

· 论著 · 方法学研究 ·

最小临床重要差异值计算方法统计性能比较研究



全旭源¹, 宋佳丽¹, 侯艳^{1, 2, 3}

1. 北京大学公共卫生学院生物统计学系 (北京 100191)
2. 北京大学肿瘤医院 (北京 100142)
3. 北京大学临床研究所 (北京 100191)

【摘要】 最小临床重要差异 (MCID) 指对患者有实际意义的最小疗效变化, 是监管机构评估药物或治疗方法临床价值的关键指标。其以患者为中心, 既体现了病情改善程度, 也体现了患者对变化的重视。然而各 MCID 计算方法在不同临床场景下的适用性尚缺乏系统比较, 临床研究中存在方法选择标准不一、估计结果差异显著等问题。既往研究表明 MCID 方法的选择需综合考虑样本量和治疗效果等特征, 不同场景下方法选择不当可能导致 MCID 估计出现严重偏倚, 影响临床研究结论的稳健性与监管决策的科学性。因此, 合理选择统计学方法是获取准确和可靠 MCID 值的关键环节。本文系统地阐述了临床研究中常用的 MCID 计算方法, 基于模拟实验比较多种锚定法和分布法的统计性能, 为 MCID 方法选择提供合理依据。

【关键词】 最小临床重要差异; 均方根误差; 锚定法; 分布法

【中图分类号】 R195.1 **【文献标志码】** A

A comparative study on the statistical performance of methods for estimating the minimal clinically important difference

QUAN Xuyuan¹, SONG Jiali¹, HOU Yan^{1,2,3}

1. Department of Biostatistics, School of Public Health, Peking University, Beijing 100191, China

2. Peking University Cancer Hospital & Institute, Beijing 100142, China

3. Peking University Clinical Research Center, Beijing 100191, China

Corresponding author: HOU Yan, Email: houyan@bjmu.edu.cn

【Abstract】 The minimal clinically important difference (MCID) refers to the smallest change in treatment effect that is meaningful to patients and serves as a key metric for regulatory agencies in evaluating the clinical value of drugs or therapeutic interventions. Rooted in a patient-centered approach, the MCID not only reflects the magnitude of clinical improvement but also embodies patients' emphasis on the change. However, the applicability of various MCID estimation methods across clinical scenarios has not been systematically compared, leading to inconsistent methodological selection and significant variability in estimated results in clinical research. Previous studies have indicated that the selection of appropriate MCID methods should comprehensively consider characteristics such as sample size and treatment effect; inappropriate

DOI: 10.12173/j.issn.1005-0698.202512063

基金项目: 国家自然科学基金面上项目 (82173615、82373682)

通信作者: 侯艳, 助理教授, 副研究员, 博士研究生导师, Email: houyan@bjmu.edu.cn

method selection across different scenarios may introduce significant bias into MCID estimates, thereby compromising the robustness of clinical research conclusions and the scientific validity of regulatory decision-making. Therefore, the appropriate selection of statistical methods is critical for obtaining accurate and reliable MCID estimates. This article systematically reviews commonly used MCID estimation methods in clinical research and evaluates the statistical performance of multiple anchor-based and distribution-based approaches through simulation experiments, providing a reasonable basis for MCID method selection.

【Keywords】 Minimal clinically important difference; Root mean square error; Anchor-based methods; Distribution-based methods

最小临床重要差异 (minimal clinically important difference, MCID) 最早由 Jachske 等^[1] 研究者提出, 指在不考虑不良反应和成本的前提下, 被患者认可的最小临床疗效评价问卷得分变化值。该定义注重以患者为中心, 既体现了改善的幅度, 也体现了患者对变化的重视^[2-3]。随着时间的推移, MCID 概念采用了多种不同的定义, 如最小重要差异、最小重要变化、最小可检测变化等^[4-5]。例如, Caroline 等^[6] 认为, 最小重要变化强调患者感知到的重要的最小变化阈值, 即在一段时间内, 患者主观感受到自己发生了重要变化的最小值; 不应与最小重要差异混淆, 最小重要差异关注的是不同患者群体之间的最小重要差异, 而非个体变化。但所有这些变体都有一个共同的目标, 即从患者角度量化具有临床实际意义的改变。从监管的角度看, 是否具有统计学意义与样本量有关, 也就是说只要样本足够大, 即使无临床意义, 也可能出现有统计学意义的结果。而 MCID 值利用了对临床实际意义的量化界定, 克服仅从统计学差异下结论的缺点, 可为监管决策提供更科学的依据^[7]。

