



## Interval estimation for minimal clinically important difference and its classification error via a bootstrap scheme

Zehua Zhou, Jiwei Zhao & Melissa Kluczynski

To cite this article: Zehua Zhou, Jiwei Zhao & Melissa Kluczynski (2020) Interval estimation for minimal clinically important difference and its classification error via a bootstrap scheme, *Statistical Theory and Related Fields*, 4:2, 135-145, DOI: [10.1080/24754269.2019.1587692](https://doi.org/10.1080/24754269.2019.1587692)

To link to this article: <https://doi.org/10.1080/24754269.2019.1587692>



Published online: 19 Mar 2019.



Submit your article to this journal [↗](#)



Article views: 36



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 1 View citing articles [↗](#)



# Interval estimation for minimal clinically important difference and its classification error via a bootstrap scheme

Zehua Zhou<sup>a</sup>, Jiwei Zhao <sup>a</sup> and Melissa Kluczynski <sup>b</sup>

<sup>a</sup>Department of Biostatistics, School of Public Health and Health Professions, State University of New York at Buffalo, Buffalo, NY, USA;

<sup>b</sup>Department of Orthopaedics, Jacobs School of Medicine and Biomedical Sciences, State University of New York at Buffalo, Buffalo, NY, USA

## ABSTRACT

With the improved knowledge on clinical relevance and more convenient access to the patient-reported outcome data, clinical researchers prefer to adopt minimal clinically important difference (MCID) rather than statistical significance as a testing standard to examine the effectiveness of certain intervention or treatment in clinical trials. A practical method to determining the MCID is based on the diagnostic measurement. By using this approach, the MCID can be formulated as the solution of a large margin classification problem. However, this method only produces the point estimation, hence lacks ways to evaluate its performance. In this paper, we introduce an  $m$ -out-of- $n$  bootstrap approach which provides the interval estimations for MCID and its classification error, an associated accuracy measure for performance assessment. A variety of extensive simulation studies are implemented to show the advantages of our proposed method. Analysis of the chondral lesions and meniscus procedures (ChAMP) trial is our motivating example and is used to illustrate our method.

## ARTICLE HISTORY

Received 3 December 2018

Revised 21 February 2019

Accepted 24 February 2019

## KEYWORDS

Minimal clinically important difference; classification error; confidence interval; non-convex optimisation; bootstrap;  $m$ -out-of- $n$  bootstrap

## 1. Introduction

Statistical significance is widely reported in clinical studies to infer treatment effect. For instance, in a randomised controlled trial to compare debridement to the observation of chondral lesions encountered during partial meniscectomy (Bisson et al., 2017), the difference of the patient outcomes before surgery and one year after is used to assess the existence of any statistically significant effect.

Although this framework based on a threshold of the  $p$ -value objectifies the research outcome, solely relying on it can have two potentially serious consequences. First, statistical significance only signifies the existence of treatment effect, no matter how large is the effect size. The statistical significance could result from a huge sample size, hence may clinically irrelevant to the patients at all. Second, a clinically importance effect could be classified as statistically non-significant due to various reasons, say, the small sample size in the study, hence be unfairly ignored. In brief statistical significance does not necessarily imply clinical importance, and vice versa.

Over the years clinical investigators are realising that the determination of a treatment's clinical importance is much more valuable and reliable than merely seeking its statistical significance. Also, the development of various patient-rated instruments contributes

huge amounts of patient-reported outcome (PRO) data, which provides the researchers with chances to study the clinical relevance. In order to study clinical importance, Jaeschke, Singer, and Guyatt (1989) proposed the concept of minimal clinically important difference (MCID). It is defined as the smallest change in an outcome that an individual patient would identify as important, therefore offers a threshold above which outcome is experienced as relevant by the patients. This avoids the problem of mere statistical significance (Wright, Hannon, Hegedus, & Kavchak, 2012). The MCID provides objective reference for clinicians and health policy-makers regarding the effectiveness of the treatment, hence has quickly gained its popularity (Erdogan, Leung, Pohl, Tennant, & Conaghan, 2016; McGlothlin & Lewis, 2014).

A variety of methods have been proposed to calculate the MCID. The anchor based method compares the changes in scores with an anchor as the reference. A popular anchor is the anchor question in the questionnaire. For instance, the short form (SF36) health survey (Ware & Sherbourne, 1992) serves this role in the ChAMP trial study (Bisson et al., 2017, 2018, 2019; Kluczynski et al., 2017). Hedayat, Wang, and Xu (2015) adopted this anchor based method and formulated the MCID as the threshold value in post-treatment change such that the probability of disagreement between the

estimated satisfaction based on the MCID and the PRO is minimised.

Although the proposal in Hedayat et al. (2015) possesses the statistical rigour and paves the way for potential extension, it has some limitations. First, Hedayat et al. (2015) rely on a testing data set with a very large sample size to have their method implemented and assessed; however, the sample size of a clinical study is usually much smaller, therefore such a testing data set is not available for real application. Second, Hedayat et al. (2015) only provide a point estimation for the MCID, which is not informative enough in most clinical studies (Cook, 2008; Erdogan et al., 2016). Without an interval estimation, it is unknown how accurate this point estimation is. Furthermore, without an interval estimation, we have no idea how to compare multiple MCIDs derived for different population subgroups, hence the population heterogeneity could not be learned.

In this paper, we aim at solving the problems mentioned above and filling in this gap in the literature. We first introduce the concept of classification error to gauge the effectiveness of the MCID. More importantly, this concept also allows us to compare MCIDs derived for different population subgroups or computed using different methods. Second, using the  $m$ -out-of- $n$  bootstrap technique, we obtain an interval estimation of the MCID, and also that of the classification error. The interval estimation makes it possible to conduct statistical inference on the MCID. It also allows us to fully learn the population heterogeneity based on the MCID.

Our proposal has two distinct features. First, different from Hedayat et al. (2015), our framework does not rely on a testing data set with a large sample size, hence can be conveniently used in various clinical studies. Second, although the bootstrap has already been a well-known and established statistical technique since Efron (1979), we stress that its conventional version cannot be directly applied in our context due to the restrictive conditions for it to be valid. Instead, the one we adopt is the  $m$ -out-of- $n$  bootstrap with its theoretical properties justified in Shao (1994), Shao (1996), Bickel, Götze, and van Zwet (1997) and among others.

In the remainder of this paper, we first introduce our motivating example, the ChAMP trial study in Section 2. Then in Section 3, we review the concept of MCID and introduce that of the classification error. Our methodology, including both simple linear and non-parametric kernel MCIDs and the bootstrap scheme to compute the confidence interval, is presented in Section 4. We show the finite sample performance of our proposed method through simulation studies in Section 5 and apply our method to the ChAMP trial study in Section 6. The mathematical details are contained in Appendix.

## 2. Motivating example: ChAMP trial

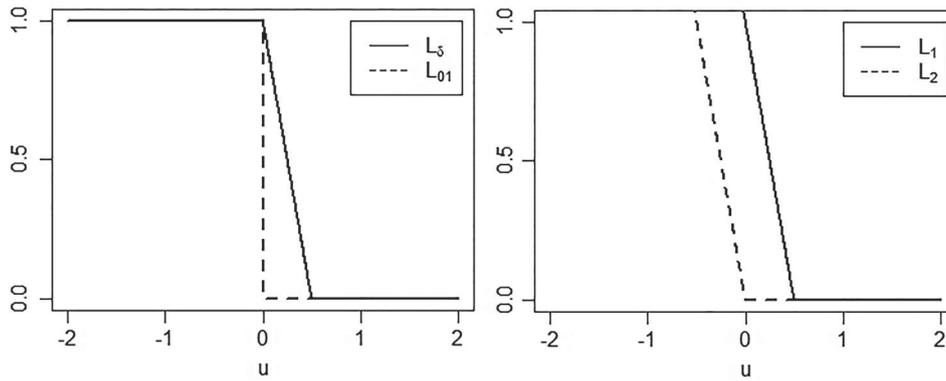
Our motivating example is the chondral lesions and meniscus procedures (ChAMP) trial that examines whether the presence of the chondral lesions surrounding the knee cartilage affects patients' recovery from the arthroscopic partial meniscectomy (APM) (Bisson et al., 2017). In the field of orthopaedics, APM is one of the most common treatment options to repair the knee damage especially for the patients with a meniscus tear. During the operations, however, the surgeons can often find the additional knee damages in the form of chondral lesions. The effect of these chondral lesions on patients' post-operative outcomes are unclear, and whether these lesions need to be treated by debridement remains an open question. Thus, the ChAMP trial is designed to help the clinical physicians better understand this relation and provide them with reasonable suggestions on preoperative evaluation and treatment option.

This study enrolled eligible patients who were  $\geq 30$  years old, diagnosed with a symptomatically consistent meniscus tear by magnetic resonance imaging, and underwent APM. Of the subjects who enrolled, 190 patients with surgically significant chondral lesions were randomised to receive debridement (CL-Deb group;  $n=98$ ) or observation (CL-noDeb group;  $n=92$ ). Outcome measures include the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) and the SF-36 health survey. Each outcome was evaluated at baseline and one-year post-operatively. The demographic data such as age and sex at baseline and surgical data including the location and type of meniscal tears were also collected.

