Minimum Clinically Important Difference in Medical Studies

A. S. Hedayat[†], Junhui Wang[†‡] and Tu Xu[§]

[†] Department of Mathematics, Statistics,

and Computer Science

University of Illinois at Chicago

Chicago, IL, 60607

[‡] Department of Mathematics

City University of Hong Kong

Kowloon, Hong Kong

[§] Amgen Inc.

Thousand Oaks, CA, 91320

SUMMARY: In clinical trials, minimum clinically important difference (MCID) has attracted increasing interest as an important supportive clinical and statistical inference tool. Many estimation methods have been developed based on various intuitions, while little theoretical justification has been established. This paper proposes a new estimation framework of the MCID using both diagnostic measurements and patient-reported outcomes (PRO's). The framework first formulates the population-based MCID as a large margin classification problem, and then extends to the personalized MCID to allow individualized thresholding value for patients whose clinical profiles may affect their PRO responses. More importantly, the proposed estimation framework is showed to be asymptotically consistent, and a finite-sample upper bound is established for its prediction accuracy compared against the ideal MCID. The advantage of our proposed method is also demonstrated in a variety of simulated experiments as well as two phase-3 clinical trials.

KEY WORDS: Fisher consistency, margin classification, minimum clinically important difference, non-convex minimization, support vector machine

1. Introduction

In clinical trials for drugs or medical devices, statistical significance is widely used to infer the treatment effect. However, there has been growing recognition that statistical significance could be misleading when evaluating treatment effect (Jacobson et al., 1984; Jacobson and Truax, 1991). First, in many trials, the statistical significance of the treatment effect may have little to do with its clinical significance. It is known that statistical significance only infers the existence of treatment effect, regardless of the effect size. Further, the statistical significance could result from a small sample variability or a huge sample size, and thus provides little information about the clinical meaningfulness of the treatment (Jacobson and Truax, 1991). Second, the statistical significance for the treatment group compared to the placebo group ignores the possible heterogeneity among individuals. For instance, in a pain reduction study, a statistically significant reduction is concluded for a test treatment while many individual patients in the treatment group actually report little improvement regarding the pain reduction (Younger et al., 2009).

Clinical significance is desired in practice as it provides a better assessment of the clinically meaningful improvement. It is often based on the patients' reports in a community according to certain external standards (Jacobson and Truax, 1991). One common approach is to collect patient-reported outcomes (PRO's; FDA, 2009), such as their satisfaction of a treatment. Some earlier practice suggested to replace the statistical significance tests by analyzing the PRO's only, which is problematic due to the subjective bias in the PRO's or unreliability of a poorly designed questionnaire. The minimum clinically important difference (MCID) was discussed in Jaeschke et al. (1989), which was intuitively defined as a thresholding value in post-treatment change, and a patient is considered experiencing a clinically meaningful improvement if her/his change exceeds the MCID. Copay et al. (2007) suggested to incorporate both certainty of effective treatment and patients' satisfactions for determining the MCID.

The concept of the MCID provides objective reference for clinicians and health policy makers regarding the effectiveness of the treatment, and has quickly gained its popularity among the practitioners. In 2012, U.S. Food and Drug Administration (FDA) hosted a special conference on the MCID for orthopaedic devices (`http://www.fda.gov/MedicalDevices/ NewsEvents/Workshops/Conferences/ucm327292.htm`). Although the importance of the MCID has been widely recognized, only a few approaches were proposed for its estimation based on receiver operating characteristic (ROC) curve. Bennett (1985), Leisenring et al. (2000) and Turner et al. (2010) discussed the trade-off of sensitivity and specificity, and formulated the MCID as the cutoff where sensitivity and specificity are equal. Shiu and Gatsonis (2008) focused on the trade-off of positive predictive value (PPV) and negative predictive value (NPV) and formulated the optimal cut as the value that maximizes the sum of PPV and NPV.

In this paper, the MCID is formulated as the thresholding value in post-treatment change such that the probability of disagreement between the estimated satisfaction based on the MCID and the PRO is minimized. With this framework, two scenarios are considered: the population-based MCID and the personalized MCID. The population-based MCID is the ideal thresholding value for the general population, and the personalized MCID allows different MCID values for individual patients based on their clinical profiles. Both scenarios can be formulated in a large margin classification framework, where the population-based MCID can be estimated via an exhaustive grid search, and the personalized MCID is modeled in a reproducing kernel Hilbert space and estimated via some non-convex optimization techniques. Most importantly, the asymptotic properties of the proposed estimation method are established for both the population-based and the personalized MCID's, and their fast convergence rates to the ideal performance are explicitly quantified.

The rest of the paper is organized as follows. In Section 2, a general framework for the

population-based MCID is presented, and its estimation algorithm and asymptotic properties are studied. Section 3 extends the framework to the personalized MCID, and discusses the appropriate large margin loss as well as the efficient non-convex optimization technique. Section 4 establishes the asymptotic properties of our proposed method for estimating the personalized MCID. Section 5 conducts numerical experiments of our proposed method in simulated examples, and Section 6 applies our proposed method to two phase-3 clinical trial datasets. Section 7 contains some discussion, and the technical proofs are deferred to supplemental materials which are available online.

2. A general framework of the MCID

2.1 *Formulating the MCID*

Suppose that a patient's diagnostic measurement $X \in \mathcal{R}^1$ is continuously connected, and the patient-reported outcome (PRO) $Y \in \{-1, 1\}$, where $Y = 1$ denotes a clinically meaningful treatment reported by the patient and $Y = -1$ otherwise. Let $f(x, y)$ and $f(x)$ be the joint density of (X, Y) and the marginal density of X, respectively. The MCID is formulated as the thresholding value c^* such that $\text{sign}(X - c^*)$ agrees with Y as much as possible, where $\text{sign}(u) = 1$ if $u \geqslant 0$ and -1 otherwise. Mathematically, c^* is defined as a solution of

$$\min_c \ P\{Y \neq \text{sign}(X - c)\} = \min_c \ \frac{1}{2} E\left\{1 - Y \text{sign}(X - c)\right\}, \tag{1}$$

where $P(\cdot)$ is evaluated with respect to both X and Y.

The formulation in (1) is closely related to the existing methods in literature. Shiu and Gatsonis (2008) aims to minimize the distance between predictive receiver operating characteristic (PROC) curve and perfect prediction, and proposes to estimate c by solving $\max_c \left\{ P(Y = 1, X \geqslant c)/P(X \geqslant c) + P(Y = -1, X < c)/P(X < c) \right\}$. Turner et al. (2010) concerns about the simple trade-off between sensitivity and specificity, and defines c as the solution of $P(Y = 1, X \geqslant c)/P(Y = 1) = P(Y = -1, X < c)/P(Y = -1)$. Note that the

proposed method in (1) can be rewritten as $\max_c \big\{ P(Y = 1, X \geqslant c) + P(Y = -1, X < c) \big\}$, which combines $P(Y = 1, X \geqslant c)$ and $P(Y = -1, X < c)$ into the agreement probability. Furthermore, the following Lemma 1 suggests that the ideal MCID based on (1) appears to be appropriate.