目前, MCID 计算主要采用锚定法和分布法^[8-9]。锚定法将患者报告的得分变化与外部锚点联系起来确定 MCID 值, 这里锚点通常指患者整体变化评分 (patient global rating of change, PGRC), 其是一种单一项目评分, 要求患者评价自基线以来的整体或特定健康状况的变化^[10-11]。分布法根据样本的分布特征进行计算, 主要基于描述变化程度的指标, 如标准差、效应量或测量的标准误等获得标准数值, 检测基线和不同时间点之间的变化超过变化预期的程度^[12-13]。然而, 现有研究未对各方法的统计性能进行深入研究, 即不熟悉各方法估计值在不同样本量和改善程度等条件下的准确性和可靠性, 存在

方法选择标准不一等问题。因此, 本研究基于模拟实验系统评估多种 MCID 计算方法在不同临床情景下的统计性能, 并以均方根误差 (root mean square error, RMSE) 为主要评价指标比较其准确性与稳健性, 旨在明确各方法的适用范围与优选策略, 为临床研究中 MCID 方法选择提供依据, 从而为临床决策及监管评价提供更可靠的支持。

1 原理与方法

1.1 MCID 概念模型

为计算 MCID 并研究不同方法的统计性能, 需深入理解 MCID 值的定义和构建过程。2022 年, Vanier 等^[14] 在健康相关生活质量变化模型基础上, 结合调查响应心理学理论, 进一步对 MCID 概念模型进行扩展。最终建立的 MCID 概念模型见图 1。基于该概念模型, 将 MCID 定义为患者的最小感知变化 (minimal perceived change, MPC)。感知变化 (perceived change,

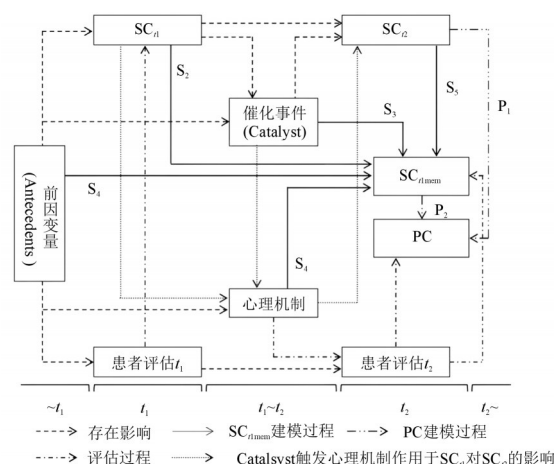


图 1 MCID 概念模型

Figure 1. MCID conceptual model

注: $SC_{i,t1}$, 患者 i 在 t_1 时的特征水平; $SC_{i,t2}$, 患者 i 在 t_2 时的特征水平; $SC_{i,tmem}$, 患者 i 在 t_2 时记忆中的 t_1 时研究特征水平; PC, 感知变化。

PC) 主要指患者主观感受到的症状缓解或生活质量改善等变化, 通常用于患者主观指标的评估, 如疼痛、焦虑等。PC 在患者报告结局 (patient-reported outcomes, PRO) 得分或指数上体现为两次测量时量表得分或指数差异。当患者认为已产生对其有重要意义的最小变化时, 对应的量表得分或指数差值即为 MCID 值。此外, 该模型描述了研究对象在两个时间点回答 PRO 条目以及在第二个时间点测量 PGRC 时, 相关变量之间的关系, 从而用于后续感知变化等建模。本研究探索了几种 MCID 计算方法在不同场景下的准确性和可靠性, 为 MCID 方法应用提供参考。