The major goal of the study is to assess whether and how the debridement group is different from the observation group in terms of the change of the WOMAC pain score from enrolment to one year after surgery, and how it relates to other covariate variables and clinical biomarkers. In our investigation, we focus on the study of the interval estimation of the MCID and its classification error. We use an anchor based method to compute the MCID and the anchor question we use is in the SF-36 health survey.

## 3. The MCID and the classification error

In the ChAMP trial, we denote each patient's reported outcome in the SF-36 health survey as a binary variable, where  $Y = 1$  if the patient reports a better health condition after the surgery and  $Y = -1$  otherwise. The difference of each patient's WOMAC pain score from baseline to one year after surgery, denoted as  $X$ , is treated as the patient's diagnostic measurement. Let a  $p$ -dimensional covariate  $\mathbf{Z}$  be the patient's clinical profile with  $\mathbf{Z} \in \mathcal{R}^p$ .



**Figure 1.** The illustration of the 0–1 loss  $L_{01}$ , its surrogate non-smooth ramp loss  $L_\delta$ , and the DC decomposition  $L_\delta = L_1 - L_2$ .

It is reasonable and of interest to consider the MCID  $c^*$  as a function of the patient’s clinical profile,  $c^*(\mathbf{z})$ . The population heterogeneity could also be learned from the knowledge of  $c^*(\mathbf{z})$ . According to Hedayat et al. (2015), the  $c^*(\mathbf{z})$  is defined as the minimiser of

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

where  $E$  is the expectation taken with respect to  $(X, Y, \mathbf{Z})$  and  $\text{sign}(\cdot)$  is the standard sign function. Given independent and identically distributed observations  $\{(x_i, y_i, \mathbf{z}_i), i = 1, \dots, n\}$ , the empirical version of the objective function in (1) becomes

$$\frac{1}{2n} \sum_{i=1}^n [1 - y_i \text{sign}\{x_i - c(\mathbf{z}_i)\}], \quad (2)$$

which involves the 0–1 loss function  $L_{01}(u) = \frac{1}{2}(1 - \text{sign}(u))$ . The direct minimisation of (2) is infeasible. In this paper, we follow Hedayat et al. (2015) to approximate the  $L_{01}$  function with the non-smooth ramp loss function. Note that the non-smooth ramp loss is defined as

$$L_\delta(u) = \begin{cases} 1 & u \leq 0, \\ 1 - \frac{u}{\delta} & 0 < u \leq \delta, \\ 0 & u > \delta, \end{cases}$$

where  $\delta > 0$  is a scalar factor. As  $\delta \rightarrow 0$ ,  $L_\delta(\cdot) \rightarrow L_{01}(\cdot)$ . Then our objective function becomes

$$\frac{1}{n} \sum_{i=1}^n L_\delta\{y_i(x_i - c(\mathbf{z}_i))\}. \quad (3)$$

Due to the non-convexity of the non-smooth ramp loss, the optimisation problem in (3) requires non-convex minimisation. Note that we can write  $L_\delta(u) = L_1(u) - L_2(u)$ , where both  $L_1(u) = (1/\delta)(\delta - u)_+$  and  $L_2(u) = (1/\delta)(-u)_+$  are convex functions. Hence, we apply the difference of convex (DC) algorithm (ThiHoaiAn & DinhTao, 1997) to minimise (3), which

has the form

$$\frac{1}{n} \sum_{i=1}^n L_1\{y_i(x_i - c(\mathbf{z}_i))\} - \frac{1}{n} \sum_{i=1}^n L_2\{y_i(x_i - c(\mathbf{z}_i))\}. \quad (4)$$

Figure 1 illustrates the relation among the  $L_{01}$ ,  $L_\delta$ ,  $L_1$  and  $L_2$  loss functions.

Once the minimiser  $\hat{c}$  is obtained, we can validate whether the debridement or the observation of the chondral lesions for each patient is indeed prescribed correctly. This is essential since it provides knowledgeable advice in future surgical practice for new patients. To re-evaluate whether the treatment is offered appropriately, we need a statistical measure to quantify the discrepancy between the patient’s PRO  $Y_0$  and its dichotomous diagnostic measure from learning his/her MCID  $\text{sign}\{X_0 - c(\mathbf{Z}_0)\}$ . Here, to avoid confusion, we use a generic notation  $\{(Y_0, X_0, \mathbf{Z}_0)\}$  to represent any patient that would like to be validated. This results in

$$E_0 \left[ \mathbb{1} \{Y_0 \neq \text{sign}\{X_0 - \hat{c}(\mathbf{Z}_0)\}\} \right], \quad (5)$$

where  $E_0$  is the expectation taken with respect to  $\{(Y_0, X_0, \mathbf{Z}_0)\}$ . While there are other alternative measures to study, this error (5), usually called the classification error, or the test error, is popularly used in the statistical machine learning literature.

The estimation for the classification error is not a trivial task (Lager & Murphy, 2011). In this paper, we concentrate on creating confidence intervals for both the MCID and the classification error.

## 4. Methodology

In this section, we first present detailed algorithms to calculating MCID. We consider both simple linear MCID and its nonparametric kernel counterpart. We then introduce our bootstrap scheme to construct confidence intervals for MCID and the classification error.

#### 4.1. Algorithms for MCID

It is of particular interest to clinicians if the MCID has a comprehensible structure. For instance,  $c(\mathbf{Z}) = \alpha + \boldsymbol{\beta}^T \mathbf{Z}$ , where  $\mathbf{Z}$  could include the treatment variable, some of the demographic variables and clinical biomarkers. This is the simple linear MCID we consider below. On the other hand, although easily interpretable, a linear structure suffers from the model misspecification issue, thereby yields a solution which may not achieve optimal performance. Therefore, we also consider the nonparametric kernel MCID adopting the reproducing kernel Hilbert space framework.

##### 4.1.1. A simple linear MCID

We assume  $c(\mathbf{Z}) = \alpha + \boldsymbol{\beta}^T \mathbf{Z}$ . We add a penalty term  $(\lambda/2)\boldsymbol{\beta}^T \boldsymbol{\beta}$  in (4) to avoid model overfitting. Let  $\boldsymbol{\omega} = (\alpha, \boldsymbol{\beta}^T)^T$ , then the objective function in (4) is  $s(\boldsymbol{\omega}) = s_1(\boldsymbol{\omega}) - s_2(\boldsymbol{\omega})$ , where

$$s_1(\boldsymbol{\omega}) = \frac{1}{n} \sum_{i=1}^n \left[ \frac{1}{\delta} \{\delta - y_i(x_i - \alpha - \boldsymbol{\beta}^T \mathbf{z}_i)\}_+ \right] + \frac{\lambda}{2} \boldsymbol{\beta}^T \boldsymbol{\beta},$$

$$s_2(\boldsymbol{\omega}) = \frac{1}{n} \sum_{i=1}^n \left[ \frac{1}{\delta} \{-y_i(x_i - \alpha - \boldsymbol{\beta}^T \mathbf{z}_i)\}_+ \right].$$

It is an iterative algorithm to minimise (4). Let  $\widehat{\boldsymbol{\omega}}^{(k)}$  be the estimator of  $\boldsymbol{\omega}$  at the  $k$ th iteration. We first approximate  $s_2(\boldsymbol{\omega})$  with its affine minorisation function  $s_2(\widehat{\boldsymbol{\omega}}^{(k)}) + \langle \boldsymbol{\omega} - \widehat{\boldsymbol{\omega}}^{(k)}, \nabla s_2(\widehat{\boldsymbol{\omega}}^{(k)}) \rangle$ , where  $\nabla s_2(\widehat{\boldsymbol{\omega}}^{(k)})$  is the subgradient of  $s_2(\boldsymbol{\omega})$  at  $\widehat{\boldsymbol{\omega}}^{(k)}$ ,

$$\nabla s_2(\widehat{\boldsymbol{\omega}}^{(k)}) = \begin{pmatrix} \frac{1}{n\delta} \sum_{i=1}^n y_i \mathbb{1}\{y_i(x_i - \alpha^{(k)} - \boldsymbol{\beta}^{(k)T} \mathbf{z}_i) < 0\} \\ \frac{1}{n\delta} \sum_{i=1}^n y_i \mathbf{z}_i \mathbb{1}\{y_i(x_i - \alpha^{(k)} - \boldsymbol{\beta}^{(k)T} \mathbf{z}_i) < 0\} \end{pmatrix}.$$

Hence

$$\begin{aligned} \widehat{\boldsymbol{\omega}}^{(k+1)} &= \arg \min_{\boldsymbol{\omega}} s_1(\boldsymbol{\omega}) - \boldsymbol{\omega}^T \nabla s_2(\widehat{\boldsymbol{\omega}}^{(k)}), \\ &= \arg \min_{\boldsymbol{\omega}} s_1(\boldsymbol{\omega}) \\ &\quad - \frac{1}{n\delta} \sum_{i=1}^n y_i (\alpha + \boldsymbol{\beta}^T \mathbf{z}_i) \mathbb{1}\{y_i(x_i - \alpha^{(k)} - \boldsymbol{\beta}^{(k)T} \mathbf{z}_i) < 0\}. \end{aligned} \quad (6)$$

To solve (6), we derive its dual problem by using the slack variable technique. The details are retained in Appendix 1. It shows that, we arrive at the dual problem

$$\min_{\boldsymbol{\tau}} \boldsymbol{\tau}^T \mathbf{Q} \boldsymbol{\tau} - \{\mathbf{b} + 2\mathbf{Q} \mathbf{t}_1(\alpha^{(k)}, \boldsymbol{\beta}^{(k)})\}^T \boldsymbol{\tau}, \quad (7)$$

subject to  $0 \leq \tau_i \leq 1$  and  $\sum_{i=1}^n (1/\delta) y_i \{\tau_i - t_{1,i}(\alpha^{(k)}, \boldsymbol{\beta}^{(k)})\} = 0$ . This optimisation problem only has simple box constraints hence can be solved by any quadratic programming method.