LEMMA 1: *Assume that $p(x) = P(Y = 1|X = x)$ is continuous and increasing in x, then the ideal MCID c^* satisfies*

$$p(c^*) = P(Y = 1|X = c^*) = \frac{1}{2}. \tag{2}$$

Furthermore, if $p(x)$ is strictly increasing in x, then c^* is the unique root of (2).

Proofs of all lemmas and theorems are given in Web Appendix A.

Note that it is reasonable to assume $p(x)$ to be increasing in x since patients with better diagnostic measurements are expected to be more likely to give positive responses. If $p(x)$ is only non-decreasing, the equation in (2) may have multiple roots and a conservative choice is to set c^* as the largest root. Furthermore, the continuity assumption of $p(x)$ can be relaxed to semi-continuity, and then the equation in (2) may have no root at all. In such scenarios, it could be proved similarly as Lemma 1 that $c^* = \operatorname{argmin}_c \{ p(c) \geqslant 1/2 \}$.

It is also known that the quality of the PRO's is largely affected by patients' subjectivity (Frost et al., 2007). Such subjectivity is accounted in the proposed formulation of MCID through $p(x)$, which can be interpreted as the probability of patient's telling the truth. For instance, Fang (2011) considered a special case of semi-continuous $p(x)$, and modeled the subjectivity explicitly as $p(x) = Q$ when $x \geqslant c^*$ and $p(x) = 1 - Q$ otherwise, where $Q > \frac{1}{2}$ measures how trustworthy the PRO's are. More importantly, the ideal MCID in (2) is less affected by the subjectivity in the PRO's, as it relies on $p(x)$ only when x is in the neighborhood of c^*. This is analogous to the Bayes rule in classification, which only relies on whether $p(x) \geqslant 1/2$ (Lin, 2002).

In addition, the MCID has an interesting connection with the median lethal dose in

toxicology research. The median lethal dose refers to the smallest dose required to kill half of the animals that receive that dose after a specified test duration. To describe the interaction between dosage and mortality rate, the logistic dose-response curve is popularly used (Williams, 1986; Alho and Valtonen, 1995; Kelly, 2001). It assumes that the mortality rate is expected to strictly increase with dose, which coincides with our assumption in Lemma 1.

2.2 *Estimating the MCID*

The primary interest of this paper is to estimate the MCID, which is in sharp contrast to the standard classification that focuses on the classification boundary. In (2), the ideal MCID c^* is defined based on $p(x)$ that is often unavailable in practice, so the MCID needs to be estimated based on the training sample $(x_i, y_i)_{i=1}^n$.

Naturally, the expectation in (1) can be approximated by its empirical version, and the estimated MCID $\widehat{c}$ is defined as a solution of

$$\min_c \ \frac{1}{2n} \sum_{i=1}^n \left\{ 1 - y_i \operatorname{sign}(x_i - c) \right\}. \tag{3}$$

Note that (3) is a simple 1-dimensional optimization problem, and the objective function remains the same for $x_{(i)} \leqslant c < x_{(i+1)}$, where $x_{(i)}$ is the i-th order statistic. Therefore, an exhaustive grid search scheme can be implemented, and the global minimizer $\widehat{c}$ is simply the x_i that yields the smallest objective function value.

THEOREM 1: *The estimated MCID $\widehat{c}$ in (3) is a consistent estimate of c^* if $p(x)$ is continuous and strictly increasing in x. Further, if there exist positive constants α_1, $\gamma_1 < 2/\alpha_1 + 4/\alpha_1^2$, a_1 and a_2, such that for sufficiently small $\xi > 0$,*

$$P(|p(X) - 1/2| \leqslant \xi) \leqslant a_1 \xi^{\alpha_1}, \tag{4}$$

$$\sup_{|x - c^*| \leqslant \xi} |p(x) - 1/2| \leqslant a_2 \xi^{\gamma_1}, \tag{5}$$

then $|\widehat{c} - c^*| = O_p\left((n \log^{-2} n)^{-1/(2(1+2/\alpha_1) - \alpha_1 \gamma_1)} \right)$.

Theorem 1 establishes the asymptotic convergence rate of $|\widehat{c} - c^*|$, and the finite sample bound for $|\widehat{c} - c^*|$ can also be obtained as in supplementary materials. In Theorem 1, (4) is similar to the low noise assumption (Polonik, 1995; Bartlett et al., 2003; Tsybakov, 2004) that describes the behavior of X in the neighborhood of c^*, and (5) is implied by a Hölder continuity condition on $p(x)$. For illustration, if X is uniformly distributed on $[a, b]$ and (5) is met with γ_1, then (4) can be verified with $\alpha_1 = 1/\gamma_1$ for sufficiently small ξ. Theorem 1 then implies that $|\widehat{c} - c^*| = O_p\big((n \log^{-2} n)^{-1/(1+4\gamma_1)}\big)$. It leads to a fast convergence rate when $p(x)$ has a steep derivative at c^* with γ_1 close to 0, and a rate of $O_p\big(n^{-1/3}(\log n)^{2/3}\big)$ when (5) holds with order $\gamma_1 = 1/2$.

3. The personalized MCID

In many clinical trials, it is commonly believed that patients' report could be influenced by various factors such as their expectation of treatment (Wise, 2004). For instance, in a shoulder pain reduction study, healthy people demonstrate a higher threshold than those with chronic conditions due to their expectation of complete recovery. In literature, covariate-adjusted ROC curve was developed to incorporate the effect of factors (Alonzo and Pepe, 2002), however very little work has been done in estimating covariate-adjusted MCID. To allow MCID to vary according to patients' clinical profiles, this section extends the estimation framework to the personalized MCID.