1.2 拟比较的 MCID 方法

1.2.1 锚定法

(1) 基于 ROC 分析的方法

基于受试者工作特征 (receiver operating characteristic, ROC) 分析方法是利用 ROC 曲线基于灵敏度 (sensitivity, Se) 和特异度 (specificity, Sp) 来计算 MCID 值。包括最佳临界点法和约登指数最大法。

①最佳临界点法, 公式如下^[15]:

$$MCID = \min(\Delta x), ((1 - Sp), Se) \rightarrow \min(d) \quad \text{公式 (1)}$$

注: d 为 ROC 曲线上的某点与坐标系中 $(1 - Sp = 1, Se = 1)$ 对应的点之间的欧几里得距离, Δx 为该距离最短时对应的临界值。

②约登指数最大法, 公式如下^[15]:

$$MCID = \min(\Delta x), ((1 - Sp), Se) \rightarrow \max(Se + Sp) \quad \text{公式 (2)}$$

注: $\max(Se + Sp)$ 为灵敏度和特异度达到最大值, Δx 为约登指数最大时对应的临界值。

(2) 基于均值变化的方法

基于均值变化法的核心思想是基于锚点中自评为“稍微变好”或“稳定”患者的平均变化值计算 MCID 值。包括最小变化组平均值法和最小变化组与稳定组平均值差法。

①最小变化组平均值法, 公式如下^[16]:

$$MCID = \overline{\Delta x} | y^2 = 1 \quad \text{公式 (3)}$$

注: $y^2 = 1$ 为最小改善组对应的人群, $\overline{\Delta x}$ 为两次测量过程中各研究对象差值的平均值。

②最小变化组与稳定组平均值差法, 公式如下^[17]:

$$MCID = (\overline{\Delta x} | y^2 = 1) - (\overline{\Delta x} | y^2 = 0) \quad \text{公式 (4)}$$

注: $y^2 = 1$ 和 $y^2 = 0$ 为最小改善组和最小稳定组对应的人群, $\overline{\Delta x}$ 为两次测量过程中各研究对象差值的平均值。

(3) 预测建模法

预测建模法利用逻辑回归来估计患者发生临床重要变化的概率。研究者以患者是否报告重要变化为因变量, 以 PRO 量表得分变化值为自变量, 建立预测模型并得到截距和回归系数, 公式如下:

$$\text{logit}(Z) = b_0 + b_1 \Delta X \quad \text{公式 (5)}$$

注: ΔX 为量表得分变化, b_0 为截距, b_1 为得分变化的回归系数。

随后, 通过求解使预测改善概率等于样本总体改善率的分数变化值, 即可得到 MCID 的估计值。公式如下^[18]:

$$MCID = \frac{\ln\left(\frac{\text{prop}(z=1)}{1 - \text{prop}(z=1)}\right) - b_0}{b_1} \quad \text{公式 (6)}$$

注: $z=1$ 为发生最小改善的研究对象, b_0 为截距, b_1 为得分变化的回归系数。

1.2.2 分布法

(1) 基于样本变异性的方法—Cohen's d 方法

Cohen's d 方法是从效应量角度考虑。小、中、大效应量对应的 Cohen's d 界值分别为 0.2、0.5 和 0.8。在没有合理先验选择 d 值的情况下, 有统计学意义的差异通常认为至少为中等效应量, 即本研究选择 $d=0.5$, 该取值源自 Cohen 对效应量的经典划分, 是当前 MCID 方法学研究中的常用基准, 为本研究提供了与同类研究可比的参照^[19]。公式如下:

$$d = \frac{\Delta x}{S_x^1} (0.2, 0.5, 0.8) \quad \text{公式 (7)}$$

注: Δx 为两个时间点测量的均值差, S_x^1 为第一次测量时的标准差。

(2) 基于测量工具精确性的方法—1 倍测量标准误差法

根据经典测量理论, PRO 得分的测量标准误差 (standard error of measurement, SEM) 源于评估工具的信度, 可视为其测量学特征之一, 因此具有较好的应用普适性^[20]。公式如下:

$$MCID = 1 \cdot s_{x^1} \sqrt{1 - \alpha_{\text{Cronbach}}} \quad \text{公式 (8)}$$