We conclude the algorithm by presenting how  $\widehat{\alpha}$  and  $\widehat{\boldsymbol{\beta}}$  can be computed. The Karush–Kuhn–Tucker (KKT) conditions associated with optimisation problem (7) are, respectively,

$$\frac{1}{n\lambda\delta} \sum_{i=1}^n y_i \{t_{1,i}(\alpha^{(k)}, \boldsymbol{\beta}^{(k)}) - \tau_i\} \mathbf{z}_i = \boldsymbol{\beta}, \quad (8)$$

$$1 \geq \tau_i \geq 0, \quad (9)$$

$$\tau_i \left\{ \xi_i - 1 + \frac{1}{\delta} y_i (x_i - \alpha - \boldsymbol{\beta}^T \mathbf{z}_i) \right\} = 0, \quad (10)$$

$$\xi_i \geq 0, \quad (11)$$

$$\xi_i - 1 + \frac{1}{\delta} y_i (x_i - \alpha - \boldsymbol{\beta}^T \mathbf{z}_i) \geq 0, \quad (12)$$

$$(1 - \tau_i) \xi_i = 0. \quad (13)$$

Therefore, we have three scenarios to discuss depending on the magnitude of  $\tau_i$ : if  $\tau_i = 0$ , by (12) and (13) we get  $\xi_i = 0$  and  $(1/\delta) y_i (x_i - \alpha - \boldsymbol{\beta}^T \mathbf{z}_i) - 1 \geq 0$ ; if  $1 > \tau_i > 0$ , by (10) and (13) we have  $\xi_i = 0$  and  $(1/\delta) y_i (x_i - \alpha - \boldsymbol{\beta}^T \mathbf{z}_i) - 1 = 0$ ; if  $\tau_i = 1$ , by (10), (12) and (13) we have  $(1/\delta) y_i (x_i - \alpha - \boldsymbol{\beta}^T \mathbf{z}_i) - 1 = -\xi_i \leq 0$ . Therefore, we can summarise the KKT conditions more concisely as

$$\frac{1}{\delta} y_i (x_i - \alpha - \boldsymbol{\beta}^T \mathbf{z}_i) - 1 = -\xi_i \geq 0 \quad \text{if } \tau_i < 1,$$

$$\frac{1}{\delta} y_i (x_i - \alpha - \boldsymbol{\beta}^T \mathbf{z}_i) - 1 = -\xi_i \leq 0 \quad \text{if } \tau_i > 0,$$

which implies

$$\frac{1}{\delta} y_i (x_i - \alpha - \boldsymbol{\beta}^T \mathbf{z}_i) - 1 = 0 \quad \text{if } 0 < \tau_i < 1. \quad (14)$$

Hence through (8) we can estimate  $\boldsymbol{\beta}^{(k+1)}$  as

$$\widehat{\boldsymbol{\beta}}^{(k+1)} = \frac{1}{n\lambda\delta} \sum_{i=1}^n y_i \{t_{1,i}(\alpha^{(k)}, \boldsymbol{\beta}^{(k)}) - \widehat{\tau}_i\} \mathbf{z}_i,$$

and through (14) we can estimate  $\alpha^{(k+1)}$  as

$$\widehat{\alpha}^{(k+1)} = \frac{1}{|\{i : 0 < \tau_i < 1\}|} \sum_{i: 0 < \tau_i < 1} \left( x_i - \frac{\delta}{y_i} - \widehat{\boldsymbol{\beta}}^{(k+1)T} \mathbf{z}_i \right).$$

Finally, we achieve the estimators  $\widehat{\alpha}$  and  $\widehat{\boldsymbol{\beta}}$  after iterative process (6) converges. For any new patient with data  $\mathbf{z}_{\text{new}}$ , his/her predicted linear MCID is

$$\widehat{c}(\mathbf{z}_{\text{new}}) = \widehat{\alpha} + \widehat{\boldsymbol{\beta}}^T \mathbf{z}_{\text{new}}.$$

##### 4.1.2. A nonparametric kernel MCID

Define a feature vector  $\boldsymbol{\phi}_i = \phi(\mathbf{z}_i)$  for the profiles of the  $i$ th patient in the enlarged feature space. We can specify a continuous, symmetric and positive-semidefinite

kernel function  $K$  corresponding to the inner product in the mapping  $\phi$ , that is,  $K(\mathbf{z}_i, \mathbf{z}_j) = \langle \phi_i, \phi_j \rangle$ . Then, we have  $c(\mathbf{z}) = w + h(\mathbf{z})$  with  $w \in \mathcal{R}$  and  $h(\mathbf{z}) \in \mathcal{H}_K$ , where  $\mathcal{H}_K$  is the reproducing kernel Hilbert space (RKHS) with a kernel function  $K(\cdot, \cdot)$ . The norm in  $\mathcal{H}_K$ , denoted by  $\|\cdot\|_K$ , is induced by the following inner product:

$$\langle f, g \rangle_K = \sum_{i=1}^n \sum_{j=1}^m v_i u_j K(\mathbf{z}_i, \mathbf{z}_j),$$

where  $f(\cdot) = \sum_{i=1}^n v_i K(\cdot, \mathbf{z}_i)$  and  $g(\cdot) = \sum_{j=1}^m u_j K(\cdot, \mathbf{z}_j)$ .

Following the representation theorem (Kimeldorf & Wahba, 1971), the nonparametric kernel MCID can be expressed as  $c(\mathbf{z}) = w + \sum_{j=1}^n v_j K(\mathbf{z}, \mathbf{z}_j)$ . Let  $\boldsymbol{\eta} = (w, \mathbf{v}^T)^T = (w, v_1, \dots, v_n)^T$ , then the objective function in (4) is  $h(\boldsymbol{\eta}) = h_1(\boldsymbol{\eta}) - h_2(\boldsymbol{\eta})$ , where

$$h_1(\boldsymbol{\eta}) = \frac{1}{n} \sum_{i=1}^n \left[ \frac{1}{\delta} \left\{ \delta - y_i (x_i - w - \sum_{j=1}^n v_j K(\mathbf{z}_i, \mathbf{z}_j)) \right\} \right]_+ + \frac{\lambda}{2} \sum_{i,j=1}^n v_i v_j K(\mathbf{z}_i, \mathbf{z}_j),$$

$$h_2(\boldsymbol{\eta}) = \frac{1}{n} \sum_{i=1}^n \left[ \frac{1}{\delta} \left\{ -y_i (x_i - w - \sum_{j=1}^n v_j K(\mathbf{z}_i, \mathbf{z}_j)) \right\} \right]_+.$$

Similar to the linear case, it is an iterative algorithm to minimise (4). Let  $\widehat{\boldsymbol{\eta}}^{(k)}$  be the estimator of  $\boldsymbol{\eta}$  at the  $k^{\text{th}}$  iteration. We first approximate  $h_2(\boldsymbol{\eta})$  with its affine minorisation function  $h_2(\widehat{\boldsymbol{\eta}}^{(k)}) + \langle \boldsymbol{\eta} - \widehat{\boldsymbol{\eta}}^{(k)}, \nabla h_2(\widehat{\boldsymbol{\eta}}^{(k)}) \rangle$ , where  $\nabla h_2(\widehat{\boldsymbol{\eta}}^{(k)})$  is the subgradient of  $h_2(\boldsymbol{\eta})$  at  $\widehat{\boldsymbol{\eta}}^{(k)}$ ,

$$\nabla h_2(\widehat{\boldsymbol{\eta}}^{(k)}) = \begin{pmatrix} \frac{1}{n\delta} \sum_{i=1}^n y_i \mathbb{1} \left\{ y_i (x_i - w^{(k)} - \sum_{j=1}^n v_j^{(k)} K(\mathbf{z}_i, \mathbf{z}_j)) < 0 \right\} \\ \frac{1}{n\delta} \sum_{i=1}^n y_i K(\mathbf{z}_i, \mathbf{z}_1) \mathbb{1} \left\{ y_i (x_i - w^{(k)} - \sum_{j=1}^n v_j^{(k)} K(\mathbf{z}_i, \mathbf{z}_j)) < 0 \right\} \\ \vdots \\ \frac{1}{n\delta} \sum_{i=1}^n y_i K(\mathbf{z}_i, \mathbf{z}_n) \mathbb{1} \left\{ y_i (x_i - w^{(k)} - \sum_{j=1}^n v_j^{(k)} K(\mathbf{z}_i, \mathbf{z}_j)) < 0 \right\} \end{pmatrix}.$$