3.1 *Formulation*

Let $\mathbf{z}$ denote patients' clinical profiles. The personalized MCID $c^*(\mathbf{z})$ is formulated as a solution of

$$\min_c P\big[Y \neq \text{sign}\{X - c(\mathbf{Z})\}\big] = \min_c \frac{1}{2} E\big[1 - Y \, \text{sign}\{X - c(\mathbf{Z})\}\big], \qquad (6)$$

where P is taken with respect to $(X, Y, \mathbf{Z})$. Similarly as in (2), we can show that $c^*(\mathbf{z})$ satisfies

$$p_{\mathbf{z}}\{c^*(\mathbf{z})\} = P\{Y = 1 | X = c^*(\mathbf{z}), \mathbf{Z} = \mathbf{z}\} = \frac{1}{2}, \qquad (7)$$

where $p_{\mathbf{z}}(x) = P(Y = 1|X = x, \mathbf{Z} = \mathbf{z})$ is assumed to be a continuous and strictly increasing function in x for any value of $\mathbf{z}$. If only semi-continuity is assumed, the personalized MCID can be formulated as $c^*(\mathbf{z}) = \operatorname{argmin}_c \{c : p_{\mathbf{z}}(c) \geqslant \frac{1}{2}\}$. It is worth pointing out that the personalized MCID in (6) differs from the Bayes rule in classification in that the candidate function in (6) has to take the form of $x - c(\mathbf{z})$ in order to estimate $c^*(\mathbf{z})$, whereas a Bayes rule in classification searches for the optimal classification function $g(x, \mathbf{z})$ that may not lead to an explicit estimator of $c^*(\mathbf{z})$.

The formulation in (6) is similar as in (1) with population-based c^*, but the difficulty arises in the estimation part. Since the empirical version of (6)

$$\min_c \frac{1}{2n} \sum_{i=1}^n \left[1 - y_i \operatorname{sign}\{x_i - c(\mathbf{z}_i)\} \right], \tag{8}$$

involves the 0-1 loss $L_{01}(u) = \frac{1}{2}(1 - \operatorname{sign}(u))$ and needs to be optimized with respect to functional $c(\mathbf{z})$, it can no longer be solved by the exhaustive grid search or any other efficient optimization techniques. Therefore, a surrogate loss function needs to be introduced to replace the 0-1 loss and facilitate the estimation. The surrogate loss has been widely studied in machine learning literature. Popularly used surrogate loss functions $L(u)$ include the hinge loss $L(u) = (1 - u)_+$ (Vapnik, 1998), the logistic loss $L(u) = \log(1 + \exp(-u))$ (Zhu and Hastie, 2005), and the ψ-loss $\min((1 - u)_+, 1)$ (Shen et al., 2003; Liu and Shen, 2006). However, all these losses are not generally Fisher consistent in estimating $c^*(\mathbf{z})$, as the candidate function in (6) is restricted to the form of $x - c(\mathbf{z})$ for estimating the personalized MCID. Counter examples are provided in Web Appendix B.

In this paper, we propose a novel surrogate loss, ψ_δ-loss, which is defined as

$$L_\delta(u) = \min \left\{ \frac{1}{\delta}(\delta - u)_+, 1 \right\}. \tag{9}$$

The ψ_δ-loss extends the ψ-loss by introducing a new parameter δ that controls the difference between the surrogate loss and the 0-1 loss. More importantly, Lemma 2 shows that the ψ_δ-loss is asymptotically Fisher consistent in estimating $c^*(\mathbf{z})$ when δ converges to 0.

LEMMA 2:   *For any given* $\mathbf{z}$, *if the conditional density* $f_{\mathbf{z}}(x)$ *is continuous and* $p_{\mathbf{z}}(x)$ *is strictly increasing in* x, *then as* $\delta \to 0$, $E\Big[L_\delta\{Y(X-c)\}|\mathbf{Z}=\mathbf{z}\Big]$ *converges to* $E\Big[L_{01}\{Y(X-c)\}|\mathbf{Z}=\mathbf{z}\Big]$ *uniformly over a compact set* $\mathcal{D}_{\mathbf{z}}$ *containing* $c^*(\mathbf{z})$ *and*

$$\operatorname*{argmin}_c E\Big[L_\delta\{Y(X-c(\mathbf{z}))\}|\mathbf{Z}=\mathbf{z}\Big] \longrightarrow c^*(\mathbf{z}).$$

With the ψ_δ-loss, the proposed estimation formulation for the personalized MCID $\widehat{c}(\mathbf{z})$ solves

$$\min_{c \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} L_\delta[y_i\{x_i - c(\mathbf{z}_i)\}] + \lambda J(c), \tag{10}$$

where λ is a tuning parameter, $J(c)$ is a penalty term, and $\mathcal{F}$ is set as a reproducing kernel Hilbert space (RKHS) (Wahba, 1990). The final estimation formulation then becomes

$$\min_{c \in \mathcal{H}_K} \frac{1}{n} \sum_{i=1}^{n} L_\delta[y_i\{x_i - c(\mathbf{z}_i)\}] + \frac{\lambda}{2}\|c\|^2_{\mathcal{H}_K}, \tag{11}$$

where $\mathcal{H}_K$ is the RKHS induced by some pre-specified kernel function $K(\cdot, \cdot)$, and $J(c) = \frac{1}{2}\|c\|^2_{\mathcal{H}_K}$ is the associated RKHS norm of $c(\mathbf{z})$. It follows from the representer theorem (Wahba, 1990) that the solution to (11) is of the form $\widehat{c}(\mathbf{z}) = b + \sum_{i=1}^{n} w_i K(\mathbf{z}_i, \mathbf{z})$, and thus $\|c\|^2_{\mathcal{H}_K} = \mathbf{w}^T\mathbf{K}\mathbf{w}$ with $\mathbf{w} = (w_1, \cdots, w_n)^T$ and $\mathbf{K} = (K(\mathbf{z}_i, \mathbf{z}_j))_{i,j=1}^{n}$.

3.2 *Non-convex optimization*

Note that the cost function in (11) is non-convex, and thus we employ the difference convex algorithm (DCA) (An and Tao, 1997) to tackle the non-convex optimization. The key idea of the DCA is to decompose the non-convex cost function into the difference of two convex functions, and then construct a sequence of subproblems by approximating the second convex function with its affine minorization function.

In particular, the ψ_δ-loss is decomposed as

$$L_\delta(u) = \min\left\{\frac{1}{\delta}(\delta - u)_+, 1\right\} = \frac{1}{\delta}(\delta - u)_+ - \frac{1}{\delta}(-u)_+.$$

Then the cost function in (11) can be decomposed as $s(\widetilde{\mathbf{w}}) = s_1(\widetilde{\mathbf{w}}) - s_2(\widetilde{\mathbf{w}})$, where

$$s(\widetilde{\mathbf{w}}) = \frac{1}{n}\sum_{i=1}^{n} L_\delta\{y_i(x_i - c(\mathbf{z}_i))\} + \frac{\lambda}{2}\|c\|^2_{\mathcal{H}_K},$$

$$s_1(\widetilde{\mathbf{w}}) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{\delta}\{\delta - y_i(x_i - c(\mathbf{z}_i))\}_+\right] + \frac{\lambda}{2}\|c\|^2_{\mathcal{H}_K},$$

$$s_2(\widetilde{\mathbf{w}}) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{\delta}\{-y_i(x_i - c(\mathbf{z}_i))\}_+\right],$$

and $\widetilde{\mathbf{w}} = (\mathbf{w}, b)$ is the coefficient vector for the RKHS representation of $c(\mathbf{z})$.

