider Dirichlet prior $D(\alpha_1, \dots, \alpha_l)$ on $(p_1, \dots, p_l)$:

$$\pi(p_1, \dots, p_l) = c(\alpha_1, \dots, \alpha_l) \prod_{i=1}^{l} p_i^{\alpha_i - 1}, \tag{41}$$

where $c(\alpha_1, \dots, \alpha_l) = \Gamma(\sum_{i=1}^{l} \alpha_i) / \prod_{i=1}^{l} \Gamma(\alpha_i)$. The posterior distribution of $(p_1, \dots, p_l)$ given the data is Dirichlet $D(1 + \alpha_1, \dots, 1 + \alpha_l)$ and is given by

$$\pi(p_1, \dots, p_l | data) = c(1 + \alpha_1, \dots, 1 + \alpha_n) \prod_{i=1}^{l} p_i^{\alpha_i}. \tag{42}$$

The posterior of the early stage sensitivity P_2 satisfies the following equation:

$$\sum_{i=1}^{l} p_i \hat{Q}_i(P_2) = 0, \tag{43}$$

where $\hat{Q}_i(P_2)$ is an estimating/influence function and $(p_1, \dots, p_l)$ follows the Dirichlet distribution $D(1 + \alpha_1, \dots, 1 + \alpha_l)$. In practice, we can generate samples of $(p_1, \dots, p_l)$ from $D(1 + \alpha_1, \dots, 1 + \alpha_l)$, and by solving Equation (43), we get the posterior samples of P_2. Based on these posterior samples, we can calculate the equal-tail credible intervals for sensitivity P_2.

Similar to Section 4.1, we consider two types of EL: EL in Equation (5) as in the plug-in EL and the influence function EL in Equation (27). We call them BpEL and Bayesian pseudo Influence Function-based EL (BpIF), respectively. For BpEL, we use $\hat{W}_{ELP}(P_1, P_2, P_3) \equiv \hat{U}_j - P_2$ to replace $\hat{Q}_i(P_2)$ in Equation (43) and consider $D(r^*, \dots, r^*)$ and $D(r^* + \frac{1}{n_2}, \dots, r^* + \frac{1}{n_2})$ as the priors (labeled as BpEL1 and BpEL2, respectively), where $r^* = \hat{r}_{P_1,P_2,P_3}$ is the estimate defined in Equation (10) in Section 2 for the scale constant. For BpIF, we use $\hat{W}_l(P_1, P_2, P_3)$ to replace $\hat{Q}_i(P_2)$ in Equation (43) and consider $D(1, \dots, 1)$ and $D(1 + \frac{1}{N}, \dots, 1 + \frac{1}{N})$ as the priors (labeled as BpIF1 and BpIF2, respectively).

5 | SIMULATION STUDY

Simulation studies are conducted to examine the finite sample performance of the proposed intervals for sensitivity of a test to early stage disease: influence function-based EL (IF) interval, Bayesian influence function EL (BIF1 and BIF2) intervals with reference priors $\pi_{IF,1}(P_2)$ and $\pi_{IF,2}(P_2)$, Bayesian EL (BEL1 and BEL2) intervals with reference priors $\pi_{EL,1}(P_2)$ and $\pi_{EL,2}(P_2)$, Bayesian pseudo influence function EL (BpIF1 and BpIF2) intervals, and Bayesian pseudo EL (BpEL1 and BpEL2) intervals. We compare them to the existing ELP and ELB intervals proposed by Dong and Tian (2015). For the purpose of comparison, we also include the normal approximation (NA) interval based on Proposition 1 by using $\tilde{P}_2 = \hat{P}_2$ as the point estimate and $\hat{\sigma}^2_{\hat{P}_2}$ in Equation (12) as the variance estimate.

We evaluate these approaches under scenarios where the underlying distributions are normal distributions, beta distributions, and the combined scenario where the normality assumptions cannot be met, that is, a gamma distribution for the nondiseased, a log-normal distribution for the early diseased, and a Weibull distribution for the fully diseased groups.

TABLE 2 Coverage probabilities of various 95% level intervals for P_2 under Normal distributions with $P_2 = 0.8$. NA, ELB, ELP, and IF intervals are confidence intervals, the other intervals are credible intervals

Sample size	NA	ELB	ELP	BEL1	BEL2	BpEL1	BpEL2	IF	BIF1	BIF2	BpIF1	BpIF2
$(\mu_1,\sigma_1)=(0,1),(\mu_2,\sigma_2)=(3,1.2),(\mu_3,\sigma_3)=(5.858,2),P_2=0.8,P_1=P_3=0.8$												
(10,10,10)	0.798	0.990	0.991	0.982	0.988	0.983	0.966	0.872	0.927	0.919	0.843	0.838
(30,30,30)	0.896	0.853	0.845	0.970	0.985	0.960	0.949	0.935	0.939	0.939	0.926	0.928
(50,30,30)	0.872	0.827	0.827	0.957	0.960	0.960	0.947	0.914	0.923	0.925	0.918	0.918
(50,50,50)	0.909	0.870	0.862	0.964	0.963	0.939	0.935	0.933	0.953	0.943	0.925	0.927
(100,100,100)	0.939	0.944	0.944	0.950	0.953	0.948	0.944	0.947	0.948	0.946	0.926	0.922
(100,50,50)	0.914	0.867	0.867	0.959	0.971	0.937	0.932	0.942	0.952	0.945	0.919	0.917
(100,100,50)	0.912	0.911	0.905	0.960	0.962	0.932	0.926	0.936	0.956	0.944	0.919	0.919
$(\mu_1,\sigma_1)=(0,1),(\mu_2,\sigma_2)=(4,1.2),(\mu_3,\sigma_3)=(7.625,2),P_2=0.8,P_1=P_3=0.9$												
(10,10,10)	0.668	0.997	0.994	0.986	0.997	0.989	0.967	0.643	0.839	0.677	0.730	0.723
(30,30,30)	0.828	0.776	0.783	0.974	0.975	0.930	0.913	0.883	0.852	0.850	0.877	0.874
(50,30,30)	0.806	0.781	0.759	0.965	0.961	0.920	0.903	0.858	0.836	0.825	0.856	0.854
(50,50,50)	0.852	0.799	0.796	0.960	0.948	0.914	0.904	0.885	0.908	0.891	0.889	0.887
(100,100,100)	0.881	0.906	0.906	0.939	0.930	0.925	0.914	0.902	0.920	0.912	0.923	0.923
(100,50,50)	0.857	0.812	0.791	0.952	0.953	0.892	0.884	0.889	0.899	0.891	0.892	0.891
(100,100,50)	0.857	0.865	0.842	0.948	0.934	0.878	0.870	0.894	0.901	0.893	0.867	0.866

Note: Abbreviations used in Tables 2–7 are defined in Table 1.

Sample sizes (n_1, n_2, n_3) are set as (10, 10, 10), (30, 30, 30), (50, 30, 30), (50, 50, 50), (100, 100, 100), (100, 50, 50), and (100, 100, 50). With a fixed 80% or 90% specificity and a fixed 80% or 90% sensitivity to the fully diseased stage, the parameters for the distributions are chosen accordingly so that P_2 equals 80% or 90%. Under each distribution scenario, there are four settings corresponding to different levels of P_1 and P_3, and the true value of P_2: (i) $P_1 = P_3 = 0.8$ and $P_2 = 0.8$, (ii) $P_1 = P_3 = 0.9$ and $P_2 = 0.8$, (iii) $P_1 = P_3 = 0.8$ and $P_2 = 0.9$, and (iv) $P_1 = P_3 = 0.9$ and $P_2 = 0.9$. Under each setting, 5000 random samples are generated. With a simulated sample, it is possible that $1 + \lambda \hat{W}_l(P_1, P_2, P_3) < 0$ for an l ($l = 1, \ldots, N$), which can occur with not insignificant rate (possibly due to (1) the poor density estimates and (2) the poor estimates for c_1 and c_2) in the simulation runs when sample sizes are small (i.e., (10,10,10), (30,30,30), (50,50,50)), and it can have some impacts in the practical use of the IF-based EL intervals (under this situation, the adjusted empirical likelihood by Chen et al., 2008, can be used for the calculation of the empirical log-likelihood ratio statistic of the sensitivity P_2. See the Remark at the end of Section 3). We also notice that the new IF-based EL method as well as the existing EL-based ELP and ELB methods all present problems of the slow convergence rate (see Figures A1–A3 in the Appendix for some QQ plots of the empirical log-likelihood ratio statistics). In our simulation study, $-2l_{IF}(P_2)$ is set to be ∞ when one of the terms $\{1 + \lambda \hat{W}_l(P_1, P_2, P_3)\}$s is negative. The simulation results are presented in Tables 2–7 and Figures 1–3.