Consequently,

$$\begin{aligned} \widehat{\boldsymbol{\eta}}^{(k+1)} &= \arg \min_{\boldsymbol{\eta}} h_1(\boldsymbol{\eta}) - \boldsymbol{\eta}^T \nabla h_2(\widehat{\boldsymbol{\eta}}^{(k)}), \\ &= \arg \min_{\boldsymbol{\eta}} h_1(\boldsymbol{\eta}) - \frac{1}{n\delta} \sum_{i=1}^n y_i (w \\ &\quad + \sum_{j=1}^n v_j K(\mathbf{z}_i, \mathbf{z}_j)) \mathbb{1} \left\{ y_i (x_i - w^{(k)} \right. \\ &\quad \left. - \sum_{j=1}^n v_j^{(k)} K(\mathbf{z}_i, \mathbf{z}_j)) < 0 \right\}. \end{aligned} \quad (15)$$

Similar to the linear case, we use the slack variable technique and reach the dual problem

$$\min_{\boldsymbol{\tau}'} \boldsymbol{\tau}'^T \mathbf{Q}' \boldsymbol{\tau}' - \{\mathbf{d} + 2\mathbf{Q}' \mathbf{t}_2(w^{(k)}, \mathbf{v}^{(k)})\}^T \boldsymbol{\tau}', \quad (16)$$

subject to  $0 \leq \tau'_i \leq 1$  and  $\sum_{i=1}^n (1/\delta) y_i \{\tau'_i - t_{2,i}(w^{(k)}, \mathbf{v}^{(k)})\} = 0$ . This optimisation problem only have simple box constraints hence can be solved by any quadratic programming method.

We conclude the algorithm by presenting how  $\widehat{w}$  and  $\widehat{\mathbf{v}}$  can be computed. The Karush–Kuhn–Tucker (KKT) conditions associated with optimisation problem (16) are, respectively,

$$\frac{1}{n\lambda\delta} y_i \{t_{2,i}(w^{(k)}, \mathbf{v}^{(k)}) - \tau'_i\} \phi_i = v_i, \quad (17)$$

$$1 \geq \tau'_i \geq 0, \quad (18)$$

$$\tau'_i \left\{ \xi'_i - 1 + \frac{1}{\delta} y_i \left( x_i - w - \sum_{j=1}^n v_j K(\mathbf{z}_i, \mathbf{z}_j) \right) \right\} = 0, \quad (19)$$

$$\xi'_i \geq 0, \quad (20)$$

$$\xi'_i - 1 + \frac{1}{\delta} y_i \left( x_i - w - \sum_{j=1}^n v_j K(\mathbf{z}_i, \mathbf{z}_j) \right) \geq 0, \quad (21)$$

$$(1 - \tau'_i) \xi'_i = 0. \quad (22)$$

Three scenarios can be consequently discussed based on the magnitude of  $\tau'_i$ : if  $\tau'_i = 0$ , by (21) and (22) we have  $\xi'_i = 0$  and  $(1/\delta) y_i (x_i - w - \sum_{j=1}^n v_j K(\mathbf{z}_i, \mathbf{z}_j)) - 1 \geq 0$ ; if  $1 > \tau'_i > 0$ , by (19) and (22) we obtain  $\xi'_i = 0$  and  $(1/\delta) y_i (x_i - w - \sum_{j=1}^n v_j K(\mathbf{z}_i, \mathbf{z}_j)) - 1 = 0$ ; if  $\tau'_i = 1$ , by (19), (21) and (22) we obtain  $(1/\delta) y_i (x_i - w - \sum_{j=1}^n v_j K(\mathbf{z}_i, \mathbf{z}_j)) - 1 = -\xi'_i \leq 0$ . Now we can

summarise the KKT conditions more concisely as

$$\begin{aligned} \frac{1}{\delta} y_i \left( x_i - w - \sum_{j=1}^n v_j K(\mathbf{z}_i, \mathbf{z}_j) \right) - 1 = -\xi'_i &\geq 0 \\ \text{if } \tau'_i < 1, \\ \frac{1}{\delta} y_i \left( x_i - w - \sum_{j=1}^n v_j K(\mathbf{z}_i, \mathbf{z}_j) \right) - 1 = -\xi'_i &\leq 0 \\ \text{if } \tau'_i > 0, \end{aligned}$$

which implies

$$\frac{1}{\delta} y_i \left( x_i - w - \sum_{j=1}^n v_j K(\mathbf{z}_i, \mathbf{z}_j) \right) - 1 = 0 \text{ if } 0 < \tau'_i < 1. \quad (23)$$

Thus for  $i = 1, \dots, n$ ,  $v_i^{(k+1)}$  can be estimated via (17) as

$$\hat{v}_i^{(k+1)} = \frac{1}{n\lambda\delta} y_i \{t_{2,i}(w^{(k)}, \mathbf{v}^{(k)}) - \hat{\tau}'_i\} \phi_i$$

and  $w^{(k+1)}$  can be estimated via (23) as

$$\hat{w}^{(k+1)} = \frac{1}{|\{i : 0 < \tau'_i < 1\}|} \sum_{i:0 < \tau'_i < 1} \left( x_i - \frac{\delta}{y_i} - \sum_{j=1}^n \hat{v}_j K(\mathbf{z}_i, \mathbf{z}_j) \right).$$

We can get the estimators  $\hat{w}$  and  $\hat{\mathbf{v}}$  after iterative process (15) converges. As a result, for any new patient with data  $\mathbf{z}_{\text{new}}$ , his/her predicted nonparametric kernel MCID is  $\hat{c}(\mathbf{z}_{\text{new}}) = \hat{w} + \sum_{j=1}^n \hat{v}_j K(\mathbf{z}_{\text{new}}, \mathbf{z}_j)$ . Note that the most common use of the nonparametric kernel function is the Gaussian radial basis function  $K(\mathbf{z}_1, \mathbf{z}_2) = \exp(-\|\mathbf{z}_1 - \mathbf{z}_2\|^2/2\sigma^2)$ , where  $\sigma$  is a positive scale parameter. If  $K(\mathbf{z}_1, \mathbf{z}_2) = \mathbf{z}_1^T \mathbf{z}_2$ , the nonparametric kernel case will reduce to the simple linear case.

#### 4.2. Bootstrap procedure for MCID and the classification error

A point estimation alone does not allow one to quantify its uncertainty hence limits the usefulness of the MCID and its classification error in real applications. Hedayat et al. (2015) postulates a testing data set with a very large sample size to quantify the effectiveness of their MCID in numerical studies, but such a testing data set is usually infeasible in clinical studies. Resampling method with the bootstrap as a representative, on the other hand, could serve as a tool to construct a confidence interval for an estimand under these situations.

To appropriately use bootstrap, we have to be cautious on the regularity conditions under which the theoretical properties can be justified (Bickel et al., 1997). If these conditions are not satisfied, the conventional

bootstrap has to be properly modified. The objective function to be minimised in (4) and the one in (5) are generally non-smooth functions. If there exists a non-negligible probability concentrated at the discontinuous points of the objective function, that is,  $P(X_0 - \hat{c}(\mathbf{Z}_0) = 0) > 0$ , it is called the irregular case under which Shao (1994) showed that the conventional bootstrap is inconsistent. Instead, we propose to adopt the  $m$ -out-of- $n$  bootstrap, a general method for remedying bootstrap inconsistency due to non-smoothness, and theoretically justified in Bickel et al. (1997), Shao (1994, 1996) and references therein.

The  $m$ -out-of- $n$  bootstrap is the conventional non-parametric bootstrap except that the resample size, historically denoted as  $m$ , is of a smaller order compared to the original sample size  $n$ . That is,  $m = m_n \rightarrow \infty$  and  $m/n \rightarrow 0$  (or  $m \log \log n/n \rightarrow 0$ ) as  $n \rightarrow \infty$  (Shao, 1994). The intuition of the  $m$ -out-of- $n$  bootstrap is to let the empirical distribution tend to the true generative distribution at a faster rate, and essentially this allows the empirical distribution to reach its limit faster hence the bootstrap samples are drawn as if they were from the true generative distribution. Intuitively the requirement of  $m/n \rightarrow 0$  (or  $m \log \log n/n \rightarrow 0$ ) is consistent with Hedayat et al. (2015) who required a very large sample size for their testing data. To some extent, in the  $m$ -out-of- $n$  bootstrap, those  $n-m$  subjects serve the role of the testing sample (Shao, 1996). Practically  $m$  is usually chosen as  $m = n^\kappa$  for some  $\kappa < 1$ . In our numerical studies, we choose  $\kappa = 0.9$ . Our algorithm is detailed below.