Next, the DCA constructs a sequence of decreasing upper envelop of $s(\widetilde{\mathbf{w}})$ by approximating $s_2(\widetilde{\mathbf{w}})$ with its affine minorization function, $s_2(\widetilde{\mathbf{w}}^{(k)}) + \langle \widetilde{\mathbf{w}} - \widetilde{\mathbf{w}}^{(k)}, \nabla s_2(\widetilde{\mathbf{w}}^{(k)})\rangle$, where $\widetilde{\mathbf{w}}^{(k)}$ is the estimated $\widetilde{\mathbf{w}}$ at the k-th iteration, and $\nabla s_2(\widetilde{\mathbf{w}}^{(k)})$ is the subgradient of $s_2(\widetilde{\mathbf{w}})$ at $\widetilde{\mathbf{w}}^{(k)}$. The updated $\widetilde{\mathbf{w}}^{(k+1)}$ is then obtained by solving

$$\widetilde{\mathbf{w}}^{(k+1)} = \underset{\widetilde{\mathbf{w}}}{\operatorname{argmin}}\ s_1(\widetilde{\mathbf{w}}) - s_2(\widetilde{\mathbf{w}}^{(k)}) - \langle \widetilde{\mathbf{w}} - \widetilde{\mathbf{w}}^{(k)}, \nabla s_2(\widetilde{\mathbf{w}}^{(k)})\rangle. \tag{12}$$

The updating scheme is iterated until convergence. Although the DCA cannot guarantee global optimum, it delivers a superior numerical performance as demonstrated in the extensive simulation study in Liu et al. (2005).

4. Asymptotic theory

This section quantifies the asymptotic behavior of $\widehat{c}(\mathbf{z})$ in estimating the personalized MCID. Denote $e_{\delta_n}(\widehat{c}, c^*) = E\left[L_{\delta_n}\{Y(X - \widehat{c}(\mathbf{Z}))\} - L_{\delta_n}\{Y(X - c^*(\mathbf{Z}))\}\right]$ with $\delta_n > 0$, where δ and λ are rewritten as δ_n and λ_n to denote their dependency on n. We make the following four technical assumptions.

Assumption A. For some positive sequence $s_n \to 0$ as $n \to \infty$, there exists $c_0(\mathbf{z}) \in \mathcal{F}$, such that for sufficiently small δ_n, $e_{\delta_n}(c_0, c^*) \leqslant s_n$. That is, $\inf_{\{c \in \mathcal{F}\}} e_{\delta_n}(c, c^*) \leqslant s_n$.

Assumption A is standard (Shen et al., 2003; Li et al., 2007), and describes the approximation error of $\mathcal{F}$ in approximating $c^*(\mathbf{z})$.

Assumption B. There exist constants $0 < \alpha_2 < +\infty$ and $a_3 > 0$ such that for any given $\mathbf{z}$,

$P(|p_{\mathbf{z}}(X) - p_{\mathbf{z}}(c^*(\mathbf{z}))| \leqslant \xi) \leqslant a_3 \xi^{\alpha_2}$ for sufficiently small ξ.

Assumption B is the low noise assumption that describes the distribution of the diagnostic

outcome X in the neighborhood of $c^*(\mathbf{z})$.

Assumption C. There exist constants $0 < \gamma_2 < +\infty$ and $a_4 > 0$ such that for any given $\mathbf{z}$,

$\sup_{|x-c^*(\mathbf{z})| \leqslant \xi} |p_{\mathbf{z}}(x) - p_{\mathbf{z}}(c^*(\mathbf{z}))| \leqslant a_4 \xi^{\gamma_2}$ for sufficiently small ξ.

Assumption C is implied by a Hölder continuity condition that describes the smoothness

of $p_{\mathbf{z}}(x)$ around $c^*(\mathbf{z})$.

Before specifying Assumption D, we first define the metric entropy for any give set. For a

given class $\mathcal{B}$ of subsets of S and any $\epsilon > 0$, $\{(G_1^l, G_1^u, \cdots, G_m^l, G_m^u)\}$ forms an ϵ-bracketing

set of $\mathcal{B}$ if for any $G \in \mathcal{B}$ there is a j such that $G_j^l \subset G \subset G_j^u$ and $\max_{\{1 \leqslant j \leqslant m\}} d(G_j^u, G_j^l) \leqslant$

ϵ, where $d(\cdot, \cdot)$ is a distance for any two subsets in S defined as $d(G_1, G_2) = P(G_1 \Delta G_2)$

and $G_1 \Delta G_2 = (G_1 \backslash G_2) \bigcup (G_2 \backslash G_1)$. Then the metric entropy $H(\epsilon, \mathcal{B})$ of $\mathcal{B}$ is defined as the

logarithm of the cardinality of the ϵ-bracketing set of $\mathcal{B}$ of the smallest size. Let $\mathcal{G}(k) =$

$\{G_c = \{(x, \mathbf{z}) : x - c(\mathbf{z}) \geqslant 0\}, c \in \mathcal{F}, J(c) \leqslant k\} \subset \mathcal{G}(\mathcal{F}) = \{G_c = \{(x, \mathbf{z}) : x - c(\mathbf{z}) \geqslant 0\}, c \in$

$\mathcal{F}, J(c) < +\infty\}$.

Assumption D. For positive constants a_5, a_6 and a_7, there exists some $\epsilon_n > 0$ such that

$$\sup_{\{k \geqslant 1\}} \phi(\epsilon_n, k) \leqslant a_5 n^{1/2},$$

where $\phi(\epsilon_n, k) = \int_{a_7 L}^{(8a_6)^{1/2} L^{\alpha/2(\alpha+\gamma)}} H^{1/2}(u^2/2, \mathcal{G}(k)) du / L$ and $L = L(\epsilon_n, C, k) = \min(\epsilon_n^2 +$

$\lambda_n J_0(k/2 - 1), 1)$.