Under the scenario with normal distributions (see Tables 2 and 3 and Figure 1), we observe that the IF-related confidence intervals (IF, BIF1, BIF2, BpIF1, and BpIF2) generally undercover P_2 (the sensitivity of a test to early stage disease). The existing confidence intervals (NA, ELP, and ELB) also undercover P_2 except the small sample size (10, 10, 10). New methods always have better performance compared to NA, ELP, and ELB methods in all settings considered here. In particular, BEL1 has the best overall performance in terms of coverage probability close to 95%. Bayesian and Bayesian pseudo approaches generally have similar or improved performance over ELP, ELB, or IF. IF-function related methods have poor performance when sample size is small. The possible reason is that small sample sizes result in the poor density estimation involved in the IF-related methods. Comparing the results from the normal distribution settings (i) and (iii) with those from the normal distribution settings (ii) and (iv) which have higher specificity(P_1) and sensitivity(P_3) to the fully diseased group, we can see that the performance of NA, ELB, ELP, and IF related methods all depends on the degree of separation of test outcomes in the fully diseased, early diseased, and nondiseased groups. Under the higher specificity and sensitivity to the fully diseased group, they have lower coverage probabilities. However, the performance of BEL1, BEL2, BpEL1, and BpEL2 does not obviously change with different P_1 and P_3. Comparing the results from the normal distribution setting (i) with (iii) or normal distribution setting (ii) with (iv) which have fixed P_1 and P_3 but higher true value of P_2, we observe that ELP, ELB, BEL2, BpEL1, IF, BIF1, and BIF2 generally have similar or poorer finite sample performance with the higher true value of P_2 and other methods perform similarly or slightly better. For example, when sample size is (100,100,100)

TABLE 3 Coverage probabilities of various 95% level intervals for P_2 under Normal distributions with $P_2 = 0.9$. NA, ELB, ELP, and IF are confidence intervals, the other intervals are credible intervals

Sample size	NA	ELB	ELP	BEL1	BEL2	BpEL1	BpEL2	IF	BIF1	BIF2	BpIF1	BpIF2
$(\mu_1,\sigma_1)=(0,1),(\mu_2,\sigma_2)=(3,1.2),(\mu_3,\sigma_3)=(6.515,2),P_2=0.9,P_1=P_3=0.8$												
(10,10,10)	0.762	0.980	0.982	0.931	0.979	0.975	0.958	0.866	0.915	0.929	0.829	0.824
(30,30,30)	0.914	0.827	0.795	0.950	0.985	0.985	0.977	0.906	0.918	0.908	0.932	0.932
(50,30,30)	0.912	0.805	0.781	0.931	0.977	0.985	0.981	0.871	0.868	0.867	0.936	0.935
(50,50,50)	0.918	0.816	0.809	0.952	0.982	0.966	0.963	0.931	0.915	0.914	0.944	0.943
(100,100,100)	0.942	0.908	0.905	0.950	0.970	0.963	0.961	0.961	0.966	0.963	0.955	0.954
(100,50,50)	0.926	0.812	0.806	0.946	0.980	0.970	0.964	0.931	0.942	0.934	0.942	0.944
(100,100,50)	0.924	0.882	0.876	0.956	0.977	0.954	0.953	0.956	0.919	0.917	0.945	0.945
$(\mu_1,\sigma_1)=(0,1),(\mu_2,\sigma_2)=(4,1.2),(\mu_3,\sigma_3)=(8.189,2),P_2=0.9,P_1=P_3=0.9$												
(10,10,10)	0.619	0.973	0.988	0.944	0.987	0.977	0.960	0.591	0.880	0.670	0.714	0.713
(30,30,30)	0.849	0.769	0.778	0.948	0.983	0.989	0.986	0.804	0.692	0.708	0.887	0.886
(50,30,30)	0.845	0.774	0.781	0.937	0.982	0.986	0.979	0.793	0.691	0.713	0.864	0.864
(50,50,50)	0.864	0.734	0.757	0.952	0.984	0.963	0.955	0.888	0.802	0.808	0.901	0.905
(100,100,100)	0.899	0.845	0.867	0.944	0.953	0.945	0.940	0.929	0.879	0.875	0.938	0.936
(100,50,50)	0.882	0.763	0.753	0.935	0.973	0.958	0.953	0.887	0.870	0.866	0.906	0.905
(100,100,50)	0.878	0.804	0.789	0.957	0.963	0.911	0.905	0.916	0.794	0.797	0.891	0.891

TABLE 4 Coverage probabilities of various 95% level intervals for P_2 under Beta distributions with $P_2 = 0.8$. NA, ELB, ELP, and IF are confidence intervals, the other intervals are credible intervals

Sample size	NA	ELB	ELP	BEL1	BEL2	BpEL1	BpEL2	IF	BIF1	BIF2	BpIF1	BpIF2
$(\alpha_1,\beta_1)=(2,6),(\alpha_2,\beta_2)=(8,6),(\alpha_3,\beta_3)=(21.3,6),P_2=0.8,P_1=P_3=0.8$												
(10,10,10)	0.865	0.990	0.986	0.973	0.987	0.988	0.980	0.885	0.938	0.912	0.855	0.847
(30,30,30)	0.931	0.907	0.894	0.968	0.976	0.968	0.960	0.961	0.956	0.953	0.950	0.949
(50,30,30)	0.925	0.904	0.887	0.959	0.975	0.970	0.960	0.962	0.956	0.953	0.942	0.943
(50,50,50)	0.930	0.926	0.934	0.964	0.961	0.953	0.946	0.948	0.961	0.951	0.941	0.942
(100,100,100)	0.944	0.972	0.962	0.963	0.957	0.955	0.954	0.954	0.962	0.957	0.963	0.964
(100,50,50)	0.936	0.946	0.938	0.963	0.969	0.964	0.957	0.955	0.957	0.949	0.954	0.953
(100,100,50)	0.940	0.970	0.961	0.964	0.957	0.946	0.944	0.948	0.973	0.957	0.961	0.961
$(\alpha_1,\beta_1)=(1,6),(\alpha_2,\beta_2)=(6,6),(\alpha_3,\beta_3)=(15.2,6),P_2=0.8,P_1=P_3=0.9$												
(10,10,10)	0.791	0.995	0.992	0.990	0.996	0.994	0.984	0.706	0.772	0.753	0.717	0.714
(30,30,30)	0.900	0.859	0.808	0.981	0.965	0.934	0.918	0.894	0.894	0.879	0.851	0.848
(50,30,30)	0.916	0.877	0.852	0.979	0.965	0.931	0.917	0.914	0.915	0.897	0.876	0.878
(50,50,50)	0.918	0.882	0.844	0.965	0.938	0.910	0.899	0.882	0.918	0.884	0.875	0.868
(100,100,100)	0.940	0.963	0.924	0.953	0.934	0.921	0.915	0.911	0.917	0.904	0.907	0.901
(100,50,50)	0.925	0.924	0.902	0.962	0.960	0.937	0.926	0.931	0.913	0.906	0.918	0.916
(100,100,50)	0.919	0.962	0.928	0.959	0.940	0.916	0.913	0.908	0.964	0.933	0.907	0.904

with $P_1 = P_3 = 0.8$, the coverage probability of ELB dropped from 0.944 to 0.906 when the true value of P_2 is changed from 0.8 to 0.9.

The simulation results under the Beta distribution setting are reported in Tables 4 and 5 and Figure 2. Similar to normal distribution settings, we observe that NA, ELP, ELB intervals, and FI-related intervals (IF, BIF1, BIF2, BpIF1, and BpIF2) generally have undercoverage problems. New intervals always have better performance, and BEL1 interval has the best overall performance in terms of coverage probability close to 95% except the first setting where IF-related intervals work well. The performance of NA, ELB, ELP, BpEL1, BpEL2, and IF related intervals all depend on the degree of separation of test outcomes in the fully diseased, early diseased, and nondiseased groups. Under higher specificity and sensitivity to the fully diseased group, they have slightly lower coverage probabilities. However, performances of BEL1 and BEL2 intervals

TABLE 5 Coverage probabilities of various 95% level intervals for P_2 under Beta distributions with $P_2 = 0.9$. NA, ELB, ELP, and IF are confidence intervals, the other intervals are credible intervals

Sample size	NA	ELB	ELP	BEL1	BEL2	BpEL1	BpEL2	IF	BIF1	BIF2	BpIF1	BpIF2
$(\alpha_1,\beta_1)=(2,6),(\alpha_2,\beta_2)=(8,6),(\alpha_3,\beta_3)=(31.8,6),P_2=0.9,P_1=P_3=0.8$												
(10,10,10)	0.728	0.985	0.981	0.929	0.988	0.988	0.983	0.813	0.921	0.893	0.777	0.772
(30,30,30)	0.920	0.817	0.822	0.955	0.984	0.987	0.987	0.850	0.850	0.850	0.921	0.920
(50,30,30)	0.923	0.824	0.805	0.947	0.974	0.983	0.981	0.860	0.848	0.849	0.914	0.916
(50,50,50)	0.914	0.815	0.815	0.951	0.982	0.969	0.965	0.918	0.909	0.902	0.919	0.918
(100,100,100)	0.940	0.922	0.919	0.955	0.969	0.959	0.956	0.956	0.962	0.954	0.963	0.963
(100,50,50)	0.940	0.843	0.836	0.954	0.980	0.961	0.956	0.936	0.967	0.950	0.947	0.948
(100,100,50)	0.937	0.910	0.917	0.953	0.975	0.964	0.964	0.952	0.923	0.921	0.963	0.961
$(\alpha_1,\beta_1)=(1,6),(\alpha_2,\beta_2)=(6,6),(\alpha_3,\beta_3)=(20.4,6),P_2=0.9,P_1=P_3=0.9$												
(10,10,10)	0.787	0.966	0.989	0.956	0.986	0.987	0.972	0.636	0.896	0.721	0.686	0.679
(30,30,30)	0.889	0.786	0.800	0.960	0.986	0.990	0.989	0.816	0.808	0.801	0.901	0.901
(50,30,30)	0.906	0.789	0.782	0.945	0.974	0.981	0.977	0.820	0.807	0.811	0.901	0.902
(50,50,50)	0.903	0.745	0.803	0.958	0.980	0.950	0.944	0.909	0.894	0.885	0.918	0.916
(100,100,100)	0.931	0.877	0.924	0.961	0.973	0.962	0.958	0.954	0.969	0.956	0.951	0.950
(100,50,50)	0.911	0.816	0.827	0.959	0.979	0.957	0.954	0.929	0.929	0.910	0.929	0.929
(100,100,50)	0.929	0.874	0.913	0.966	0.968	0.954	0.955	0.944	0.917	0.915	0.941	0.939

TABLE 6 Coverage probabilities of various 95% level intervals for P_2 under combined distributions with $P_2 = 0.8$. NA, ELB, ELP, and IF are confidence intervals, the other intervals are credible intervals