注: S_x^1 为第一次测量时的标准差, α_{Cronbach} 为第一次测量时量表的 Cronbach's α 值。

2 模拟实验

为探索 MCID 值的几种计算方法的统计学特性, 本研究需模拟出 n 个研究对象在 2 个时间点对 PRO

条目的回答,以及在 t_2 时刻对PGRC的回答。数据生成机制如下(关键参数设置依据见附件1):

2.1 PRO 条目数据

在2个测量时间点上,基于已知潜在特征 θ 的分布,模拟研究对象对PRO各条目的回答。需满足:①已知 θ 在总体中的分布;②模拟的PRO数据要具备足够的可靠性和结构效度,且前提假设为单维性。

公式(9)模拟出个体在时间点 t 对某条目 j 某选项 h 的回答概率。其将研究对象 i 在 t 时间点,对条目 j 回答 h 类别的概率建模为“潜在特征 $\theta_i^{(t)}$ 的水平”和“项目阈值参数 δ_{jl} ”的函数。

$$P(X_{ij}^{(t)} = h | \theta_i^{(t)}, \delta_{j1}, \dots, \delta_{jmj}) = \frac{\exp(h\theta_i^{(t)} - \sum_{l=1}^{m^0} \delta_{jl})}{\sum_{c=0}^{m_j} \exp(c\theta_i^{(t)} - \sum_{l=1}^c \delta_{jl})} \quad \text{公式(9)}$$

注: N 为患者 i ($i=1, \dots, N$); J 为PRO条目 j ($j=1, \dots, J$); t 为测量时间点 t ($t=1, 2$); h 为对PRO第 j 项条目的具体选择类别; $\theta_i^{(t)}$ 为患者 i 在 t 时刻的潜在特征水平,代表患者 i 在 t 时刻的某种心理或能力水平; δ_{jl} 为PRO第 j 项条目的项目阈值 l ($l=1, \dots, m_j$),代表了条目 j 在不同回答类别 h 的难度或区分度; $\theta \sim N\left(\begin{pmatrix} \sigma^{2(1)}=1 & \rho \\ \rho & \sigma^{2(2)} \end{pmatrix}, \begin{pmatrix} \sigma^{2(1)}=1 & \rho \\ \rho & \sigma^{2(2)} \end{pmatrix}\right)$,即潜在特征 θ 均服从正态分布,且两个时间的潜在特征具有相关性; $\theta^{(1)} \sim N(\alpha^{(1)}, \sigma_{\theta^{(1)}}^2)$; $\theta^{(2)} \sim N(\alpha^{(2)}, \sigma_{\theta^{(2)}}^2)$; $\text{cor}(\theta^{(1)}, \theta^{(2)}) = \rho = 0.7$ 。

2.2 个体感知变化

公式(10)模拟生成患者PC数据。其为患者在第2次测量时潜在特征 $\theta^{(2)}$ 水平减去记忆的基线时潜在特征 θ 水平。

$$PC = \theta^{pc(2)} = \theta^{(2)} - \theta^{(1)} \quad \text{公式(10)}$$

注: $\theta^{(1)}$ 为潜在特征 θ 在 t 时间点的水平,如生活质量,疼痛等在 t 时刻的水平; $\theta^{(2)}$ 为第2次测量时潜在特征的水平; $\theta^{pc(2)}$ 为感知到的潜在特征 θ 的变化水平; $\theta^{(2)}$ 为在第2次测量时,记忆的基线时潜在特征 θ 水平。

对 $\theta^{(2)}$ 进行建模,如公式(11):

$$\theta^{(2)} = \beta_1 \theta^{(1)} + \beta_2 \theta^{(2)} + \beta_3 U \quad \text{公式(11)}$$

注: $\beta_1 = \{0.3, 0.5, 1\}$, $\beta_2 = \{0, 0.3\}$, $\beta_1 + \beta_2 + \beta_3 = 1$; $\beta_1, \beta_2, \beta_3$: $\theta^{(1)}, \theta^{(2)}, U$ 在决定 $\theta^{(2)}$ 时的相对重要性; $U \sim N(\mu_u = 0, \sigma_u^2 = 1)$; $\theta^{(2)} \sim N(\mu_{\theta^{(2)}} = \beta_2 \alpha^{(2)}, \sigma_{\theta^{(2)}}^2 = \beta_1^2 + \beta_2^2 + \beta_3^2 + 2\beta_1 \beta_2 \rho)$ 。