For  $b = 1, \dots, B$ , we generate bootstrap samples with size  $m$  as  $\{(x_j^{(b)}, y_j^{(b)}, \mathbf{z}_j^{(b)}), j = 1, \dots, m\}$ . The MCID can be derived as  $\hat{c}^{(b)}$  based on the method in Section 4. To be more specific, the simple linear MCID is  $\hat{c}_l^{(b)} = \hat{\alpha}^{(b)} + \hat{\beta}^{(b)T} \mathbf{z}_0$  and the nonparametric kernel MCID is  $\hat{c}_n^{(b)} = \hat{w}^{(b)} + \sum_{i=1}^n \hat{v}_i^{(b)} K(\mathbf{z}_0, \mathbf{z}_i^{(b)})$ , where  $\mathbf{z}_0$  denotes the profile for a new patient. Accordingly, the classification error based on MCID  $\hat{c}^{(b)}$  is computed as

$$\widehat{\text{err}}^{(b)} = \frac{1}{m} \sum_{j=1}^m \mathbb{1}\{y_j^{(b)} \neq \text{sign}(x_j^{(b)} - \hat{c}(\mathbf{z}_j^{(b)}))\}.$$

Similarly,  $\widehat{\text{err}}^{(b)}$  can also be distinguished as  $\widehat{\text{err}}_l^{(b)}$  and  $\widehat{\text{err}}_n^{(b)}$  for simple linear and nonparametric kernel cases.

We repeat the above procedure in total  $B$  times. If we approach the simple linear MCID, we obtain  $\{(\hat{\alpha}^{(b)}, \hat{\beta}^{(b)}, \hat{c}_l^{(b)}, \widehat{\text{err}}_l^{(b)})\}, b = 1, \dots, B$ . Let  $\hat{l}_\alpha$  and  $\hat{u}_\alpha$  be the  $\alpha/2$ th and  $(1 - \alpha/2)$ th quantiles of  $\{\hat{\alpha}^{(b)}, b = 1, \dots, B\}$ ,  $\hat{l}_\beta$  and  $\hat{u}_\beta$  be the  $\alpha/2$ th and  $(1 - \alpha/2)$ th quantiles of  $\{\hat{\beta}^{(b)}, b = 1, \dots, B\}$ ,  $\hat{l}_{c_l}$  and  $\hat{u}_{c_l}$  be the  $\alpha/2$ th and  $(1 - \alpha/2)$ -th quantiles of  $\{\hat{c}_l^{(b)}, b = 1, \dots, B\}$ , and  $\hat{l}_{\text{err}_l}$  and  $\hat{u}_{\text{err}_l}$  be the  $\alpha/2$ th and  $(1 - \alpha/2)$ th quantiles of  $\{\widehat{\text{err}}_l^{(b)}, b = 1, \dots, B\}$ . Thus, the  $(1 - \alpha)$ th confidence

interval of  $\alpha$  is given by  $[\widehat{l}_\alpha, \widehat{u}_\alpha]$ , the  $(1 - \alpha)$ th confidence interval of  $\beta$  is given by  $[\widehat{l}_\beta, \widehat{u}_\beta]$ ,  $(1 - \alpha)$ th confidence interval of  $c_i$  is given by  $[\widehat{l}_{c_i}, \widehat{u}_{c_i}]$ , and  $(1 - \alpha)$ th confidence interval of  $\text{err}$  is given by  $[\widehat{l}_{\text{err}_i}, \widehat{u}_{\text{err}_i}]$ . If we approach the nonparametric kernel MCID, we have  $\{(\widehat{c}_n^{(b)}, \widehat{\text{err}}_n^{(b)})\}, b = 1, \dots, B$ . Similarly the  $(1 - \alpha)$ th confidence interval of  $c_n$  is given by  $[\widehat{l}_{c_n}, \widehat{u}_{c_n}]$ , and  $(1 - \alpha)$ th confidence interval of  $\text{err}$  is given by  $[\widehat{l}_{\text{err}_n}, \widehat{u}_{\text{err}_n}]$ .

### 5. Simulation studies

In this section, we apply the proposed method to provide confidence intervals for the MCID and the classification error via extensive numerical studies based on simulated data.

We consider two scenarios. In the first scenario, we generate a random sample consisting of independent and identically distributed observations  $\{(X_i, Y_i, \mathbf{Z}_i), i = 1, \dots, n\}$ , where we first generate patient's clinical profile  $\mathbf{Z}_i$  from a bivariate normal distribution  $N_2(\boldsymbol{\mu}, \mathbf{I}_2)$ , where  $\boldsymbol{\mu} = (0, 0)^T$  and  $\mathbf{I}_2 = \text{diag}(1, 1)$ . Then, we generate  $X_i$  from  $N(\alpha + \boldsymbol{\beta}^T \mathbf{z}_i, 1)$ , where  $\alpha = 0$  and  $\boldsymbol{\beta} = (1, 2)^T$ . Finally, we generate the binary patient reported outcome  $Y_i \in \{-1, 1\}$  from  $\text{Bern}(F(x_i))$ , where  $F(x_i) = P(X_i \leq x_i)$ . Note that under this scenario, the linear MCID is the underlying truth. We also generate a new observation  $(Y_{\text{new}}, X_{\text{new}}, \mathbf{Z}_{\text{new}})$  with  $Y_{\text{new}} = 1, X_{\text{new}} = -0.3376$  and  $\mathbf{Z}_{\text{new}} = (-1.3577, -1.3643)$  from the same distribution as all  $\{(Y_i, X_i, \mathbf{Z}_i)\}$ 's. The true value of the MCID for this patient is  $-4.0862$ .

The data under the second scenario are generated similarly to the first, except that the  $X_i$  is generated from  $N(\alpha + \boldsymbol{\beta}^T \mathbf{z}_i - \boldsymbol{\beta}^T \mathbf{z}_i^2, 1)$ . Hence the linear structure for the MCID is misspecified in this setting. Similar to the first scenario, we also generate a new observation  $(Y_{\text{new}}, X_{\text{new}}, \mathbf{Z}_{\text{new}})$  with  $Y_{\text{new}} = 1, X_{\text{new}} = -2.1809$  and  $\mathbf{Z}_{\text{new}} = (-1.3577, -1.3643)$ . The true value of the MCID for this new patient is  $-9.6519$ .

In each of the two scenarios, we apply both simple linear and nonparametric kernel MCID methods. The nonparametric kernel we apply is the Gaussian kernel defined as  $K(\mathbf{z}_1, \mathbf{z}_2) = \exp(-\|\mathbf{z}_1 - \mathbf{z}_2\|^2/2\sigma^2)$ . The scale parameter can be found by setting it as the median of pairwise Euclidean distances within the observations  $(\mathbf{z}_1, \mathbf{z}_2)$  used to estimate the prediction rule (Hedayat et al., 2015). The  $m$ -out-of- $n$  bootstrap samples are generated 1000 times for each case. For simplicity we set  $\delta = 0.01$  in our numerical studies and we use the  $m$ -fold cross validation method to determine the tuning parameter  $\lambda$ . We report two different sample sizes:  $n = 500$  and  $n = 1000$  for each situation.

Based on 500 simulation replications, our results are summarised in Tables 1 and 2. It can be seen that, in scenario 1, the length of the confidence interval is much shorter and the coverage is much more accurate when

**Table 1.** Confidence interval for MCID in simulation studies.

Sample size	Kernel	MCID			
		Lower	Upper	Length	CP
Scenario 1					
$n = 500$	Linear (Correct)	-4.3040	-3.3198	0.9842	0.934
	Kernel	-4.6326	-2.3341	2.2985	0.928
$n = 1000$	Linear (Correct)	-4.2669	-3.6130	0.6539	0.952
	Kernel	-4.5890	-1.5640	3.0250	0.982
Scenario 2					
$n = 500$	Linear (Incorrect)	-7.5186	-4.1525	3.3661	0.004
	Kernel	-10.7414	-7.4439	3.2974	0.990
$n = 1000$	Linear (Incorrect)	-7.5371	-4.8906	2.6465	0.000
	Kernel	-10.5402	-4.7963	5.7439	0.998

Note: CP: coverage probability.

**Table 2.** Confidence interval for the classification error in simulation studies.

Sample Size	Kernel	Classification Error			
		Lower	Upper	Length	CP
Scenario 1					
$n = 500$	Linear (Correct)	0.1951	0.3068	0.1117	1.000
	Kernel	0.1985	0.3198	0.1213	0.996
$n = 1000$	Linear (Correct)	0.2080	0.2871	0.0791	0.992
	Kernel	0.2063	0.3428	0.1365	1.000
Scenario 2					
$n = 500$	Linear (Incorrect)	0.3595	0.4861	0.1266	0.000
	Kernel	0.2041	0.3335	0.1294	0.996
$n = 1000$	Linear (Incorrect)	0.3762	0.4687	0.0924	0.000
	Kernel	0.2086	0.3883	0.1797	1.000

Note: CP: coverage probability.

the correct linear structure and a larger sample size are used. The length is much larger and the coverage is much broader when the nonparametric kernel function is used. In scenario 2, the coverage using the nonparametric kernel is much broader. More importantly we find that the estimation of the MCID using the incorrect simple linear structure is biased so that it gives very poor coverage. This issue cannot be uncovered if a confidence interval is not available as in Hedayat et al. (2015). It also reinforces the necessity and importance of developing interval estimation for the MCID.

We notice that the coverage for the classification error is very broad and Xu et al. (2015) also noted the similar phenomenon. Due to the computational burden, we only explore our method up to the sample size of 1000. Under the situation with a much larger sample size, the coverage would become more accurate.

### 6. ChAMP trial analysis

In the ChAMP trial study, 190 patients with chondral lesions undergoing APM are randomised to either treatment (debridement) group ( $n = 98$ ) or control (no debridement, observation) group ( $n = 92$ ). It is of interest to investigate whether the debridement treatment on the chondral lesions would encourage the recovery from the surgery of repairing knee damage.