Sample size	NA	ELB	ELP	BEL1	BEL2	BpEL1	BpEL2	IF	BIF1	BIF2	BpIF1	BpIF2
$Gamma(\alpha,\beta)=(4,10),LN(\mu,\sigma)=(1,0.5),Weibull(a,b)=(4.07,6),P_2=0.8,P_1=P_3=0.8$												
(10,10,10)	0.852	0.983	0.982	0.957	0.979	0.986	0.970	0.876	0.914	0.895	0.869	0.871
(30,30,30)	0.920	0.902	0.911	0.968	0.977	0.952	0.940	0.944	0.936	0.934	0.936	0.934
(50,30,30)	0.921	0.903	0.910	0.970	0.970	0.947	0.938	0.940	0.929	0.934	0.938	0.932
(50,50,50)	0.926	0.930	0.938	0.951	0.959	0.948	0.942	0.948	0.958	0.948	0.945	0.944
(100,100,100)	0.938	0.957	0.960	0.961	0.957	0.957	0.953	0.956	0.959	0.958	0.950	0.948
(100,50,50)	0.928	0.924	0.941	0.964	0.952	0.941	0.939	0.939	0.959	0.955	0.938	0.937
(100,100,50)	0.914	0.946	0.956	0.964	0.959	0.954	0.957	0.953	0.950	0.941	0.948	0.952
$Gamma(\alpha,\beta)=(4,10),LN(\mu,\sigma)=(0.5,0.5),Weibull(a,b)=(2.8,6),P_2=0.8,P_1=P_3=0.9$												
(10,10,10)	0.754	0.996	0.997	0.991	0.995	0.993	0.989	0.719	0.795	0.721	0.788	0.787
(30,30,30)	0.886	0.850	0.852	0.991	0.981	0.955	0.949	0.928	0.922	0.916	0.908	0.910
(50,30,30)	0.870	0.840	0.849	0.982	0.973	0.946	0.934	0.910	0.897	0.896	0.901	0.897
(50,50,50)	0.907	0.886	0.905	0.971	0.967	0.945	0.939	0.934	0.949	0.940	0.928	0.928
(100,100,100)	0.917	0.954	0.952	0.970	0.962	0.949	0.946	0.944	0.955	0.949	0.937	0.938
(100,50,50)	0.891	0.871	0.890	0.968	0.953	0.931	0.926	0.928	0.939	0.930	0.921	0.920
(100,100,50)	0.887	0.926	0.916	0.970	0.947	0.932	0.929	0.930	0.938	0.925	0.924	0.924

do not obviously change when P_1 and P_3 increase. Comparing the results from the beta distribution setting (i) with (iii) or beta distribution setting (ii) with (iv) which have fixed P_1 and P_3 but higher true value of P_2, we note that BEL1, BpEL1, and BpEL2 intervals generally have similar or better finite sample performance with higher true value of P_2, and other intervals perform similarly or slightly worse except BpIF1 and BpIF2, which have no obvious trend. Specifically, BpIF1 and BpIF2 intervals perform very well when $P_1 = P_3 = 0.8$ and true $P_2 = 0.8$ and have similar or slightly lower coverage probabilities than that with $P_2 = 0.9$. However, BpIF1 and BpIF2 intervals perform better when the true value of P_2 is changed from 0.8 to 0.9.

The simulation results under the combined distribution settings are reported in Tables 6 and 7 and Figure 3. Clearly, the new methods always have better performance compared to ELB and ELP. However, BEL1 does not always have the best

TABLE 7 Coverage probabilities of various 95% level intervals for P_2 under combined distributions with $P_2 = 0.9$. NA, ELB, ELP, and IF are confidence intervals, the other intervals are credible intervals

Sample size	NA	ELB	ELP	BEL1	BEL2	BpEL1	BpEL2	IF	BIF1	BIF2	BpIF1	BpIF2
$Gamma(\alpha,\beta)=(4,10), LN(\mu,\sigma)=(1,0.5), Weibull(a,b)=(4.07,7.49), P_2=0.9, P_1=P_3=0.8$												
(10,10,10)	0.689	0.977	0.990	0.923	0.978	0.976	0.970	0.764	0.836	0.805	0.745	0.736
(30,30,30)	0.911	0.820	0.810	0.974	0.990	0.995	0.994	0.838	0.831	0.827	0.932	0.933
(50,30,30)	0.910	0.793	0.829	0.952	0.982	0.986	0.984	0.821	0.805	0.804	0.927	0.928
(50,50,50)	0.923	0.833	0.853	0.946	0.968	0.950	0.949	0.914	0.892	0.897	0.936	0.939
(100,100,100)	0.940	0.946	0.922	0.967	0.974	0.965	0.960	0.965	0.964	0.962	0.960	0.959
(100,50,50)	0.928	0.818	0.849	0.956	0.982	0.952	0.948	0.903	0.925	0.924	0.933	0.932
(100,100,50)	0.900	0.917	0.917	0.961	0.972	0.957	0.956	0.955	0.884	0.884	0.948	0.948
$Gamma(\alpha,\beta)=(4,10), LN(\mu,\sigma)=(1,0.5), Weibull(a,b)=(4.25,6), P_2=0.9, P_1=P_3=0.9$												
(10,10,10)	0.629	0.987	0.989	0.937	0.990	0.985	0.981	0.583	0.765	0.605	0.670	0.668
(30,30,30)	0.906	0.798	0.816	0.959	0.983	0.990	0.988	0.805	0.745	0.751	0.910	0.910
(50,30,30)	0.900	0.812	0.805	0.958	0.984	0.991	0.988	0.801	0.740	0.751	0.914	0.914
(50,50,50)	0.904	0.822	0.807	0.958	0.979	0.957	0.952	0.904	0.852	0.855	0.925	0.924
(100,100,100)	0.926	0.924	0.929	0.958	0.961	0.948	0.942	0.944	0.932	0.930	0.940	0.939
(100,50,50)	0.910	0.822	0.819	0.960	0.977	0.949	0.949	0.884	0.842	0.849	0.929	0.929
(100,100,50)	0.895	0.905	0.918	0.973	0.965	0.947	0.947	0.946	0.834	0.839	0.937	0.939

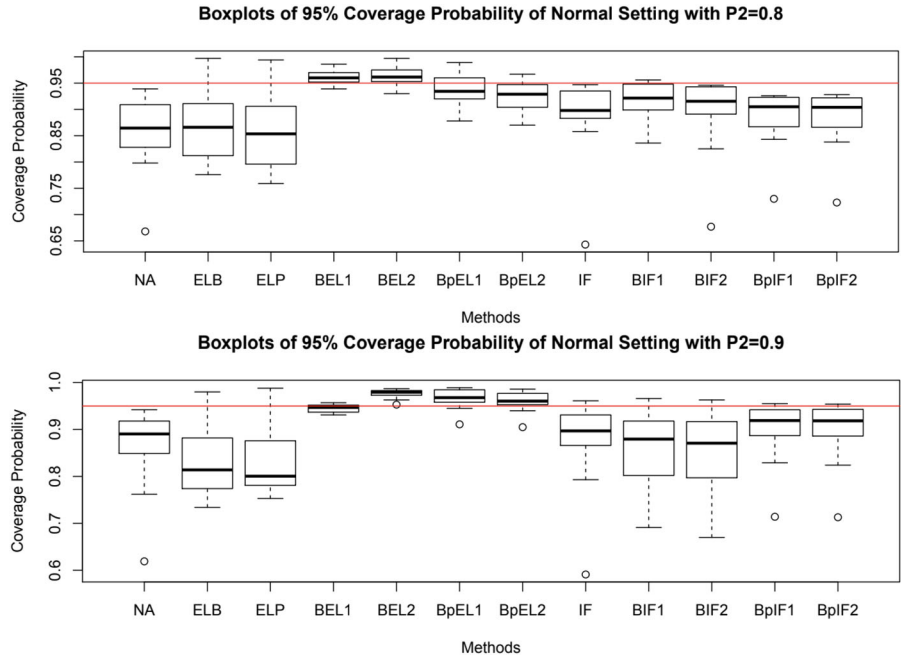

FIGURE 1 Boxplots of coverage probabilities under normal distribution setting.

overall performance. Bayesian pseudo empirical likelihood methods(BpEL1, BpEL2, BpIF1, and BpIF2) also perform well in most of the settings considered here. IF-related methods also work well when true $P_2 = 0.8$. The performance of ELB, ELP, and IF related methods all depend on the degree of separation of test outcomes in the fully diseased, early diseased, and nondiseased groups. Under higher specificity and sensitivity to the fully diseased group, they have slightly lower coverage probabilities. The performances of NA, BEL1, BEL2, BpEL1, and BpEL2 intervals do not obviously change when P_1 and P_3 increase. Comparing the results from the combined distribution setting (i) with (iii) or combined distribution setting (ii) with (iv) which have fixed P_1 and P_3 but higher true value of P_2, we can see that BEL1, BEL2, BpEL1, BpEL2, BpIF1, and BpIF2 intervals generally have similar or better finite sample performance with higher true value of P_2, and