其中, β_1 反映个体对基线状态的真实记忆能力,高 β_1 意味着能独立于当前感受回顾过去, β_2 反映个体用当前状态重构过去的倾向,高 β_2 意味着个体对基线状态的回忆显著受到当前状态的影响,体现了记忆建构过程中现时信息对过往体验的重塑作用。二者共同刻画了自传体记忆中记忆保持与现时重构的过程。 β_3 反映其他偶然因素对回忆的影响。当 β_1 高时,PGRC更能反映真实变

化;当 β_2 高时,PGRC与真实变化脱节,锚定法可能出现偏倚。

2.3 PGRC 数据

PGRC 阈值 (T_s) 是将连续的PC转换为PGRC中不同类别之间的分界点。 T_s 公式如下:

$$T_s \sim N\left(\lambda_s = \pm \frac{2\sigma_{\theta^{pc(2)}}(2S-1)}{K-1}, \zeta_s^2 = \left(\frac{D}{K-1}\right)^2\right) \quad \text{公式(12)}$$

注: T_s 是PGRC的类别边界(阈值),即将连续的感知变化(PC)划分为PGRC不同类别之间的分界点; D 为离散因子,控制PGRC阈值(T_s)分布范围,即决定不同患者变化敏感度的差异有多大; λ_s 是PGRC阈值 T_s 在人群中的均值,即某个特定类别边界在人群中的平均水平。

在确定PGRC阈值分布的基础上,假设:①阈值围绕0点(即PC为零的点)对称分布;②从0点开始,PC分布的同一侧上,相邻阈值之间间隔相等;③阈值的分布是PGRC响应类别数量 K 的函数, K 是一个至少等于3的奇数自然数;④给定的 K 值下,每个阈值的方差相同。

可得MPC,如公式(13):

$$MPC = \lambda_1 = \frac{2\sigma_{\theta^{pc(2)}}}{K-1} \quad \text{公式(13)}$$

对于患者 i ,在第2次测量时对PGRC的反应 $Y_i^{(2)}$ 如下:

$$Y_i^{(2)} = \begin{cases} \text{if } \tau_{-1,i} < \theta_i^{pc(2)} \leq \tau_{1,i} \rightarrow 0 \\ \text{else if } \theta_i^{pc(2)} \leq \tau_{\min(s),i} \rightarrow -\frac{K-1}{2} \\ \text{else if } \tau_{\max(s),i} \leq \theta_i^{pc(2)} \rightarrow -\frac{K-1}{2} \\ \text{else if } \tau_{s-1,i} < \theta_i^{pc(2)} \leq \tau_{s,i} \rightarrow s \quad (s < 0) \\ \text{else if } \tau_{s-1,i} < \theta_i^{pc(2)} \leq \tau_{s,i} \rightarrow s-1 \quad (s > 0) \end{cases}$$

注: $\tau_{s,i}$ 是 K 在不同取值下(如 $K=3$ 时, $S = \{-1, 1\}$),患者 i 的PGRC阈值。

2.4 真实MCID值

通过模拟100 000个个体样本在两次测量时对PRO量表的回答,且在限制真实变化等于潜在特征 θ 度量中模拟的真实MCID值的前提下,即将MCID真实值建模为基于 K, J, M, β_1 和 β_2 的函数,满足公式(14):

$$\alpha^{(2)} = \frac{2\sigma_{\theta^{pc(2)}}}{K-1} \quad \text{公式(14)}$$

然后,计算该模拟数据集中的PRO总分平均值。每次测量时PRO总分分数的平均差异是分数度量中真实MCID值的估计值。对各参数组合下的每一组重复此过程,以获得相应参数组合下的