THEOREM 2: *Suppose that Assumptions A-D are met and $p_{\mathbf{z}}(x)$ is strictly increasing in*

x. For the estimated personalized MCID $\widehat{c}(\mathbf{z})$, there exists positive constants a_8 and a_9 such

that

$$P\left[|\widehat{c}(\mathbf{Z}) - c^*(\mathbf{Z})| \geqslant \{\beta_n^2 \log(1/\beta_n^2)\}^{\frac{\alpha_2}{\alpha_2+2}}\right] \leqslant 3.5 \exp\left\{-a_8 n (\lambda_n J(c_0))^{\frac{\alpha_2+2}{\alpha_2+1}}\right\} + a_9 \{\log(1/\beta_n^2)\}^{-1}.$$

provided that $\beta_n^2 \geqslant 4\lambda_n \max(J(c_0), 1)$ *with* $\beta_n^2 = \min(\max(\epsilon_n^2, 2s_n + 2a_3 a_4^{\alpha_2} \delta_n^{\alpha_2 \gamma_2}), 1)$ *and* $f_z(c^*(\mathbf{z}))$ *is bounded away from 0.*

COROLLARY 1: *Under the assumptions of Theorem 2,* $|\widehat{c}(\mathbf{Z}) - c^*(\mathbf{Z})| = O_p\big[\{\beta_n^2 \log(1/\beta_n^2)\}^{\frac{\alpha_2}{\alpha_2 + 2}}\big]$, *provided that* $n\{\lambda_n J(c_0)\}^{\frac{\alpha_2 + 2}{\alpha_2 + 1}}$ *is bounded away from 0.*

Theorem 2 and Corollary 1 develop upper bounds for the estimation accuracy of the estimated $\widehat{c}(\mathbf{z})$. The convergence rate $\beta_n^{\frac{2\alpha_2}{\alpha_2 + 2}}$ in Corollary 1 depends on the values of δ_n, ϵ_n^2, s_n and λ_n. More importantly, such results can be difficult to establish for the standard classification function $g(x, \mathbf{z})$ due to its lack of explicit estimation of $c^*(\mathbf{z})$. For illustration, supposed that $c^*(Z) = Z$, where Z is uniformly distributed on $[-100, 100]$. For any given z, X is uniformly distributed on $[z - 1, z + 1]$ and $p_z(x) = P(X \leqslant x | Z = z)$, and $\mathcal{F}$ is set as $\{c(z) = wz + b : w, b \in \mathcal{R}\}$. All assumptions can be verified with $s_n = 1/n$, $\alpha_2 = 1$, $\gamma_2 = 1$, $\epsilon_n = (\log n / n)^{1/2}$ and $\beta_n^2 = \log n / n$ when δ_n is sufficiently small. Then applying Corollary 1 yields that $|\widehat{c}(Z) - c^*(Z)| = O_p(\log^2 n / n)^{1/3}$. More details could be found in Xu (2013).

5. Simulation

This section examines the proposed estimation methods for estimating the MCID using simulated examples. Two scenarios are considered. Scenario I focuses on the population-based MCID for all patients, and scenario II focuses on the personalized MCID that varies among patients and relies on each patient clinical profile. To assess the estimation performance, we report the estimated MCID as well as the misclassification error (MCE) based on the testing set, which is defined as

$$MCE(\widehat{c}) = \frac{1}{n_{test}} \sum_{i \in testing\ set} I\{y_i \neq \text{sign}(x_i - \widehat{c}(\mathbf{z}_i))\},$$

where n_{test} denotes the size of the testing set, and $\widehat{c}(\mathbf{z}_i) = \widehat{c}$ for the population-based MCID. In addition, the sensitivity, specificity, PPV, and NPV are also reported to assess the classification performance.

5.1 *Scenario I: the population-based MCID*

Two simulated examples are examined.

Example 1. A random sample $\{(X_i, Y_i); i = 1, \cdots, n + 2000\}$ is generated as follows. First, X_i is generated from $Unif(-1, 1)$ and then Y_i is generated from $Bern((x_i + 1)/2)$. Next, a sample of size n is randomly selected for training and the remaining 2000 observations are for testing.

Example 2. A random sample $\{(X_i, Y_i); i = 1, \ldots, n + 2000\}$ is generated as follows. First, X_i is generated from the mixture of two Gaussian distributions $0.7N(-1, 1) + 0.3N(1, 1)$ and then Y_i is generated from $Bern(0.75F(x_i))$, where $F(x_i) = P(X \leqslant x_i)$. Next, a sample of size n is randomly selected for training and the remaining 2000 observations are for testing.

The training sizes are set as $n = 250, 500$ and 1000, and both examples are replicated 100 times. The averaged performance measures of our proposed method, Turner et al. (2010) and Shiu and Gatsonis (2008) are summarized in Table 1. In addition, the ideal MCID's and their corresponding performance measures are used as baseline for the comparison in Table 1.

[Table 1 about here.]

In both examples, our proposed method yields accurate MCID estimates that are very close to the ideal MCID's. The resulting MCE's are also close to the MCE's produced by using the ideal MCID's. In Example 1, the performance of Turner et al. (2010) is also competitive due to the fact that $P(X \geqslant c^*|Y = 1) = P(X < c^*|Y = -1)$. However, in Example 2 where the equality is no longer true, the performance of Turner et al. (2010) becomes less satisfactory for all measures. The MCE of Shiu and Gatsonis (2008) appears to be less competitive. Even with a large sample size $n = 1000$, their estimated MCID's are still considerably different from the ideal MCID's. Note that Shiu and Gatsonis (2008) is designed to maximize PPV+NPV, but Table 1 shows that our proposed method also yields competitive PPV+NPV across all

scenarios. In addition, the estimated sensitivity and specificity further demonstrate that our proposed method is superior to this method.

5.2 *Scenario II: the personalized MCID*

For the personalized MCID, the MCE by using our proposed method with linear and Gaussian kernels are examined. The linear kernel is defined as $K(\mathbf{z}_1, \mathbf{z}_2) = \mathbf{z}_1^T \mathbf{z}_2$, and the Gaussian kernel is defined as $K(\mathbf{z}_1, \mathbf{z}_2) = e^{-\|\mathbf{z}_1 - \mathbf{z}_2\|^2/2\sigma^2}$, where the scale parameter σ^2 is set as the median of pairwise Euclidean distances within the training set. To optimize the performance of our proposed method, a grid search by 5-fold cross validation is employed to select the tuning parameter λ. The grid for all examples is set as $\{10^{(s-31)/10}; s = 1, \cdots, 61\}$. For illustration, three simulated examples are examined with $\delta = 0.1$.

Example 1. A random sample $\{(X_i, Y_i, \mathbf{Z}_i); i = 1, \cdots, n\}$ is generated as follows. First, $\mathbf{Z}_i$'s are independently generated from $N_2(\boldsymbol{\mu}, \boldsymbol{I}_2)$ with $\boldsymbol{\mu} = (0,0)^T$ and $\boldsymbol{I}_2 = diag(1,1)$. Second, X_i's are independently generated from $N(b + \boldsymbol{\beta}^T \mathbf{z}_i, 1)$, where $b = 0$ and $\boldsymbol{\beta} = (1,2)^T$. Next, the response Y_i is generated from $Bern(F(x_i))$, where $F(x_i) = P(X_i \leqslant x_i)$.