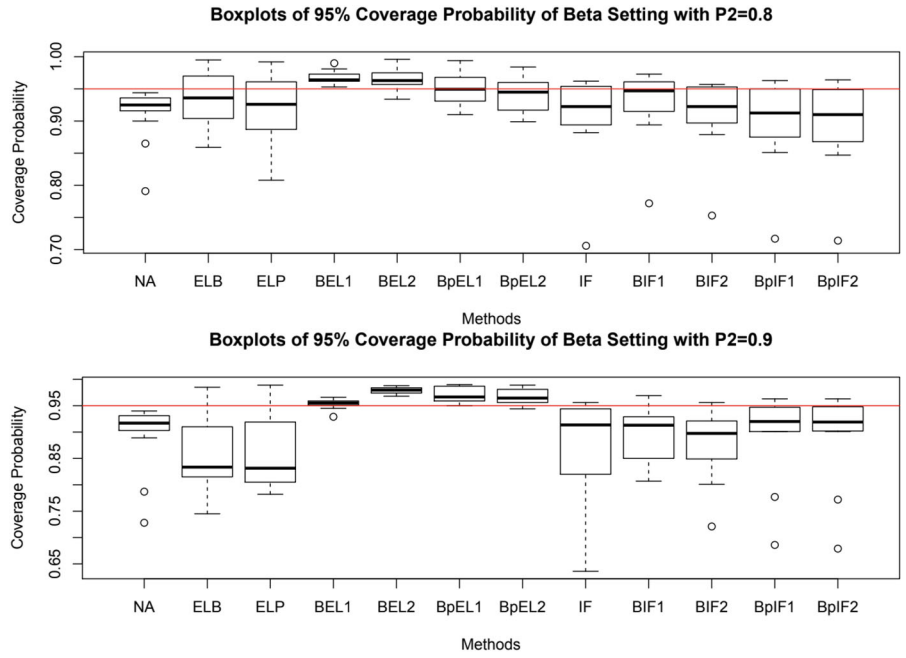

FIGURE 2 Boxplots of 95% coverage probabilities under beta distribution setting

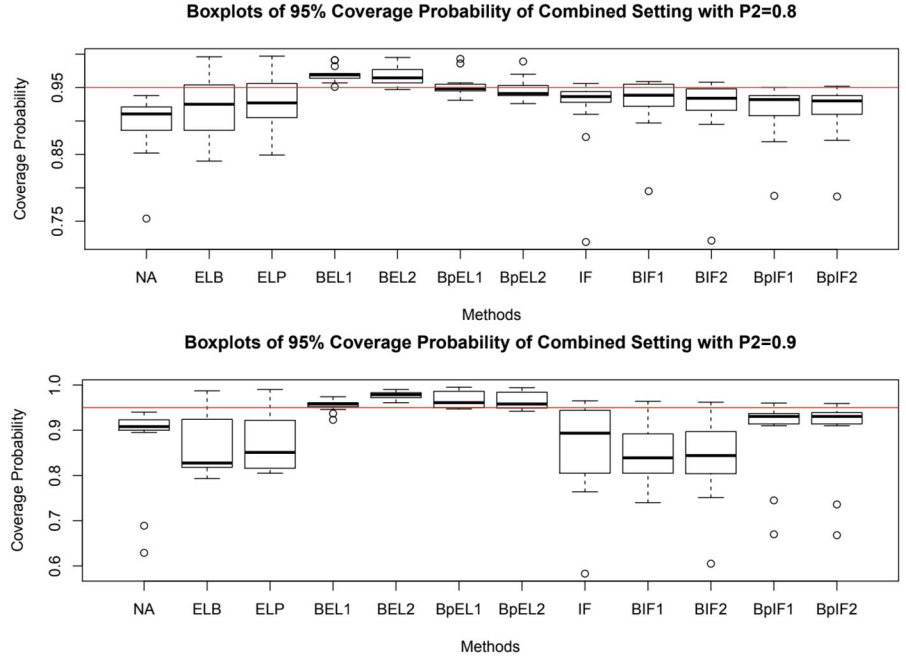

FIGURE 3 Boxplots of 95% coverage probabilities under the combined distribution setting

performances of other intervals are obviously worse when the true value of P_2 is changed from 0.8 to 0.9, especially the IF and Bayesian IF intervals. The possible reason might be it is more difficult to obtain a better density estimate when the true value of P_2 is higher (which means a higher degree of separation of early diseased test outcomes from other groups).

In summary, new Bayesian and Bayesian pseudo empirical likelihood intervals, especially BEL1 interval, are consistent and have coverage probabilities closer to the nominal confidence level than other intervals in all settings. The performance of IF and Bayesian IF methods is acceptable in some settings.

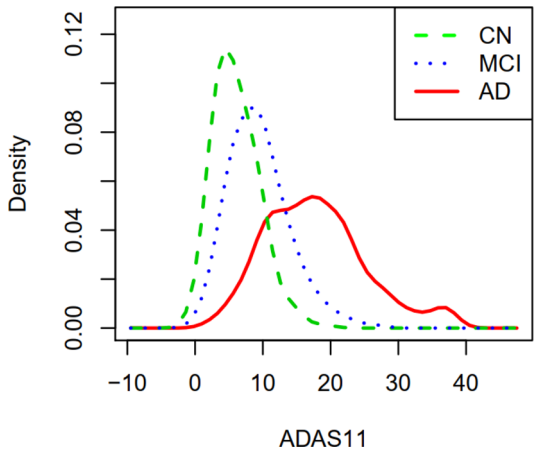

FIGURE 4 Estimated densities of ADAS11 values in the ADNI data

TABLE 8 Point estimates and 95% level credible intervals for the sensitivity P_2 of ADAS11 to MCI patients

	Point estimate	CI
	$P_1 = P_3 = 0.3$	
BEL1	0.923	(0.846 , 0.931)
BEL2	0.923	(0.851 , 0.934)
	$P_1 = P_3 = 0.4$	
BEL1	0.864	(0.759 , 0.871)
BEL2	0.864	(0.763 , 0.874)
	$P_1 = P_3 = 0.8$	
BEL1	0.280	(0.120 , 0.299)
BEL2	0.280	(0.113 , 0.290)

6 | A REAL EXAMPLE IN THE DETECTION OF ALZHEIMER'S DISEASE

In this section, we illustrate the application of the proposed methods to assess the diagnostic accuracy of biomarkers in the detection of Alzheimer's disease (AD). The data used in this section were obtained from the ADNI database (adni.loni.usc.edu). The goal of the ADNI study is to track the progression of the diseases, MCI, and AD, using biomarkers and clinical measures.

We apply the proposed methods to a small subset of a data-freeze named "QT-PAD Project Data," which was downloaded on June 29, 2017. It is available in the "Test Data/Data for Challenges" section of the LONI website (ADNI database). Here, we only consider nonmissing records based on the commonly used biomarker: Alzheimer's Disease Assessment Scale 11 (ADAS11). The dataset we used consists of 203 control subjects (CN), 389 MCI, and 237 AD patients. We consider MCI as the early stage of AD. Figure 4 presents the estimated density curves of ADAS11 values from these three groups, respectively. We note that the results of MCI patients are similar to those of control groups. Therefore, we can expect that the sensitivities of this biomarker to MCI patients cannot be high enough.

In Table 8, we report point estimates and 95% level BEL1 and BEL2 intervals for the sensitivity P_2 of this biomarker to MCI patients when $P_1 = P_3 = 0.3$, $P_1 = P_3 = 0.4$, and $P_1 = P_3 = 0.8$. As expected, the estimated sensitivity of ADS11 to MCI (the early stage of AD) patients is low when $P_1 = P_3 = 0.8$. We observe that ADAS11 has relatively high sensitivities (0.864–0.923) to the MCI patients when P_1 and P_3 are low (0.4–0.3).

7 | DISCUSSION

In this article, we propose an influence function-based EL method, several Bayesian and Bayesian pseudo EL methods for inference on sensitivity of a test to early stage disease. Our simulation results show that the proposed intervals have

better coverage accuracy than the existing intervals. The proposed BEL and BpEL intervals have the best performance among all intervals discussed in this paper. The influence function-based intervals perform slightly worse than the BEL and BpEL intervals.

In practice, clinicians sometimes need to compare the sensitivities of two tests to early stage disease at the same specificity and sensitivity to the late stage of disease, denoted as P_{21} and P_{22}. Similar to the proposed Bayesian approach, we can generate posterior samples of P_{21} and P_{22} separately to obtain posterior samples of $(P_{21} - P_{22})$. Based on these posterior samples, Bayesian credible intervals for the difference of the sensitivities of two tests to early stage disease can be easily constructed. In addition, the influence function techniques can be extended immediately to make inference on the difference between the sensitivities of two tests since the influence function of the difference is the difference between the influence functions of two sensitivities to early stage disease.

ACKNOWLEDGMENTS

We thank the associate editor and two reviewers for their helpful comments. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

CONFLICT OF INTEREST

The authors have declared no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in the Alzheimer's Disease Neuroimaging Initiative (ADNI) database at adni.loni.usc.edu.

OPEN RESEARCH BADGES

This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available in the Supporting Information section.

This article has earned an open data badge "**Reproducible Research**" for making publicly available the code necessary to reproduce the reported results. The results reported in this article were reproduced partially due to computational complexity.

ORCID

Gengsheng Qin https://orcid.org/0000-0002-8058-3124

REFERENCES

Aarsland, D., & Kurz, M. W. (2010). The epidemiology of dementia associated with Parkinson's disease. *Brain Pathology*, *20*, 633–639.

Adimari, G., & Guolo, A. (2010). A note on the asymptotic behaviour of empirical likelihood statistics. *Statistical Methods & Applications*, *19*, 463–476.

Berger, J. O., & Bernardo, J. M. (1992). On the development of reference priors. *Bayesian Statistics*, *4*, 35–60.

Berger, J. O., Bernardo, J. M., & Sun, D. (2009). The formal definition of reference priors. *The Annals of Statistics*, *37*, 905–938.

Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference. *Journal of the Royal Statistical Society. Series B (Methodological)*, *41*, 113–147.
Chen, J., Variyath, A. M., & Abraham, B. (2008). Adjusted empirical likelihood and its properties. *Journal of Computational and Graphical Statistics*, *17*, 426–443.
Clarke, B., & Yuan, A. (2010). Reference priors for empirical likelihoods. In M. H. Chen, P. Mueller, D. Sun, K. Ye, & D. K. Dey (Eds.). *Frontiers of statistical decision making and Bayesian analysis: In honor of James O. Berger*, pp. 56–68). Springer.
Dong, T., Tian, L., Hutson, A., & Xiong, C. (2011). Parametric and non-parametric confidence intervals of the probability of identifying early disease stage given sensitivity to full disease and specificity with three ordinal diagnostic groups. *Statistics in Medicine*, *30*, 3532–3545.
Dong, T., & Tian, L. (2015). Confidence interval estimation for sensitivity to the early diseased stage based on empirical likelihood. *Journal Biopharm Statics*, *25*, 1215–1233.
Ghosh, J. K. (1971). A new proof of the Bahadur representation of quantiles and an application. *The Annals of Mathematical Statistics*, *42*, 1957–1961.
Gong, Y., Peng, L., & Qi, Y. (2010). Smoothed jackknife empirical likelihood method for ROC curve. *Journal of Multivariate Analysis*, *6*, 1520–1531.
Hai, Y., Min, X., & Qin, G. (2020). Bayesian and influence function-based empirical likelihoods for inference of sensitivity in diagnostic tests. *Statistical Methods in Medical Research*, *29*, 3457–3491.
Hall, P., & La Scala, B. (1990). Methodology and algorithm of empirical likelihood. *International Statistical Review*, *58*, 109–127.
Heckerling, P. S. (2001). Parametric three-way receiver operating characteristic surface analysis using Mathematica. *Medical Decision Making*, *21*, 409–417.
Lazar, N. A. (2003). Bayesian empirical likelihood. *Biometrika*, *90*, 319–326.
Li, J., & Zhou, X. H. (2009). Nonparametric and semiparametric estimation of the three way receiver operating characteristic surface. *Journal of Statistical Planning and Inference*, *139*, 4133–4142.
Mossman, D. (1999). Three-way ROCs. *Medical Decision Making*, 1999; *19*, 78–79.
Nakas, C. T., & Yiannoutsos, C. T. (2004). Ordered multiple-class ROC analysis with continuous measurements. *Statistics in Medicine*, *23*, 3437–3449.
Owen, A. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics*, *18*, 90–120.
Owen, A. (2001). *Empirical likelihood*. Chapman & Hall/CRC.
Petersen, R. C. (2004). Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine*, *256*, 183–194.
Rao, J. N. K., & Wu, C. (2010). Bayesian pseudo-empirical-likelihood intervals for complex surveys. *Journal of the Royal Statistical Society. Series B (Methodological)*, *72*, 533–544.
Scurfield, B. K. (1996). Multiple-event forced-choice tasks in the theory of signal detectability. *Journal of Mathematical Psychology*, *40*, 253–269.
Silverman, B. W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *The Annals of Statistics*, *6*(1), 177–184.
Wand, M. P., & Jones, M. C. (1995). *Kernel smoothing*. Chapman & Hall/CRC.
Xiong, C., Belle, G. van, Philip, M., & John, C. M. (2006). Measuring and estimating diagnostic accuracy when there are three ordinal diagnostic groups. *Statistics in Medicine*, *25*, 1251–1273.
Yu, W., Zhao, Z., & Zheng, M. (2012). Empirical likelihood methods based on influence functions. *Statistics and Its Interface*, *5*, 355–366.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Hai, Y., Shi, S., & Qin for the Alzheimer's Disease Neuroimaging, G. (2023). Bayesian and influence function-based empirical likelihoods for inference of sensitivity to the early diseased stage in diagnostic tests. *Biometrical Journal*, *65,* 2200021. https://doi.org/10.1002/bimj.202200021

APPENDIX A

A.1 | Proof of propositions and theorem

Proof of Proposition 1. From (23), we only need to prove that

$$\frac{1}{\sigma\sqrt{N}}\sum_{l=1}^{N} W_l(P_1, P_2, P_3) \xrightarrow{d} N(0,1). \tag{A.1}$$

From (24), we have that

$$\frac{1}{\sqrt{N}}\sum_{l=1}^{N} W_l(P_1,P_2,P_3) = \sqrt{N}\Big\{\frac{1}{n_1}\frac{f_2(c_1)}{f_1(c_1)}\sum_{j=1}^{n_1}[I(Y_{1,j}\le c_1)-P_1] + \frac{1}{n_2}\sum_{j=1}^{n_2}[I(c_1<Y_{2,j}\le c_2)-P_2] + \frac{1}{n_3}\frac{f_2(c_2)}{f_3(c_2)}\sum_{j=1}^{n_3}[I(Y_{3,j}>c_2)-P_3]\Big\}. \tag{A.2}$$

Since $I(Y_{1,j}\le c_1)\text{s} \overset{iid}{\sim} Binomial(1,P_1)$, $I(c_1<Y_{2,j}\le c_2)\text{s} \overset{iid}{\sim} Binomial(1,P_2)$, and $I(Y_{3,j}>c_2)\text{s} \overset{iid}{\sim} Binomial(1,P_3)$, by the central limit theorem, we have that

$$\begin{aligned}
&\frac{1}{\sqrt{n_1}}\sum_{j=1}^{n_1}[I(Y_{1,j}\le c_1)-P_1] \xrightarrow{d} N(0,P_1(1-P_1)),\\
&\frac{1}{\sqrt{n_2}}\sum_{j=1}^{n_2}[I(c_1<Y_{2,j}\le c_2)-P_2] \xrightarrow{d} N(0,P_2(1-P_2)),\\
&\frac{1}{\sqrt{n_3}}\sum_{i=1}^{n_3}[I(Y_{3,j}>c_2)-P_3] \xrightarrow{d} N(0,P_3(1-P_3)).
\end{aligned} \tag{A.3}$$

Hence, (A.1) and Proposition 1 follow immediately from (23) and the independence of $Y_{1,i}$s, $Y_{2,j}$s, and $Y_{3,k}$s.

We need the following lemma for the proof of Theorem 1. □

Lemma A.1. *Under the conditions in Theorem 1, we have that*

(i) $\frac{1}{\sigma\sqrt{N}}\sum_{l=1}^{N}\hat{W}_l(P_1,P_2,P_3) \xrightarrow{d} N(0,1)$.

(ii) $\frac{1}{N}\sum_{l=1}^{N}\hat{W}_l^2(P_1,P_2,P_3) \xrightarrow{p} \sigma^2$.

Proof. (i) From (A.1), we only need to prove that

$$\frac{1}{\sqrt{N}}\sum_{l=1}^{N}\hat{W}_l(P_1,P_2,P_3) = \frac{1}{\sqrt{N}}\sum_{l=1}^{N}W_l(P_1,P_2,P_3)+o_p(1). \tag{A.4}$$

We have the following decomposition:

$$\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_{l=1}^{N}\hat{W}_l(_1,P_2,P_3)\\
&= \frac{1}{\sqrt{N}}\sum_{l=1}^{N}W_l(P_1,P_2,P_3) + \sqrt{N}\Big\{\frac{1}{n_2}\sum_{j=1}^{n_2}[I(\hat{c}_1<Y_{2,j}\le\hat{c}_2)-I(c_1<Y_{2,j}\le c_2)]\Big\}\\
&+ \sqrt{N}\Big\{\frac{1}{n_1}\frac{\hat{f}_2(\hat{c}_1)}{\hat{f}_1(\hat{c}_1)}\sum_{j=1}^{n_1}[I(Y_{1,j}\le\hat{c}_1)-P_1] - \frac{1}{n_1}\frac{f_2(c_1)}{f_1(c_1)}\sum_{j=1}^{n_1}[I(Y_{1,j}\le c_1)-P_1]\Big\}\\
&+ \sqrt{N}\Big\{\frac{1}{n_3}\frac{\hat{f}_2(\hat{c}_2)}{\hat{f}_3(\hat{c}_2)}\sum_{j=1}^{n_3}[I(Y_{3,j}>\hat{c}_2)-P_3] - \frac{1}{n_3}\frac{f_2(c_2)}{f_3(c_2)}\sum_{j=1}^{n_3}[I(Y_{3,j}>c_2)-P_3]\Big\}.
\end{aligned} \tag{A.5}$$

As we mentioned in Section 3, using the Bahadur representation of the sample quantiles $\hat{c}_1$ and $\hat{c}_2$ (Ghosh, 1971),

$$\begin{aligned}\hat{c}_1 - c_1 &= \frac{P_1 - \frac{1}{n_1}\sum_{j=1}^{n_1} I(Y_{1,j} \le c_1)}{f_1(c_1)} + o_p(n_1^{-\frac{1}{2}}),\\ \hat{c}_2 - c_2 &= \frac{\frac{1}{n_3}\sum_{j=1}^{n_3} I(Y_{3,j} > c_2) - P_3}{f_3(c_2)} + o_p(n_3^{-\frac{1}{2}}),\end{aligned} \tag{A.6}$$

and Equations (21) and (22), we get that

$$\begin{aligned}&\frac{1}{n_2}\sum_{j=1}^{n_2}[I(\hat{c}_1 < Y_{2,j} \le \hat{c}_2) - I(c_1 < Y_{2,j} \le c_2)]\\ &= \frac{1}{n_2}\sum_{j=1}^{n_2}[I(Y_{2,j} \le \hat{c}_2) - I(Y_{2,j} \le \hat{c}_1)] - \frac{1}{n_2}\sum_{j=1}^{n_2}[I(Y_{2,j} \le c_2) - I(Y_{2,j} \le c_1)]\\ &= \frac{1}{n_2}\sum_{j=1}^{n_2}[I(Y_{2,j} \le \hat{c}_2) - I(Y_{2,j} \le c_2)] - \frac{1}{n_2}\sum_{j=1}^{n_2}[I(Y_{2,j} \le \hat{c}_1) - I(Y_{2,j} \le c_1)]\\ &= [\hat{F}_2(\hat{c}_2) - \hat{F}_2(c_2)] - [\hat{F}_2(\hat{c}_1) - \hat{F}_2(c_1)]\\ &= \frac{1}{n_3}\frac{f_2(c_2)}{f_3(c_2)}\sum_{j=1}^{n_3}[I(Y_{3,j} > c_2) - P_3] + \frac{1}{n_1}\frac{f_2(c_1)}{f_1(c_1)}\sum_{j=1}^{n_1}[I(Y_{1,j} \le c_1) - P_1] + o_p(N^{-1/2}).\end{aligned} \tag{A.7}$$