**Table 3.** Interval estimation for the ChAMP trial analysis.

	MCID				Classification error		
	Point Estimation	Interval estimation			Interval estimation		
		Lower	Upper	Length	Lower	Upper	Length
No model: $c(\mathbf{z}) = c$							
MCID_all	1.4199	-0.9868	2.3224	3.3092	0.3050	0.4421	0.1371
MCID_trt = 1	2.0216	-0.9944	3.8266	4.8210			
MCID_trt = 0	-0.0843	-0.9868	1.7207	2.7076			
MCID_diff	2.1059	-1.5042	3.9109	5.4151			
Model 1: $c(\mathbf{z}) = \alpha + \beta_1$ treatment							
$\alpha$	-0.0009	-2.3885	1.6182	4.0067	0.2842	0.5263	0.2421
$\beta_1$	1.9008	-1.4109	3.4050	4.8159			
MCID_all	0.9676	-1.0497	1.8041	2.8538			
MCID_trt = 1	1.8999	-0.8843	2.5207	3.4050			
MCID_trt = 0	-0.0009	-2.3885	1.6182	4.0067			
MCID_diff	1.9008	-1.4109	3.4050	4.8159			
Model 2: $c(\mathbf{z}) = \alpha + \beta_1$ treatment + $\beta_2$ age + $\beta_3$ sex + $\beta_4$ damage							
$\alpha$	-6.7261	-21.4810	11.3416	32.8226	0.2737	0.5158	0.2421
$\beta_1$	2.4068	-1.9171	3.9081	5.8252			
$\beta_2$	0.1680	-0.1918	0.3768	0.5686			
$\beta_3$	-0.9713	-3.4509	2.2291	5.6800			
$\beta_4$	-1.1430	-1.8663	1.3246	3.1909			
MCID_all	0.2467	-1.6964	2.0215	3.7178			
MCID_trt = 1	1.6050	-1.6592	3.3212	4.9804			
MCID_trt = 0	-1.1645	-3.1225	1.7468	4.8693			
MCID_diff	2.7694	-1.6298	4.0097	5.6395			
Model 3: $c(\mathbf{z})$ is nonparametric							
MCID_all	-0.3465	-1.5073	1.5785	3.0858	0.0526	0.4424	0.3897
MCID_trt = 1	-0.2987	-1.4469	1.8478	3.2947			
MCID_trt = 0	-0.3962	-1.7511	1.5378	3.2890			
MCID_diff	0.0975	-0.5132	1.2298	1.7430			

As we mentioned in the previous section, the binary variable  $Y$  is derived from the anchor question in the SF-36 health survey. The patient's diagnostic measurement  $X$  is the difference in WOMAC pain score between baseline and one year after surgery. The score is scaled from 0 (extreme problem) to 100 (no problem). Patient's clinical profile  $\mathbf{Z}$  includes their age (continuous), treatment assignment (binary), sex (binary) and knee damage (four level categorical). The knee damage variable is the total number of types of meniscus tears that the patient suffers from. It reflects the severity of patient's knee damage.

After excluding the missing data, our analysis contains 157 patients. Among them, 80 patients are assigned in the debridement group and 77 others observation group. In our analysis, we first compute the MCID for the whole population, for the subpopulation within the debridement group (treatment = 1), for the subpopulation within the observation group (treatment = 0), and their difference, respectively, where the MCID  $c(\mathbf{z})$  only includes an intercept term. We call this 'No model' in Table 3. Then based on the structure

$$c(\mathbf{z}) = \alpha + \beta_1 \text{ treatment},$$

which we call 'Model 1', the structure

$$c(\mathbf{z}) = \alpha + \beta_1 \text{ treatment} + \beta_2 \text{ age} + \beta_3 \text{ sex} + \beta_4 \text{ damage},$$

which we call 'Model 2', and the nonparametric  $c(\mathbf{z})$  which we call 'Model 3' in Table 3, separately, we implement our proposed estimation procedure. The results are summarised in Table 3. Note that although we concentrate on interval estimations, we also list the 'point estimation' column mainly for the purpose of comparing with the results in Hedayat et al. (2015). Here we generated the  $m$ -out-of- $n$  bootstrap samples  $B = 1000$  times with  $m = n^{0.9} \approx 95$ . Also, a sensitivity analysis on the value of  $B$  in Table 4 demonstrates that the results are not sensitive when it varies from around 500 to 1500.

From Table 3, the point estimation of MCID for the treatment = 1 subgroup 2.0216 looks quite different from that for the treatment = 0 subgroup -0.0843. Without the interval estimation, one would believe that they are different however with no further evidence on the degree of how they are different. Now with interval estimation, we see the confidence interval for each of them covers zero, and the confidence interval for the difference also covers zero. The similar phenomenon could also be observed under either 'Model 1' or 'Model 2' or 'Model 3'.

Across the first three models, each of the MCID quantities has slightly different estimates, but each of their confidence intervals covers zero. 'Model 3' has difference MCID estimates than its linear counterpart, indicating the potential model interpretability

**Table 4.** Sensitivity analysis for the ChAMP trial.

		Bootstrap resampling size (B)			
		600	900	1200	1500
<b>No model: <math>c(\mathbf{z}) = c</math></b>					
MCID_all	Lower	-0.9868	-0.9868	-0.9868	-0.9868
	Upper	2.3600	2.3224	2.3224	2.3224
MCID_trt = 1	Lower	-0.9868	-0.9868	-1.2877	-1.2877
	Upper	3.8266	3.8266	3.8266	3.8266
MCID_trt = 0	Lower	-0.9868	-0.9868	-0.9868	-0.9868
	Upper	1.7207	1.7207	1.7207	1.7207
MCID_diff	Lower	-1.5042	-1.5042	-1.5042	-1.5042
	Upper	3.9109	3.9109	3.9109	3.9109
Classification Error	Lower	0.3053	0.2997	0.3053	0.2947
	Upper	0.4526	0.4476	0.4526	0.4526
<b>Model 1: <math>c(\mathbf{z}) = \alpha + \beta_1</math> treatment</b>					
$\alpha$	Lower	-2.2125	-2.3885	-2.3885	-2.3885
	Upper	1.6182	1.6182	1.6182	1.6182
$\beta_1$	Lower	-1.2034	-1.3565	-1.4084	-1.2992
	Upper	3.4050	3.4050	3.4050	3.3092
MCID_all	Lower	-0.8760	-1.0214	-1.0473	-1.0458
	Upper	1.7648	1.7624	1.8041	1.9027
MCID_trt = 1	Lower	-0.7909	-0.8843	-0.8843	-0.8843
	Upper	2.5223	2.5207	2.5208	2.5208
MCID_trt = 0	Lower	-2.2125	-2.3885	-2.3885	-2.3885
	Upper	1.6182	1.6182	1.6182	1.6182
MCID_diff	Lower	-1.2034	-1.3565	-1.4084	-1.2992
	Upper	3.4050	3.4050	3.4050	3.3092
Classification Error	Lower	0.2842	0.2842	0.2842	0.2842
	Upper	0.5161	0.5263	0.5263	0.5263
<b>Model 2: <math>c(\mathbf{z}) = \alpha + \beta_1</math> treatment + <math>\beta_2</math> age + <math>\beta_3</math> sex + <math>\beta_4</math> damage</b>					
$\alpha$	Lower	-22.5595	-21.4412	-21.3968	-21.1874
	Upper	9.7443	10.3888	10.7200	11.2746
$\beta_1$	Lower	-1.9235	-1.8822	-1.9171	-1.9289
	Upper	4.0760	3.8862	3.9378	3.9518
$\beta_2$	Lower	-0.1830	-0.1842	-0.1830	-0.2019
	Upper	0.4176	0.3760	0.3751	0.3715
$\beta_3$	Lower	-3.5843	-3.4407	-3.5837	-3.4926
	Upper	2.3425	2.2252	2.2467	2.3068
$\beta_4$	Lower	-1.7888	-1.8370	-1.8925	-1.8103
	Upper	1.3664	1.3070	1.3238	1.3833
MCID_all	Lower	-1.7405	-1.7091	-1.6945	-1.7689
	Upper	2.0215	2.0263	1.9981	2.0001
MCID_trt = 1	Lower	-1.7345	-1.6240	-1.5910	-1.5944
	Upper	3.3077	3.3142	3.3077	3.2833
MCID_trt = 0	Lower	-3.2409	-3.1269	-3.0683	-3.1346
	Upper	1.8563	1.7076	1.6739	1.6121
MCID_diff	Lower	-1.6303	-1.5644	-1.6227	-1.6262
	Upper	4.2119	3.9789	4.0669	4.0933
Classification Error	Lower	0.2734	0.2737	0.2737	0.2737
	Upper	0.5055	0.5158	0.5053	0.5158
<b>Model 3: <math>c(\mathbf{z})</math> is nonparametric</b>					
MCID_all	Lower	-1.4763	-1.4921	-1.5270	-1.5201
	Upper	1.5442	1.5951	1.5785	1.6139
MCID_trt = 1	Lower	-1.4830	-1.4605	-1.5064	-1.4834
	Upper	1.7903	1.8483	1.8195	1.8338
MCID_trt = 0	Lower	-1.7379	-1.7465	-1.7623	-1.7518
	Upper	1.5213	1.5583	1.5378	1.6119
MCID_diff	Lower	-0.4841	-0.5003	-0.5468	-0.5488
	Upper	1.2007	1.2294	1.2000	1.1966
Classification Error	Lower	0.0526	0.0526	0.0526	0.0526
	Upper	0.4526	0.4476	0.4421	0.4421

insufficiency contributed by the linear component of the MCID.