Example 2. A random sample $\{(X_i, Y_i, \mathbf{Z}_i); i = 1, \cdots, n\}$ is generated as follows. First, $\mathbf{Z}_i$'s are independently generated from $N_2(\boldsymbol{\mu}, \boldsymbol{I}_2)$ with $\boldsymbol{\mu} = (0,0)^T$. Second, X_i's are independently generated from $N(b + \boldsymbol{\beta}^T \mathbf{z}_i - \boldsymbol{\beta}^T \mathbf{z}_i^2, 1)$, where $b = 0$ and $\boldsymbol{\beta} = (1,2)^T$. Next, the response Y_i is generated from $Bern(F(x_i))$, where $F(x_i) = P(X_i \leqslant x_i)$.

Example 3. A random sample $\{(X_i, Y_i, \mathbf{Z}_i); i = 1, \cdots, n\}$ is generated as follows. First, Z_i's are independently generated from $N_3(\boldsymbol{\mu}, \boldsymbol{I}_3))$ with $\boldsymbol{\mu} = (0,0,0)^T$ and $\boldsymbol{I}_3 = diag(1,1,1)$. Second, X_i's are independently generated from $N(b + \cos(\boldsymbol{\beta}^T \mathbf{z}_i), 1)$, where $b = 0$ and $\boldsymbol{\beta} = (1, 1.5, 2)^T$. Next, the response Y_i is generated from $Bern(F(x_i))$, where $F(x_i) = P(X_i \leqslant x_i)$.

For each example, the training sizes are set as 100, 250, 500 and the testing size is set as 2000. All examples are replicated 50 times, and the averaged test errors are reported in Table 2.

[Table 2 about here.]

Our proposed method delivers satisfactory performance in estimating the personalized MCID in all three examples. In addition, the linear kernel yields slightly better performance than the Gaussian kernel in Example 1 as the true classification boundary is linear, and it is outperformed by the Gaussian kernel in the other two examples with nonlinear boundaries. Therefore, the Gaussian kernel would be suggested if no prior knowledge about the boundary is available.

For estimating the personalized MCID, the choice of δ may impact the performance of our proposed method. By Theorem 2, large δ leads to less accurate prediction while computational instability may occur when small δ is used for the estimation. For illustration, we conducted a sensitivity analysis on the values of δ in Example 1 with training size 250. Table 3 consists of the averaged MCE, the estimated coefficients, and the computational time based on 100 replications.

[Table 3 about here.]

The estimated coefficients and prediction error in a random selected replication as functions of $\delta \in [0.01, 2]$ are displayed in Figure 1.

[Figure 1 about here.]

It is clear that the numerical performance is relatively stable when δ is reasonably close to 0. However, when δ is too large or too small, the estimation of $c(\mathbf{z})$ deviates away from the truth and leads to deteriorated estimation performance. It is also interesting to note that when δ is too small, the computational instability often leads to early stop of the algorithm, and thus shorter computing time but less competitive performance. Based on the sensitivity study, we recommend to set δ as 0.1 in practice.

6. Real applications

In this section, our proposed method is applied to a phase-3 woman heavy menstrual blood loss dataset (WHMBL) and a phase-3 hot flush dataset (Hot Flush).

The WHMBL clinical trial aims to develop a treatment for reducing the amount of blood loss during a menstrual cycle in excessive bleeding women. The primary efficacy variable is the change from baseline in blood loss volume. The blood loss of each patient is measured per menstrual cycle and the PRO's are collected based on a questionnaire answered by each patient at a post-treatment visit. The WHMBL trial dataset consists of 481 patients administered either placebo or active doses. Patient profile contains the information of age, body mass index (BMI), alcohol (Yes/No), tobacco (Yes/No) and baseline value of blood loss. The 481 patients were randomly split into a training set of 240 patients and a testing set of 241 patients.

The hot flush clinical trial aims to develop a treatment for reducing hot flush in women due to menopause. The primary efficacy variable is the change from baseline in average daily frequency of hot flushes, and the PRO's are collected based on a questionnaire answered by each patient at a post-treatment visit. The hot flush clinical trial dataset consists of 1684 patients administered either placebo or active doses. Patient profile contains the information for age, BMI, race and baseline hot flushes. 300 patients were selected randomly to form the training set and the remaining 1384 patients were used as the testing set.

Here, $\delta = 0.1$ is used for simplicity and the tuning parameter λ is selected as in Section 5.2. Each example is replicated 50 times, and Table 4 summarizes the averaged performance measures of the method by Shiu and Gatsonis (2008), Turner et al. (2010), the population-based MCID, and the personalized MCID with the linear and Gaussian kernels.

[Table 4 about here.]

In both scenarios, our proposed method delivers competitive performance in comparison

with the methods by Shiu and Gatsonis (2008) and by Turner et al. (2010). In WHMBL trial, the method by Shiu and Gatsonis (2008) yields a negative MCID which is clinically misleading. It is also interesting to notice that for the WHMBL trial, the personalized MCID yields larger MCE when compared with the population-based MCID. It could be due to the homogeneity among the enrolled patients. For the hot flush trial, patients' satisfaction on treatment effect is more accurately estimated when the clinical profiles are included. A closer investigation of the fitted classification function implies that patients' satisfaction is highly affected by the baseline hot flushes. This is reasonable as patients with higher baseline hot flushes tend to expect better treatment effect.

7. Closing remarks

The concept of MCID has attracted much attention in clinical trials, while little statistical work has been done for appropriately determining MCID. This paper proposes a general framework for formulating as well as estimating the population-based and the personalized MCID's. Our proposed method unifies both the population-based and the personalized MCID's into a large margin classification framework, and delivers superior estimation performance in both simulated examples and real applications to two phase-3 clinical trials. More importantly, the asymptotic properties of our proposed method are established for both the population-based and the personalized MCID's.

As for potential extensions, the proposed estimation framework could be extended to scenarios with multiple diagnostic test outcomes, where a prognostic score function can be constructed to combine the multiple test outcomes. It is also of interest to extend the framework to estimation of other types of optimal cut points based on ROC curve such as the Youden index.

8. Supplementary materials

The Web Appendices referenced in Sections 2 and 3, and the implemented codes are available with this paper at the Biometrics website on Wiley Online Library.

References

U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, Center for Devices and Radiological Health (2009). Guidance for industry. Patient-report outcome measures: use in medical product development to support labeling claims. *U.S. Department of Health and Human Services, Rockville, MD.*

Alho, J. and Valtonen, E. (1995). Interval estimation of inverse dose-response. *Biometrics*, **51**, 491-501.

Alonzo, T. and Pepe, M. (2002). Distribution-free ROC analysis using binary regression techniques. *Biometrics*, **3**, 421-432.

An, L. and Tao, P. (1997). Solving a class of linearly constrained indefinite quadratic problems by D.C. algorithms. *Journal of Global Optimization*, **11**, 253-285.