In addition, we have that

$$\begin{aligned}&\frac{1}{n_1}\sum_{j=1}^{n_1}[I(Y_{1,j} \le \hat{c}_1) - P_1]\\ &= \frac{1}{n_1}\sum_{j=1}^{n_1}[I(Y_{1,j} \le \hat{c}_1) - I(Y_{1,j} \le c_1)] + \frac{1}{n_1}\sum_{j=1}^{n_1}[I(Y_{1,j} \le c_1) - P_1]\\ &= \int [I(y \le \hat{c}_1) - I(y \le c_1)]d\hat{F}_1(y) + \frac{1}{n_1}\sum_{j=1}^{n_1}[I(Y_{1,j} \le c_1) - P_1]\\ &= \int [I(y \le \hat{c}_1) - I(y \le c_1)]dF_1(y)\\ &\quad + n_1^{-1/2}\int [I(y \le \hat{c}_1) - I(y \le c_1)]d(n_1^{1/2}(\hat{F}_1(y) - F_1(y))) + \frac{1}{n_1}\sum_{j=1}^{n_1}[I(Y_{1,j} \le c_1) - P_1].\\ &\equiv D_1 + D_2 + D_3.\end{aligned} \tag{A.8}$$

From $n_1^{1/2}(\hat{F}_1(y) - F_1(y)) \xrightarrow{d} B(y)$ which is a Gaussian process, and $I(y \le \hat{c}_1) - I(y \le c_1) \to 0$ a.s., it follows that

$$D_2 = n_1^{-1/2}\int [I(y \le \hat{c}_1) - I(y \le c_1)]d(n_1^{1/2}(\hat{F}_1(y) - F_1(y))) = o_p(n_1^{-1/2}). \tag{A.9}$$

Using the Bahadur representation for the sample quantile $\hat{c}_1$, $\hat{c}_1 - c_1 = O_p(n_1^{-1/2})$, and $o(\hat{c}_1 - c_1) = o_p(n_1^{-1/2})$, we get that

$$\begin{aligned}
&\frac{1}{n_1}\sum_{j=1}^{n_1}[I(Y_{1,j} \le \hat{c}_1) - P_1] = D_1 + o_p(n_1^{-1/2}) + D_3 \\
&= \int [I(y \le \hat{c}_1) - I(y \le c_1)]dF_1(y) + o_p(n_1^{-1/2}) + \frac{1}{n_1}\sum_{j=1}^{n_1}[I(Y_{1,j} \le c_1) - P_1] \\
&= f_1(c_1)(\hat{c}_1 - c_1) + o(\hat{c}_1 - c_1) + \frac{1}{n_1}\sum_{j=1}^{n_1}[I(Y_{1,j} \le c_1) - P_1] + o_p(n_1^{-1/2}) \\
&= f_1(c_1)\left(\frac{P_1 - \frac{1}{n_1}\sum_{j=1}^{n_1} I(Y_{1,j} \le c_1)}{f_1(c_1)} + o_p(n_1^{-\frac{1}{2}})\right) + \frac{1}{n_1}\sum_{j=1}^{n_1}[I(Y_{1,j} \le c_1) - P_1] + o_p(n_1^{-1/2}), \\
&= f_1(c_1) \cdot o_p(n_1^{-1/2}) + o_p(n_1^{-1/2}) = o_p(n_1^{-1/2}).
\end{aligned} \tag{A.10}$$

Similarly, we have that

$$\begin{aligned}
&\frac{1}{n_3}\sum_{j=1}^{n_3}[I(Y_{3,j} > \hat{c}_2) - P_3] \\
&= \frac{1}{n_3}\sum_{j=1}^{n_3}[I(Y_{3,j} > \hat{c}_2) - I(Y_{3,j} > c_2)] + \frac{1}{n_3}\sum_{j=1}^{n_3}[I(Y_{3,j} > c_2) - P_3] \\
&= -\int [I(y \le \hat{c}_2) - I(y \le c_2)]d\hat{F}_3(y) + \frac{1}{n_3}\sum_{j=1}^{n_3}[I(Y_{3,j} > c_2) - P_3] \\
&= -\int [I(y \le \hat{c}_2) - I(y \le c_2)]dF_3(y) + o_p(n_3^{-1/2}) + \frac{1}{n_3}\sum_{j=1}^{n_3}[I(Y_{3,j} > c_2) - P_3] \\
&= -f_3(c_2)(\hat{c}_2 - c_2) + \frac{1}{n_3}\sum_{j=1}^{n_3}[I(Y_{3,j} > c_2) - P_3] + o_p(n_3^{-1/2}) \\
&= -f_3(c_2)\left(\frac{\frac{1}{n_3}\sum_{j=1}^{n_3} I(Y_{3,j} > c_2) - P_3}{f_3(c_2)} + o_p(n_3^{-\frac{1}{2}})\right) + \frac{1}{n_3}\sum_{j=1}^{n_3}[I(Y_{3,j} > c_2) - P_3] + o_p(n_3^{-1/2}) \\
&= -f_3(c_2) \cdot o_p(n_3^{-1/2}) + o_p(n_3^{-1/2}) = o_p(n_3^{-1/2}),
\end{aligned} \tag{A.11}$$

Therefore,

$$\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_{l=1}^{N}\hat{W}_l(_1, P_2, P_3) = \frac{1}{\sqrt{N}}\sum_{l=1}^{N} W_l(P_1, P_2, P_3) \\
&\quad + \frac{\sqrt{N}}{n_1}\frac{\hat{f}_2(\hat{c}_1)}{\hat{f}_1(\hat{c}_1)}\sum_{j=1}^{n_1}[I(Y_{1,j} \le \hat{c}_1) - P_1] + \frac{\sqrt{N}}{n_3}\frac{\hat{f}_2(\hat{c}_2)}{\hat{f}_3(\hat{c}_2)}\sum_{j=1}^{n_3}[I(Y_{3,j} > \hat{c}_2) - P_3] + o_p(1) \\
&\quad = \frac{1}{\sqrt{N}}\sum_{l=1}^{N} W_l(P_1, P_2, P_3) + o_p(1).
\end{aligned} \tag{A.12}$$

Under the assumptions in Theorem 1, kernel density estimators for $\hat{f}_i$s are almost surely and uniformly consistent (see Silverman, 1978). The last equality holds by the uniform consistency of the density estimates $\hat{f}_1$, $\hat{f}_2$, and $\hat{f}_3$, and $\frac{f_2(c_1)}{f_1(c_1)} = O(1)$, $\frac{f_2(c_2)}{f_3(c_2)} = O(1)$. Lemma A.1(i) is thus proved.

(ii) Since

$$
\begin{aligned}
&\frac{1}{N}\sum_{l=1}^{N} W_l^2(P_1,P_2,P_3)\\
&= \frac{N}{n_1^2}\frac{f_2^2(c_1)}{f_1^2(c_1)}\sum_{j=1}^{n_1}[I(Y_{1,j}\le c_1)-P_1]^2 + \frac{N}{n_2^2}\sum_{j=1}^{n_2}[I(c_1<Y_{2,j}\le c_2)-P_2]^2\\
&\quad + \frac{N}{n_3^2}\frac{f_2^2(c_2)}{f_3^2(c_2)}\sum_{j=1}^{n_3}[I(Y_{3,j}>c_2)-P_3]^2\\
&= (1+\rho_1^{-1}+\rho_3^{-1})\frac{f_2^2(c_1)}{f_1^2(c_1)}E[I(Y_{1,j}\le c_1)-P_1]^2 + (1+\rho_1+\rho_2)E[I(c_1<Y_{2,j}\le c_2)-P_2]^2\\
&\quad + (1+\rho_2^{-1}+\rho_3)\frac{f_2^2(c_2)}{f_3^2(c_2)}E[I(Y_{3,j}>c_2)-P_3]^2 + o_p(1)\\
&= (1+\rho_1^{-1}+\rho_3^{-1})P_1(1-P_1)\frac{f_2^2(c_1)}{f_1^2(c_1)} + (1+\rho_1+\rho_2)P_2(1-P_2)\\
&\quad + (1+\rho_2^{-1}+\rho_3)\frac{f_2^2(c_2)}{f_3^2(c_2)}P_3(1-P_3) + o_p(1) = \sigma^2 + o_p(1),
\end{aligned}
\tag{A.13}
$$

we only need to prove that

$$
\frac{1}{N}\sum_{l=1}^{N}\hat{W}_l^2(P_1,P_2,P_3) = \frac{1}{N}\sum_{l=1}^{N} W_l^2(P_1,P_2,P_3) + o_p(1). \tag{A.14}
$$

Under the assumptions in Theorem 1, using the uniform consistency of the density estimate $\hat{f}_1$ (Silverman, 1978) and the strong consistency of the sample quantile $\hat{c}_1$, we get that

$$
\begin{aligned}
\left|\hat{f}_1(\hat{c}_1)-f_1(c_1)\right| &\le \left|\hat{f}_1(\hat{c}_1)-f_1(\hat{c}_1)\right| + \left|f_1(\hat{c}_1)-f_1(c_1)\right|\\
&\le \sup_x\left|\hat{f}_1(x)-f_1(x)\right| + o_p(1) = o_p(1).
\end{aligned}
\tag{A.15}
$$

So, $\hat{f}_1(\hat{c}_1) = f_1(c_1) + o_p(1)$. Similarly, we have $\hat{f}_3(\hat{c}_2) = f_3(c_2) + o_p(1)$, $\hat{f}_2(\hat{c}_1) = f_2(c_1) + o_p(1)$, and $\hat{f}_2(\hat{c}_2) = f_2(c_2) + o_p(1)$. By Slutsky's theorem, we have that $\frac{\hat{f}_2^2(c_1)}{\hat{f}_1^2(c_1)} = \frac{f_2^2(c_1)}{f_1^2(c_1)} + o_p(1)$ and $\frac{\hat{f}_2^2(c_2)}{\hat{f}_3^2(c_2)} = \frac{f_2^2(c_2)}{f_3^2(c_2)} + o_p(1)$.