真实MCID值。公式(14)本质上是将MCID值定义为人群中从“稳定”感知状态跨越到“轻微变化”感知状态所需的变化阈值之均值。在数据生成过程中,通过设定整体变化评分的类别边界在人群中的分布参数,使真实MCID成为由感知变异幅度和PGRC中类别数共同决定的已知参数,从而为评价不同估计方法的偏倚提供了具有理论锚点的评价基准。

2.5 真数据标准化

真实MCID值是基于潜在特征 θ 的度量表示,MCID估计值是基于PRO量表得分表示,非同一度量尺度。因此需统一两者的度量标准。PRO总分需标准化为0到100的尺度。此外,需将真实MCID值映射到与PRO总分相同的度量标准上。

3 模拟结果

3.1 样本量对统计性能的影响

该模拟实验是研究样本量 n 对RMSE值的影响(n 取值包括50、100、200、500);以及不同样本量 n 下,最佳的MCID方法。具体而言,在一个量表条目数 $J=20$ 、选项类别数 $M=4$ 、PGRC类别数 $K=5$ 的数据库,研究相同 β_1 、 β_2 、 $\alpha^{(2)}$ 、 D 的参数组合下, n 与RMSE值的关系。图2描述了在4个样本量的情景下,不同MCID计算方法的RMSE值。其中,最佳临界值法、约登指数最大法、最小变化组平均值法、最小变化组与稳定组平均值差法和预测建模法的5种锚定法均呈现出随着 n 的增大, RMSE值连续降低的趋势。

3.2 改善程度对统计性能的影响

该模拟实验是研究改善程度 $\alpha^{(2)}$ 对RMSE值的影响(其中 $\alpha^{(2)}$ 取值包括0、0.2、0.5),以及不同改善程度 $\alpha^{(2)}$ 下,最佳的MCID方法。具体而言,1个量表条目数 $J=20$ 、选项类别数 $M=4$ 、PGRC类别数 $K=5$ 数据库,研究相同 β_1 、 β_2 、 n 、 D 的参数组合下, $\alpha^{(2)}$ 与RMSE值的关系。图3描述了在3个改善程度的情景下,不同MCID计算方法的RMSE值。在改善程度 $\alpha^{(2)}=0.5$ 时,预测建模法和SEM法具有较好的准确性和可靠性, $M(P_{25}, P_{75})$ 分别为3.16(1.45, 3.91)和1.56(0.86, 3.07)。

3.3 PGRC类别数 K 对统计性能的影响

该模拟实验是研究 K 对RMSE值的影响(其中 K 取值包括3、5、7)。具体而言,在一个量表

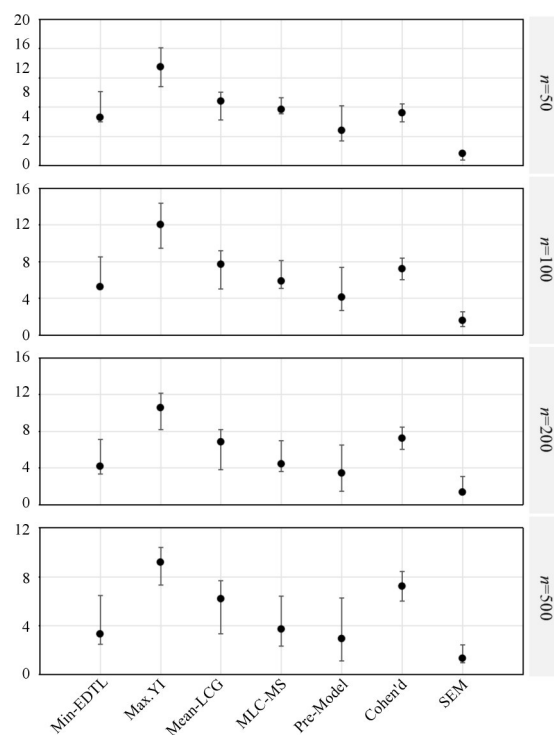


图2 相同参数组合下不同样本对MCID方法RMSE值的影响

Figure 2. Influence of different sample sizes on the RMSE of MCID estimation methods under identical parameter settings