Bartlett, P., Jordan, M. and McAuliffe, J. (2003). Large margin classifiers: convex loss, low noise, and convergence rates. *In proceeding of: Advances in Neural Information Processing Systems 16, NIPS.*

Bennett, B. (1985). On tests for equality of predictive values for t diagnostic procedures. *Statistics in Medicine*, **4**, 535-539.

Copay, A., Subach, B., Glassman, S.,Polly, Jr D. and Schuler, T. (2007). Understanding the minimum clinically important difference: a review of concepts and methods. *The Spine Journal*, **7**, 541-546.

Fang, X. (2011). A new statistical method for estimating clinically meaningful threshold. *Joint Statistical Meetings Proceedings 2011.*

Frost, M., Reeve, B., Liepa, A., Stauffer, J., Hays, R. and Mayo/FDA Patient-Reported Outcomes Consensus Meeting Group (2007). What is sufficient evidence for the reliability and validity of patient-reported outcome measures? *Value Health*, **10(2S)**, S94 - S105.

Jacobson, N., Follette, W. and Revenstorf, D. (1984). Psychotherapy outcome research: methods for reporting variability and evaluating clinical significance. *Behavior Therapy.* **15**, 336-352.

Jacobson, N. and Truax, P. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, **59**, 12-19.

Jaeschke, R., Singer, J. and Guyatt, G. (1989). Measurement of health status: ascertaining the minimal clinically important difference. *Controlled Clinical Trails*, **10**, 407-415.

Kelly, G. (2001). The median lethal dose-design and estimation. *The Statistician*, **50**, 41-50.

Leisenring, W., Alonzo, T. and Pepe, M. (2000). Comparisons of predictive values of binary

medical diagnostic tests for paired designs. *Biometrics*, **56**, 345-351.

Li, Y., Liu, Y. and Zhu, J. (2007). Quantile regression in reproducing kernel Hilbert spaces. *Journal of the American Statistical Association*, **102**, 255-268.

Lin, Y. (2002). Support vector machines and the Bayes rule in classification. *Data Mining and Knowledge Discovery*, **6**, 259-275.

Liu, S., Shen, X. and Wong W. (2005). Computational development of ψ-learning. *Proceedings of the SIAM International Conference on Data Mining*, Newport, CA, 1-12.

Liu, Y. and Shen, X. (2006). Multicategory psi-learning. *Journal of the American Statistical Association*, **101**, 500-509.

Polonik, W. (1995). Measuring mass concentrations and estimating density contour clusters - an excess mass approach. *The Annals of Statistics*, **23**, 855-881.

Shen, X. and Wong, W. (1994). Convergence rate of sieve estimates. *The Annals of Statistics*, **22**, 580-615.

Shen, X., Tseng, G., Zhang, X. and Wong, W. (2003). On ψ-learning. *Journal of the American Statistical Association*, **98**, 724-734.

Shiu, S. and Gatsonis, C. (2008). The predictive receiver operating characteristic curve for the joint assessment of the positive and negative predictive values. *Philosophical Transactions of The Royal Society A*, **366**, 2313-2333.

Tsybakov, A. (2004). Optimal aggregation of classifiers in statistical learning. *The Annals of Statistics*, **32**, 135-166.

Turner, D., Schünemann, H., Griffith, L., Beaton, D., Griffiths, A., Critch, J. *et al.* (2010). The minimal detectable change cannot reliably replace the minimal important difference. *Journal of Clinical Epidemiology*, **63**, 28-36.

Vapnik, V. (1998). *Statistical learning theory*. Chichester, UK, Wiley.

Wahba, G. (1990) Spline models for observational data. Philadelphia: SIAM.

Williams, D. (1986) Interval estimation of the median lethal dose. *Biometrics*, **42**, 641-645.

Wise, E. (2004) Methods for analyzing psychotherapy outcomes: a review of clinical significance, reliable change, and recommendations for future directions. *Journal of Personality Assessment*, **82**, 50-59.

Xu, T. (2013) New developments of minimum clinically important difference: theory and methodology. *Ph.D. Dissertation*, University of Illinois at Chicago.

Younger, J., McCue, R. and Mackey, S. (2009) Pain outcomes: a brief review of instruments and techniques. *Current Pain and Headache Reports*, **13**, 39-43.

Zhu, J. and Hastie, T. (2005). Kernel logistic regression and the import vector machine. *Journal of Computational and Graphical Statistics*, **14**, 185-205.

Figure 1. Sensitivity analysis of δ in a randomly selected replication of Example 1 in Scenario II with $n = 250$. β_1, β_2 and b are the coefficients of $c(\mathbf{Z})$.
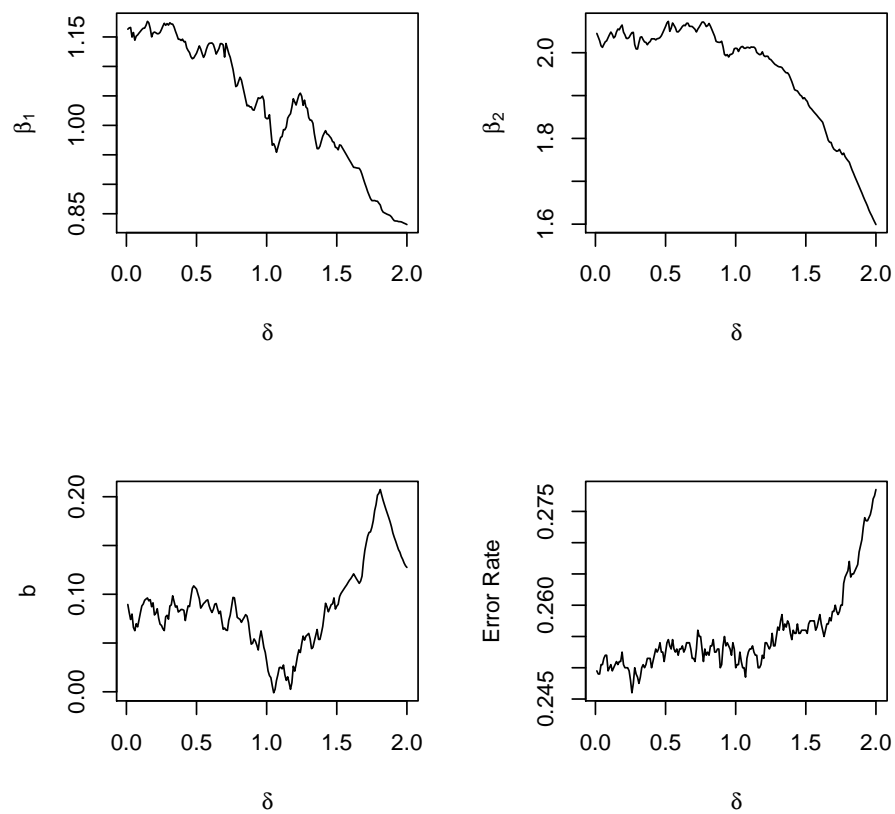
Table 1