From Equation (A.7) and central limit theorem (CLT), it follows that

$$
\begin{aligned}
&\frac{1}{n_2}\sum_{j=1}^{n_2}\left[I(c_1<Y_{2,j}\le c_2) - I(\hat{c}_1<Y_{2,j}\le\hat{c}_2)\right]\\
&= -\frac{1}{n_3}\frac{f_2(c_2)}{f_3(c_2)}\sum_{j=1}^{n_3}[I(Y_{3,j}>c_2)-P_3] - \frac{1}{n_1}\frac{f_2(c_1)}{f_1(c_1)}\sum_{j=1}^{n_1}[I(Y_{1,j}\le c_1)-P_1] + o_p(N^{-1/2})\\
&= O_p(n_1^{-1/2}) + O_p(n_3^{-1/2}) + o_p(N^{-1/2})\\
&= O_p(N^{-1/2}).
\end{aligned}
\tag{A.16}
$$

Similarly, from Equation (21) and Bahadur representation of the sample quantile, we have that

$$\begin{aligned}\frac{1}{n_1}\sum_{j=1}^{n_1}[I(Y_{1,j}\le c_1)-I(Y_{1,j}\le \hat{c}_1)] &= f_1(c_1)(c_1-\hat{c}_1)+o_p(n_1^{-1/2})\\ &= \frac{1}{n_1}\sum_{j=1}^{n_1}[I(Y_{1,j}\le c_1)-P_1]+o_p(n_1^{-1/2})\\ &= O_p(n_1^{-1/2})\end{aligned} \tag{A.17}$$

and

$$\frac{1}{n_3}\sum_{j=1}^{n_3}[I(Y_{3,j}> c_2)-I(Y_{3,j}> \hat{c}_2)] = O_p(n_3^{-1/2}). \tag{A.18}$$

Therefore,

$$\begin{aligned}&\Big|\frac{1}{N}\sum_{l=1}^{N}W_l^2(P_1,P_2,P_3)-\frac{1}{N}\sum_{l=1}^{N}\hat{W}_l^2(P_1,P_2,P_3)\Big|\\ &= (N)\Big|\frac{1-2P_2}{n_2^2}\sum_{j=1}^{n_2}\left[I(c_1<Y_{2,j}\le c_2)-I(\hat{c}_1<Y_{2,j}\le \hat{c}_2)\right]+\frac{P_1^2}{n^2}\left[\frac{f_2^2(c_1)}{f_1^2(c_1)}-\frac{\hat{f}_2^2(c_1)}{\hat{f}_1^2(c_1)}\right]\\ &+\frac{1-2P_1}{n_1^2}\sum_{j=1}^{n_1}\left\{\left(\frac{f_2^2(c_1)}{f_1^2(c_1)}-\frac{\hat{f}_2^2(c_1)}{\hat{f}_1^2(c_1)}\right)I(Y_{1,j}\le c_1)+\frac{\hat{f}_2^2(c_1)}{\hat{f}_1^2(c_1)}\left[I(Y_{1,j}\le c_1)-I(Y_{1,j}\le \hat{c}_1)\right]\right\}\\ &+\frac{1-2P_3}{n_3^2}\sum_{j=1}^{n_3}\left\{\left(\frac{f_2^2(c_2)}{f_3^2(c_2)}-\frac{\hat{f}_2^2(c_2)}{\hat{f}_3^2(c_2)}\right)I(Y_{3,j}> c_2)+\frac{\hat{f}_2^2(c_2)}{\hat{f}_3^2(c_2)}\left[I(Y_{3,j}> c_2)-I(Y_{3,j}> \hat{c}_2)\right]\right\}\\ &+\frac{P_3^2}{n^3}\left[\frac{f_2^2(c_2)}{f_3^2(c_2)}-\frac{\hat{f}_2^2(c_2)}{\hat{f}_3^2(c_2)}\right]\Big|\\ &\le o_p(1),\end{aligned} \tag{A.19}$$

and Lemma A.1(ii) is proved. □

Proof of Theorem 1. From (24), we have that $E(W_l(P_1,P_2,P_3))=0$, and

$$Var(W_l(P_1,P_2,P_3)) = \begin{cases}\dfrac{N^2}{n_1^2}\dfrac{f_2^2(c_1)}{f_1^2(c_1)}P_1(1-P_1), & l=1,\dots,n_1,\\ \dfrac{N^2}{n_2^2}P_2(1-P_2), & l=n_1+1,\dots,n_1+n_2,\\ \dfrac{N^2}{n_3^2}\dfrac{f_2^2(c_2)}{f_3^2(c_2)}P_3(1-P_3), & l=n_1+n_2+1,\dots,N.\end{cases} \tag{A.20}$$

From (A.20) and the assumptions in Theorem 1, it follows that $Var(W_l(P_1,P_2,P_3))<\infty$. Then, using the similar techniques in the proofs of Lemma 11.1 and Theorem 3.2 in Owen (2001), we get that zero is inside the convex hull of $W_l(P_1,P_2^0,P_3)$s with probability tending to 1. From the strong consistency of the density estimates $\hat{f}_i$s and the strong consistency of the sample quantile $\hat{c}_i$s, it follows that $\hat{W}_l(P_1,P_2^0,P_3)=W_l(P_1,P_2^0,P_3)+o(1),\ a.s.$ Hence, zero is inside the convex hull of $\hat{W}_l(P_1,P_2^0,P_3)$s with probability tending to 1.

Using Proposition 1 and Lemma A.1, Theorem 1 can be proved by using the similar techniques in the proof of Theorem 2 in Gong et al. (2010) or an application of Theorem 1 and subsequent corollaries in Adimari and Guolo (2010). The details for the proof are omitted here. □

Proof of Proposition 2. We first briefly introduce the approach of Clarke and Yuan (2010). Define the outer product matrix $\Omega = E[g(Z_j, P_2)g'(Z_j, P_2)]$, Jacobian matrix $D(P_2) = E[\partial g(Z_j, P_2)/\partial P_2]$ and the matrix $\Lambda(P_2) = D'(P_2)\Omega^{-1}(P_2)D(P_2)$, where $g(Z_j, P_2)$ is an estimating function.

For our BEL approach, $g(Z_j, P_2) = U(Y_{2,j}) - P_2$, where $U(Y)$ is defined in Section 2, and $\Omega(P_2) = E[g(Z_j, P_2)]^2 = E[U(Y_{2,j}) - P_2]^2 = P_2(1 - P_2)$. Thus we have

$$\Lambda(P_2) = D'(P_2)\Omega^{-1}(P_2)D(P_2) = \frac{1}{P_2(1 - P_2)}. \tag{A.21}$$

So the reference prior for the EL under the relative entropy is $\pi_{EL,1}(P_2) \propto |\Lambda^{-1}(P_2)|^{1/2} = \sqrt{P_2(1 - P_2)}$, that is, $\pi_{EL,1}(P_2) = \beta(\frac{3}{2}, \frac{3}{2})$, and the reference prior for the hybrid EL under Hellinger distance is

$$\pi_{EL,2}(P_2) \propto |\Lambda(P_2)|^{1/2} = \frac{1}{\sqrt{P_2(1 - P_2)}}, \tag{A.22}$$

that is, $\pi_{EL,2}(P_2) = \beta(\frac{1}{2}, \frac{1}{2})$. □

A.2 | QQ plots for IF-, ELB-, and ELP-based empirical log-likelihood ratio statistics with normal distributions as the underlying distributions