注: Min-EDTL. 最佳临界点法; Max. YI. 约登指数最大法; Mean-LCG. 最小变化组平均值法; MLC-MS. 最小变化组与稳定组平均值差法; Pre-Model. 预测建模法。

条目数 $J=20$ 、选项类别数 $M=4$ 、样本量 $n=200$ 的数据库,研究相同 β_1 、 β_2 、 $\alpha^{(2)}$ 、 D 的参数组合下, K 与RMSE值的关系。图4描述了在PGRC中不同类别数 K 的情景下,不同MCID计算方法的RMSE值。最佳临界值法、预测建模法、Cohen's d法和SEM法在各种情景下呈现出, $K=7$ 对应的RMSE值普遍小于 $K=5$ 和3的RMSE值;约登指数最大法、最小变化组平均值法、最小变化组与稳定组平均值差法中, K 与RMSE值的关系在各种情景下未呈现出明显的规律或一致的趋势。

3.4 统计性能比较

在临床中, n 反映样本量的大小, $\alpha^{(2)}$ 反映治疗的改善程度。因此,可根据 n 和 $\alpha^{(2)}$ 对临床场景进行分类,并在每类场景下对比各方法的RMSE值。基于 n 和 $\alpha^{(2)}$ 划分出的4种临床场景。其中, $n=200$ 或500, $\alpha^{(2)}=0$ 代表了治疗无效且大样本的场景; $n=200$ 或500, $\alpha^{(2)}=0.2$ 或0.5代表了治疗改善或有效,且大样本的场景; $n=50$,

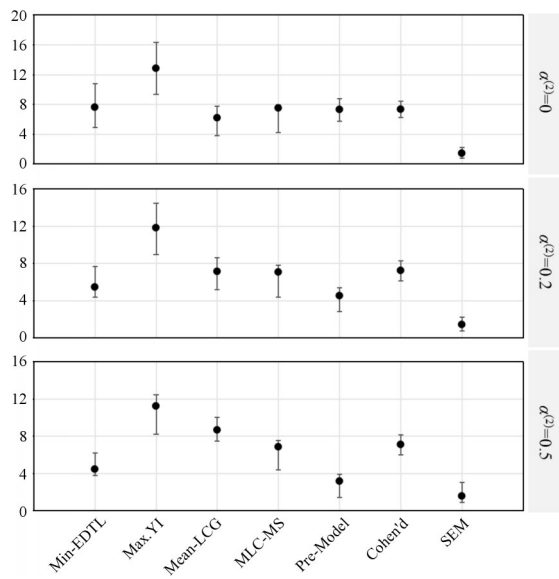


图3 相同参数组合下不同改善程度对 MCID 方法 RMSE 值的影响

Figure 3. Influence of varying degrees of improvement on the RMSE of MCID estimation methods under identical parameter settings

注：Min-EDTL. 最佳临界点法；Max. YI. 约登指数最大法；Mean-LCG. 最小变化组平均值法；MLC-MS. 最小变化组与稳定组平均值差法；Pre-Model. 预测建模法。

$\alpha^{(2)} = 0$ 代表了治疗无效且小样本的场景； $n=50$, $\alpha^{(2)} = 0.2$ 或 0.5 代表了治疗改善或有效，且小样本的场景。治疗无效指 t_1 和 t_2 时间点 PRO 得分差异无统计学意义，治疗改善或有效则指有统计学意义。在一个量表条目数 $J=20$ 、选项类别数 $M=4$ 、 $K=7$ 、 $D=1$ 的模拟数据库中，相同 n 、 β_1 、 β_2 、 $\alpha^{(2)}$ 参数组合下，比较各 MCID 方法的 RMSE 值。