Simulation I. Averaged MCID, the misclassification error (MCE), sensitivity (SEN), specificity (SPE), positive predictive value (PPV), negative predictive value (NPV), and their standard errors (in parentheses) for our method (OUR), the method by Shiu and Gatsonis (SG) and the method by Turner et al. (TU) based on 100 replications. The ideal performance is included as the baseline for comparison.

		n=250	n=500	n=1000	Ideal
Example 1					
MCID	OUR	0.055(0.0116)	-0.021(0.0058)	0.004(0.0032)	
	SG	0.078(0.0387)	-0.065(0.0290)	-0.080(0.0222)	0.000
	TU	0.004(0.0003)	0.002(0.0002)	0.003(0.0001)	
MCE	OUR	0.260(0.0010)	0.255(0.0005)	0.253(0.0003)	
	SG	0.344(0.0045)	0.355(0.0033)	0.374(0.0024)	0.250
	TU	0.251(0.0005)	0.251(0.0004)	0.250(0.0003)	
SEN	OUR	0.713(0.0061)	0.756(0.0030)	0.746(0.0016)	
	SG	0.618(0.0203)	0.677(0.0152)	0.666(0.0113)	0.750
	TU	0.748(0.0017)	0.748(0.0010)	0.748(0.0007)	
SPE	OUR	0.768(0.0057)	0.735(0.0029)	0.748(0.0016)	
	SG	0.696(0.0193)	0.613(0.0145)	0.586(0.0114)	0.750
	TU	0.751(0.0017)	0.750(0.0011)	0.751(0.0007)	
PPV	OUR	0.762(0.0029)	0.744(0.0015)	0.750(0.0008)	
	SG	0.769(0.0097)	0.733(0.0072)	0.730(0.0055)	0.750
	TU	0.751(0.0011)	0.750(0.0007)	0.750(0.0005)	
NPV	OUR	0.736(0.0031)	0.755(0.0016)	0.750(0.0009)	
	SG	0.731(0.0097)	0.767(0.0073)	0.770(0.0055)	0.750
	TU	0.749(0.0012)	0.749(0.0008)	0.749(0.0005)	
Example 2					
MCID	OUR	0.181(0.0269)	0.151(0.0153)	0.141(0.0083)	
	SG	2.374(0.0949)	2.470(0.0622)	2.762(0.0445)	0.142
	TU	-0.329(0.0071)	-0.338(0.0034)	-0.334(0.0018)	
MCE	OUR	0.298(0.0009)	0.296(0.0005)	0.295(0.0004)	
	SG	0.367(0.0035)	0.366(0.0020)	0.378(0.0013)	0.291
	TU	0.303(0.0007)	0.304(0.0005)	0.304(0.0003)	
SEN	OUR	0.542(0.0075)	0.568(0.0046)	0.553(0.0027)	
	SG	0.122(0.0156)	0.134(0.0131)	0.102(0.0082)	0.551
	TU	0.695(0.0025)	0.697(0.0011)	0.698(0.0008)	
SPE	OUR	0.800(0.0046)	0.784(0.0030)	0.797(0.0016)	
	SG	0.940(0.0105)	0.915(0.0102)	0.935(0.0064)	0.802
	TU	0.698(0.0022)	0.695(0.0014)	0.695(0.0006)	
PPV	OUR	0.626(0.0024)	0.619(0.0017)	0.625(0.0009)	
	SG	0.714(0.0097)	0.732(0.0089)	0.736(0.0071)	0.626
	TU	0.581(0.0012)	0.579(0.0007)	0.579(0.0005)	
NPV	OUR	0.748(0.0024)	0.755(0.0015)	0.751(0.0008)	
	SG	0.656(0.0046)	0.655(0.0043)	0.654(0.0027)	0.749
	TU	0.793(0.0010)	0.793(0.0007)	0.793(0.0004)	

Table 2

Simulation II. Estimated means and standard deviations (in parentheses) of the misclassification error by using our proposed method with linear and Gaussian kernels based on 50 replications.

	n=100	n=250	n=500	Ideal
Example 1				
Linear	0.256(0.0119)	0.254(0.0112)	0.250(0.0108)	0.250
Gaussian	0.280(0.0177)	0.270(0.0146)	0.259(0.0130)	
Example 2				
Linear	0.412(0.0146)	0.408(0.0140)	0.408(0.0095)	0.250
Gaussian	0.290(0.0169)	0.274(0.0133)	0.260(0.0118)	
Example 3				
Linear	0.315(0.0132)	0.313(0.0129)	0.318(0.0103)	0.250
Gaussian	0.323(0.0182)	0.308(0.0122)	0.293(0.0109)	

Table 3
Sensitivity analysis of δ of Example 1 in Scenario II with $n = 250$. The averaged coefficients of $c(\mathbf{Z})$, β_1, β_2 and b, the averaged MCE, and the averaged computing time (sec) are reported.

δ	MCE	β_1	β_2	b	time
10^{-10}	0.437(0.0037)	1.393(0.3734)	1.769(0.3028)	-0.099(0.0540)	1.546(0.0157)
10^{-5}	0.255(0.0011)	0.985(0.0078)	1.970(0.0094)	0.005(0.0095)	2.081(0.0251)
0.1	0.254(0.0012)	0.979(0.0096)	1.971(0.0078)	0.000(0.0120)	3.534(0.0726)
0.5	0.266(0.0016)	0.988(0.0245)	1.917(0.0176)	-0.051(0.0234)	4.302(0.1403)
1	0.319(0.0050)	0.765(0.0784)	1.553(0.0765)	-0.102(0.0937)	6.787(0.3805)
2	0.469(0.0017)	0.473(0.4243)	1.376(0.3528)	-0.302(0.3867)	4.621(0.1201)

Table 4

Real applications. Averaged MCID and misclassification error (MCE) and their standard errors(in parenthesis) by using the method by Shiu and Gatsonis (SG), Turner et al. (TU), the population-based MCID (OUR) and the personalized MCID with linear kernel (OUR_L) and Gaussian kernel (OUR_G) based on 50 replications.

	SG	TU	OUR	OUR_L	OUR_G
			WHMBL		
MCID	-45.004(3.3011)	37.383(0.4323)	20.610(0.4905)	-	-
MCE	0.436(0.0016)	0.368(0.0014)	0.358(0.0014)	0.365(0.0186)	0.376(0.0185)
			Hot Flush		
MCID	5.426(0.4453)	7.073(0.0127)	6.060(0.0229)	-	-
MCE	0.399(0.0049)	0.302(0.0005)	0.282(0.0005)	0.260(0.0054)	0.268(0.0031)