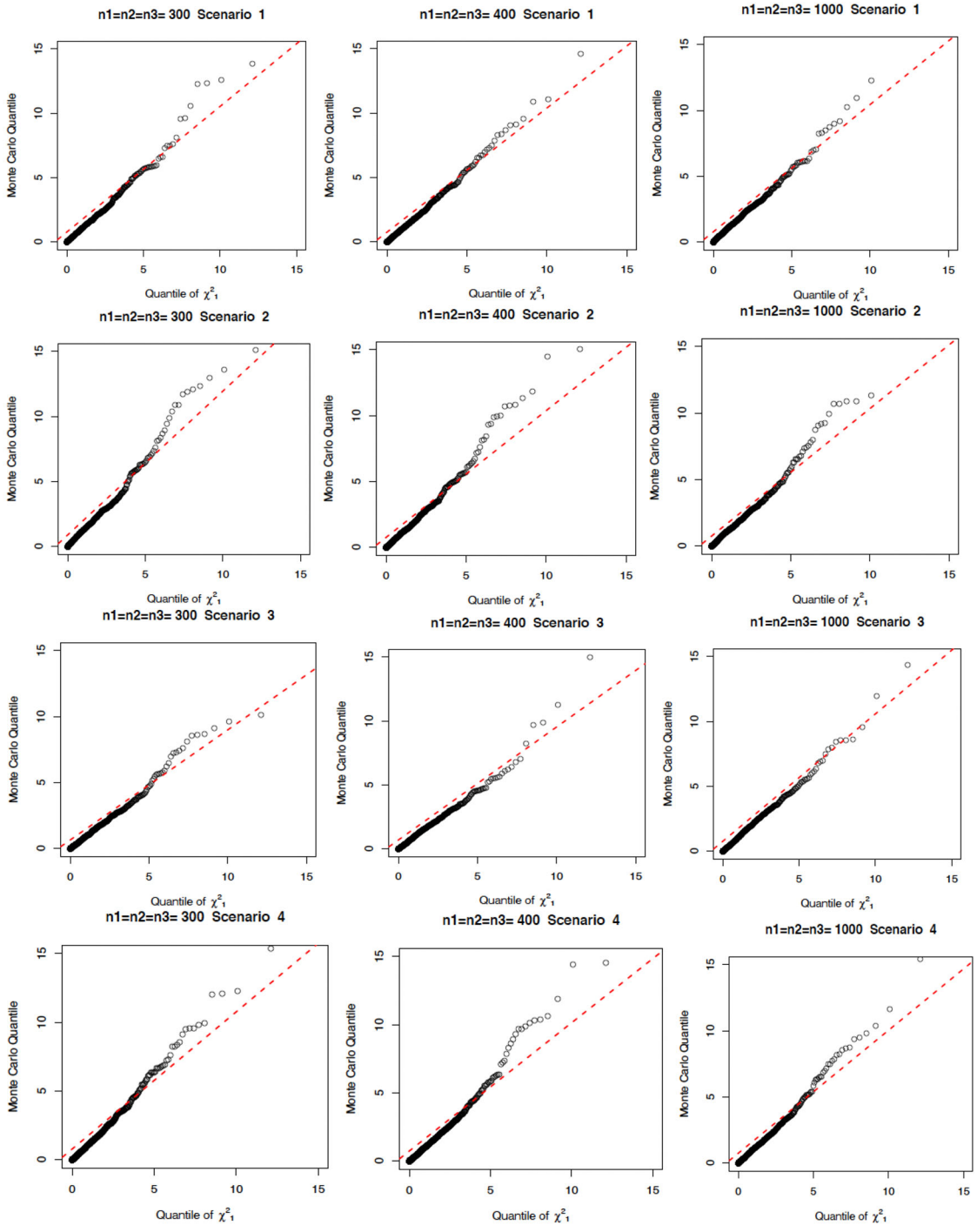

FIGURE A1 Four scenarios for (P_1, P_2, P_3): (1) $P_1 = P_3 = 0.8$ and $P_2 = 0.8$; (2) $P_1 = P_3 = 0.9$ and $P_2 = 0.8$; (3) $P_1 = P_3 = 0.8$ and $P_2 = 0.9$; (4) $P_1 = P_3 = 0.9$ and $P_2 = 0.9$

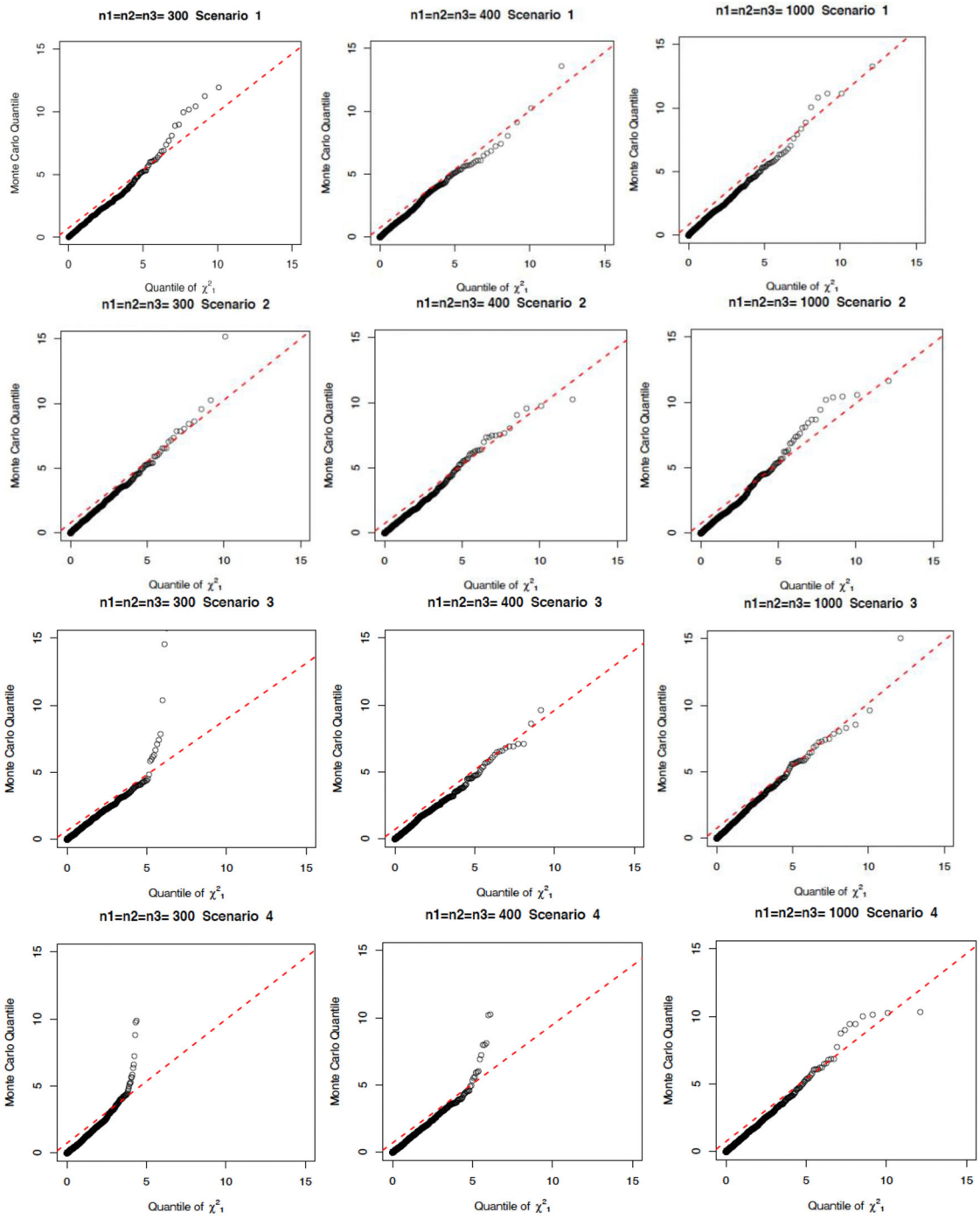

FIGURE A2 Four scenarios for (P_1, P_2, P_3): (1) $P_1 = P_3 = 0.8$ and $P_2 = 0.8$; (2) $P_1 = P_3 = 0.9$ and $P_2 = 0.8$; (3) $P_1 = P_3 = 0.8$ and $P_2 = 0.9$; (4) $P_1 = P_3 = 0.9$ and $P_2 = 0.9$

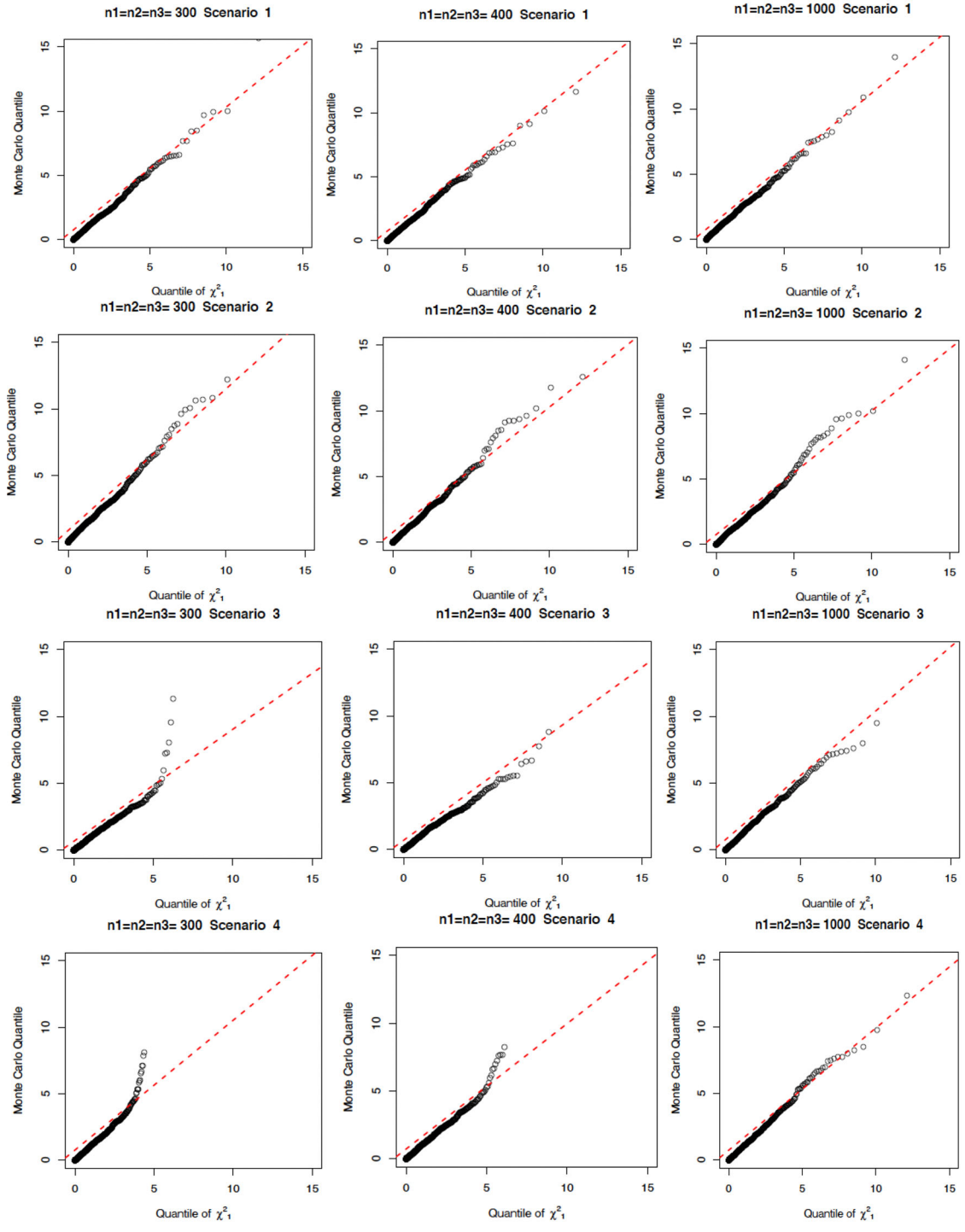

FIGURE A3 Four scenarios for (P_1, P_2, P_3): (1) $P_1 = P_3 = 0.8$ and $P_2 = 0.8$; (2) $P_1 = P_3 = 0.9$ and $P_2 = 0.8$; (3) $P_1 = P_3 = 0.8$ and $P_2 = 0.9$; (4) $P_1 = P_3 = 0.9$ and $P_2 = 0.9$