Also for the treatment effect  $\beta_1$  in ‘Model 1’ and ‘Model 2’, although their estimates are different, each of their confidence intervals also covers zero. As we can see, although ‘Model 3’ looks more flexible than its linear counterpart, it lacks clear interpretability of the treatment effect. The different results from ‘No model’

and ‘Model 3’ also indicates that there may also exist some other unobserved covariate variables that may play a role to quantifying the population heterogeneity in terms of MCID.

The classification errors for the first three models are roughly similar with that of ‘Model 1’ slightly greater than ‘Model 2’. This is reasonable since ‘Model 2’ controls more covariates hence should have a greater ability for model interpretability. ‘Model 3’ has smaller lower bound and upper bound of its confidence interval than ‘Model 2’ because it is generally acknowledged that a nonparametric kernel model entails fewer model assumptions than its linear counterpart model hence is regarded as more flexible.

In all, besides the point estimation of the MCID in each population of interest, our proposed method also provides its interval estimation which gives us a better understanding of the scale of MCID hence could facilitate us to compare with some certain historical values or values from other populations or other diseases so that a better, more convincing health policy decision could be made. The ChAMP study shows that the MCID between the debridement group and the observation group has some difference, but that difference is not significant. One of the major findings of the ChAMP trial study is that to debride the chondral lesions does not have a statistically significant effect hence recommends that not to debride the chondral lesions in future surgical practice (Bisson et al., 2017). Using the proposed method for the MCID in this paper, we reach the same conclusion. This could have a big impact on the orthopaedics surgical practice since the additional debridement of chondral lesions would bring a significant medical cost to the patients.

### Acknowledgments

The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

### Disclosure statement

No potential conflict of interest was reported by the authors.

### Funding

This work was supported by the National Center for Advancing Translational Sciences of the National Institutes of Health under award number UL1TR001412.

### Notes on contributors

**Zehua Zhou** is PhD candidate from the Department of Biostatistics at SUNY Buffalo. Before that, he obtained his Bachelor degree in Bioengineering from China Pharmaceutical University in 2014 and his Master degree in Biostatistics from SUNY Buffalo in 2017. He is interested in statistical machine learning methods, missing data analysis, and personalized medicine.

**Jiwei Zhao** is Assistant Professor from the Department of Biostatistics at SUNY Buffalo. He earned his PhD degree in Statistics from the University of Wisconsin-Madison in 2012. He is generally interested in using statistical and machine learning techniques to solve challenging problems in biomedical data science.

**Melissa Kluczynski** is Clinical Research Associate from the Department of Orthopaedics at SUNY Buffalo. She is also affiliated with the UBMD Orthopaedics and Sports Medicine.

## ORCID

Jiwei Zhao  <http://orcid.org/0000-0002-9298-9412>

Melissa Kluczynski  <http://orcid.org/0000-0003-3947-4222>

## References

- Bickel, P., Götze, F., & van Zwet, W. (1997). Resampling fewer than  $n$  observations: Gains, losses, and remedies for losses. *Statistica Sinica*, 7, 1–31.
- Bisson, L. J., Kluczynski, M. A., Wind, W. M., Fineberg, M. S., Bernas, G. A., Rauh, M. A., . . . Zhao, J. (2017). Patient outcomes after observation versus debridement of unstable chondral lesions during partial meniscectomy. *The Journal of Bone and Joint Surgery*, 99, 1078–1085.
- Bisson, L. J., Kluczynski, M. A., Wind, W. M., Fineberg, M. S., Bernas, G. A., Rauh, M. A., . . . Zhao, J. (2018). How does the presence of unstable chondral lesions affect patient outcomes after partial meniscectomy? The ChAMP randomized controlled trial. *The American Journal of Sports Medicine*, 46, 590–597.
- Bisson, L., Phillips, P., Matthews, J., Zhou, Z., Zhao, J., Wind, W., Fineberg, M., Bernas, G., Rauh, M., Marzo, J. and Kluczynski, M. (2019). Association Between Bone Marrow Lesions, Chondral Lesions, and Pain in Patients Without Radiographic Evidence of Degenerative Joint Disease Who Underwent Arthroscopic Partial Meniscectomy. *Orthopaedic Journal of Sports Medicine*, 7(3), 2325967119830381.
- Cook, C. E. (2008). Clinimetrics corner: The minimal clinically important change score (MCID): A necessary pretense. *Journal of Manual & Manipulative Therapy*, 16, 82E–83E.
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7, 1–26.
- Erdogan, B. D., Leung, Y. Y., Pohl, C., Tennant, A., & Conaghan, P. G. (2016). Minimal clinically important difference as applied in rheumatology: An OMERACT Rasch Working Group systematic review and critique. *The Journal of Rheumatology*, 43, 194–202.
- Hedayat, A., Wang, J., & Xu, T. (2015). Minimum clinically important difference in medical studies. *Biometrics*, 71, 33–41.
- Jaeschke, R., Singer, J., & Guyatt, G. H. (1989). Measurement of health status: Ascertaining the minimal clinically important difference. *Controlled Clinical Trials*, 10, 407–415.
- Kimeldorf, G., & Wahba, G. (1971). Some results on Tchebycheffian spline functions. *Journal of Mathematical Analysis and Applications*, 33, 82–95.
- Kluczynski, M. A., Marzo, J. M., Wind, W. M., Fineberg, M. S., Bernas, G. A., M. A. Rauh, . . . Bisson, L. J. (2017). The effect of body mass index on clinical outcomes in patients without radiographic evidence of degenerative joint disease after arthroscopic partial meniscectomy. *Arthroscopy: The Journal of Arthroscopic & Related Surgery*, 33, 2054–2063.
- Laber, E. B., & Murphy, S. A. (2011). Adaptive confidence intervals for the test error in classification. *Journal of the American Statistical Association*, 106, 904–913.
- McGlothlin, A. E., & Lewis, R. J. (2014). Minimal clinically important difference: Defining what really matters to patients. *JAMA*, 312, 1342–1343.
- Shao, J. (1994). Bootstrap sample size in nonregular cases. *Proceedings of the American Mathematical Society*, 122, 1251–1262.
- Shao, J. (1996). Bootstrap model selection. *Journal of the American Statistical Association*, 91, 655–665.
- Thi Hoai An, L., & Dinh Tao, P. (1997). Solving a class of linearly constrained indefinite quadratic problems by DC algorithms. *Journal of Global Optimization*, 11, 253–285.
- Ware, J. E., Jr, & Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. *Medical Care*, 30, 473–483.
- Wright, A., Hannon, J., Hegedus, E. J., & Kavchak, A. E. (2012). Clinimetrics corner: A closer look at the minimal clinically important difference (MCID). *Journal of Manual & Manipulative Therapy*, 20, 160–166.
- Xu, Y., Yu, M., Zhao, Y.-Q., Li, Q., Wang, S., & Shao, J. (2015). Regularized outcome weighted subgroup identification for differential treatment effects. *Biometrics*, 71, 645–653.

## Appendices

### Appendix 1

To solve (6), we need to derive its dual problem by replacing the loss function in  $s_1$  with slack variables  $\xi_i$ ,  $i = 1, \dots, n$ , and adding two sets of constraints. This leads to

$$\min_{\omega, \xi} \frac{1}{n} \sum_{i=1}^n \left\{ \xi_i - \frac{1}{\delta} t_{1,i}(\alpha^{(k)}, \beta^{(k)}) y_i (\alpha + \beta^T \mathbf{z}_i) \right\} + \frac{\lambda}{2} \beta^T \beta, \quad (\text{A1})$$

subject to  $\xi_i \geq 0$  and  $\xi_i \geq 1 - (1/\delta) y_i (x_i - \alpha - \beta^T \mathbf{z}_i)$ ,  $i = 1, \dots, n$ , where  $t_{1,i}(\alpha^{(k)}, \beta^{(k)}) = \mathbb{1}\{y_i(x_i - \alpha^{(k)} - \beta^{(k)T} \mathbf{z}_i) < 0\}$ . Then its primary Lagrangian is

$$L_p = \frac{1}{n} \sum_{i=1}^n \left\{ \xi_i - \frac{1}{\delta} t_{1,i}(\alpha^{(k)}, \beta^{(k)}) y_i (\alpha + \beta^T \mathbf{z}_i) \right\} + \frac{\lambda}{2} \beta^T \beta - \frac{1}{n} \sum_{i=1}^n \tau_i \left\{ \xi_i - 1 + \frac{1}{\delta} y_i (x_i - \alpha - \beta^T \mathbf{z}_i) \right\} - \frac{1}{n} \sum_{i=1}^n \gamma_i \xi_i,$$

where  $\boldsymbol{\tau} = (\tau_1, \dots, \tau_n)^T$  and  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_n)^T$  are vectors of non-negative Lagrange multipliers, corresponding to the two sets of constraints in (A1). Setting derivatives of the Lagrangian with respect to the primary space variables  $\omega$  and  $\xi$  to 0, we get

$$0 = \frac{1}{\delta} \sum_{i=1}^n y_i \{ \tau_i - t_{1,i}(\alpha^{(k)}, \beta^{(k)}) \},$$

$$\beta = \frac{1}{n\lambda\delta} \sum_{i=1}^n y_i \{ t_{1,i}(\alpha^{(k)}, \beta^{(k)}) - \tau_i \} \mathbf{z}_i,$$

$$1 = \tau_i + \gamma_i, \quad i = 1, \dots, n.$$

Plugging them back to  $L_p$ , we have

$$\begin{aligned}
 L_p &= -\frac{1}{n\delta} \sum_{i=1}^n \tau_i y_i x_i + \frac{1}{n} \sum_{i=1}^n \tau_i - \frac{\lambda}{2} \boldsymbol{\beta}^T \boldsymbol{\beta}, \\
 &\propto -2 \sum_{i=1}^n \delta y_i \tau_i x_i + 2 \sum_{i=1}^n \delta^2 \tau_i \\
 &\quad - \frac{1}{n\lambda} \left\{ \sum_{i=1}^n y_i \tau_i \mathbf{z}_i^T \sum_{i=1}^n y_i \tau_i \mathbf{z}_i \right. \\
 &\quad \left. - 2 \sum_{i=1}^n y_i t_{1,i}(\boldsymbol{\alpha}^{(k)}, \boldsymbol{\beta}^{(k)}) \mathbf{z}_i^T \sum_{i=1}^n y_i \tau_i \mathbf{z}_i \right\}, \\
 &= -\frac{1}{n\lambda} \sum_{i=1}^n y_i \tau_i \mathbf{z}_i^T \sum_{i=1}^n y_i \tau_i \mathbf{z}_i + 2 \sum_{i=1}^n (\delta^2 - \delta y_i x_i) \tau_i \\
 &\quad + \frac{2}{n\lambda} \sum_{i=1}^n y_i t_{1,i}(\boldsymbol{\alpha}^{(k)}, \boldsymbol{\beta}^{(k)}) \mathbf{z}_i^T \sum_{i=1}^n y_i \tau_i \mathbf{z}_i, \\
 &= -\boldsymbol{\tau}^T \mathbf{Q} \boldsymbol{\tau} + \mathbf{b}^T \boldsymbol{\tau} + 2\mathbf{t}_1(\boldsymbol{\alpha}^{(k)}, \boldsymbol{\beta}^{(k)})^T \mathbf{Q} \boldsymbol{\tau},
 \end{aligned}$$

where  $\mathbf{Q}$  is a square matrix with  $[i, j]$ th element as  $\langle y_i \mathbf{z}_i, y_j \mathbf{z}_j \rangle / (n\lambda)$ ,  $\mathbf{t}_1(\boldsymbol{\alpha}^{(k)}, \boldsymbol{\beta}^{(k)}) = \{t_{1,i}(\boldsymbol{\alpha}^{(k)}, \boldsymbol{\beta}^{(k)})\}_{i=1}^n$ ,  $\mathbf{b} = 2\{\delta^2 - \delta y_i x_i\}_{i=1}^n$ .

## Appendix 2

To solve (15), we need to derive its dual problem by replacing the loss function in  $h_1$  with slack variables  $\xi'_i, i = 1, \dots, n$ , and adding two sets of constraints. This results in

$$\begin{aligned}
 \min_{\boldsymbol{\eta}, \boldsymbol{\xi}'} \quad & \frac{1}{n} \sum_{i=1}^n \left\{ \xi'_i - \frac{1}{\delta} t_{2,i} \right. \\
 & \left. (w^{(k)}, \mathbf{v}^{(k)}) y_i \left( w + \sum_{j=1}^n v_j K(\mathbf{z}_i, \mathbf{z}_j) \right) \right\} \\
 & + \frac{\lambda}{2} \sum_{i,j=1}^n v_i v_j K(\mathbf{z}_i, \mathbf{z}_j), \tag{A2}
 \end{aligned}$$

subject to  $\xi'_i \geq 0$  and  $\xi'_i \geq 1 - (1/\delta) y_i (x_i - w - \sum_{j=1}^n v_j K(\mathbf{z}_i, \mathbf{z}_j))$ , where  $t_{2,i}(w^{(k)}, \mathbf{v}^{(k)}) = \mathbb{1}\{y_i(x_i - w^{(k)} - \sum_{j=1}^n v_j K(\mathbf{z}_i, \mathbf{z}_j)) < 0\}$ . Then its primary Lagrangian is

$$\begin{aligned}
 L'_p &= \frac{1}{n} \sum_{i=1}^n \left\{ \xi'_i - \frac{1}{\delta} t_{2,i}(w^{(k)}, \mathbf{v}^{(k)}) y_i \right. \\
 & \left. \left( w + \sum_{j=1}^n v_j K(\mathbf{z}_i, \mathbf{z}_j) \right) \right\} + \frac{\lambda}{2} \sum_{i,j=1}^n v_i v_j K(\mathbf{z}_i, \mathbf{z}_j)
 \end{aligned}$$

$$\begin{aligned}
 & - \frac{1}{n} \sum_{i=1}^n \tau'_i \left\{ \xi'_i - 1 + \frac{1}{\delta} y_i \right. \\
 & \left. \left( x_i - w - \sum_{j=1}^n v_j K(\mathbf{z}_i, \mathbf{z}_j) \right) \right\} - \frac{1}{n} \sum_{i=1}^n \gamma'_i \xi'_i,
 \end{aligned}$$

where  $\boldsymbol{\tau}' = (\tau'_1, \dots, \tau'_n)^T$  and  $\boldsymbol{\gamma}' = (\gamma'_1, \dots, \gamma'_n)^T$  are vectors of non-negative Lagrange multipliers, corresponding to the two sets of constraints in (A2). Setting derivatives of the Lagrangian with respect to the primary space variables  $\boldsymbol{\eta}$  and  $\boldsymbol{\xi}'$  to 0, we get

$$\begin{aligned}
 0 &= \frac{1}{\delta} \sum_{i=1}^n y_i \{ \tau'_i - t_{2,i}(w^{(k)}, \mathbf{v}^{(k)}) \}, \\
 \mathbf{v} &= \frac{1}{n\lambda\delta} \sum_{i=1}^n y_i \{ t_{2,i}(w^{(k)}, \mathbf{v}^{(k)}) - \tau'_i \} \boldsymbol{\phi}_i, \\
 1 &= \tau'_i + \gamma'_i, i = 1, \dots, n.
 \end{aligned}$$

Plugging them back to  $L'_p$ , we have

$$\begin{aligned}
 L'_p &= -\frac{1}{n\delta} \sum_{i=1}^n \tau'_i y_i x_i + \frac{1}{n} \sum_{i=1}^n \tau'_i - \frac{\lambda}{2} \sum_{i=1}^n \sum_{j=1}^n v_i v_j K(\mathbf{z}_i, \mathbf{z}_j), \\
 &\propto -2 \sum_{i=1}^n \delta y_i \tau'_i x_i + 2 \sum_{i=1}^n \delta^2 \tau'_i \\
 &\quad - \frac{1}{n\lambda} \left\{ \sum_{i=1}^n y_i \tau'_i \boldsymbol{\phi}_i^T \sum_{i=1}^n y_i \tau'_i \boldsymbol{\phi}_i \right. \\
 &\quad \left. - 2 \sum_{i=1}^n y_i t_{2,i}(w^{(k)}, \mathbf{v}^{(k)}) \boldsymbol{\phi}_i^T \sum_{i=1}^n y_i \tau'_i \boldsymbol{\phi}_i \right\}, \\
 &= -\frac{1}{n\lambda} \sum_{i=1}^n y_i \tau'_i \boldsymbol{\phi}_i^T \sum_{i=1}^n y_i \tau'_i \boldsymbol{\phi}_i + 2 \sum_{i=1}^n (\delta^2 - \delta y_i x_i) \tau'_i \\
 &\quad + \frac{2}{n\lambda} \sum_{i=1}^n y_i t_{2,i}(w^{(k)}, \mathbf{v}^{(k)}) \boldsymbol{\phi}_i^T \sum_{i=1}^n y_i \tau'_i \boldsymbol{\phi}_i, \\
 &= -\boldsymbol{\tau}'^T \mathbf{Q}' \boldsymbol{\tau}' + \mathbf{d}^T \boldsymbol{\tau}' + 2\mathbf{t}_2(w^{(k)}, \mathbf{v}^{(k)})^T \mathbf{Q}' \boldsymbol{\tau}',
 \end{aligned}$$

where  $\mathbf{Q}'$  is a square matrix with  $[i, j]$ th element as  $\langle y_i \boldsymbol{\phi}_i, y_j \boldsymbol{\phi}_j \rangle / (n\lambda)$ ,  $\mathbf{t}_2(w^{(k)}, \mathbf{v}^{(k)}) = \{t_{2,i}(w^{(k)}, \mathbf{v}^{(k)})\}_{i=1}^n$ ,  $\mathbf{d} = 2\{\delta^2 - \delta y_i x_i\}_{i=1}^n$ .