图5展示了4种临床场景下，不同MCID方法比较结果。在 $n=200$ 或 500 , $\alpha^{(2)} = 0$ 的临床场景中，SEM法的RMSE中位数最小 [1.93 (0.71, 3.47)]，其次为最小变化组平均值法 [4.90 (2.76, 5.56)] 和最小变化组与稳定组平均值差法 [5.01 (3.45, 5.96)]。在 $n=200$ 或 500 , $\alpha^{(2)} = 0.2$ 或 0.5 的临床场景中，预测建模法和SEM法的RMSE值普遍较低， $M(P_{25}, P_{75})$ 分别为 2.51 (1.46, 4.41) 和 1.89 (0.73, 3.41)。在 $n=50$, $\alpha^{(2)} = 0$ 的临床场景中，SEM法的RMSE值的 $M(P_{25}, P_{75})$ 为 2.67 (1.33, 4.23)。在 $n=50$, $\alpha^{(2)} = 0.2$ 或 0.5 的临床场景中，SEM法的RMSE值普遍较低， $M(P_{25}, P_{75})$ 为 2.62 (1.24, 4.18)；预测建模法、Cohen's d法在 $\beta_1=0.5$ 和 $\beta_2=0.3$ ，以及 $\beta_1=1$ 和 $\beta_2=0$ 的参数组合下RMSE值亦普遍较低， $M(P_{25}, P_{75})$ 分别为 4.01 (2.78, 5.63) 和 3.99

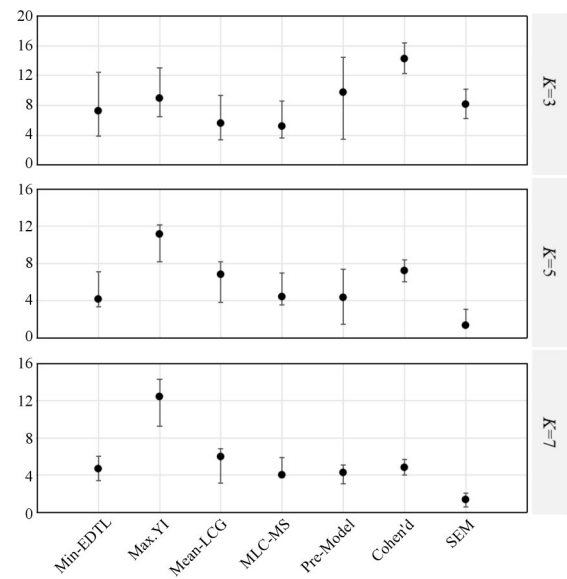


图4 相同参数组合下不同类别数对 MCID 方法 RMSE 值的影响

Figure 4. Influence of different numbers of categories on the RMSE of MCID estimation methods under identical parameter settings

注：Min-EDTL. 最佳临界点法；Max. YI. 约登指数最大法；Mean-LCG. 最小变化组平均值法；MLC-MS. 最小变化组与稳定组平均值差法；Pre-Model. 预测建模法。

(2.51, 5.28)。不同方法的偏倚值（真实MCID值与模拟MCID值差值）和模型MCID值的方差结果与RMSE结果基本一致，详见附件2“模拟实验汇总结果”。

4 实例分析

4.1 背景介绍

术前焦虑是指患者在接受手术前，由于对手术过程、麻醉风险等因素产生的一种紧张、恐惧和担忧的情绪状态^[21]。术前焦虑与信息需求量表 (Amsterdam Preoperative Anxiety and Information Scale, APAIS) 因其简便高效的特点被广泛应用于术前焦虑评估，该量表由 Moerman 等研究者^[22] 开发。2025年，Salzmann 等研究者^[23] 基于单中心前瞻性随机对照研究，比较个性化信息干预联合标准麻醉咨询与单纯标准咨询的效果。结果显示，干预组的麻醉及手术相关焦虑评分显著低于对照组，差异具有统计学及临床意义，其MCID分别为1.03分和1.13分。

术前焦虑患者 APAIS 的 MCID 值研究的实例分析共包含 314 名患者，在术前不同时点调查患者的麻醉和手术相关焦虑条目，以及针对麻醉和

手术的PGRC (K=7) 回答。

4.2 分析结果

表1展示了314名患者麻醉和手术相关条目的得分统计。表2展示了患者在麻醉和手术相关

的焦虑方面对PGRC的回答。两个维度中 t_1 和 t_2 时间点比较差异均有统计学意义，且属于大样本。因此，选择预测建模法和SEM法计算MCID值(表3展示了MCID值计算结果)。经计算，预测建

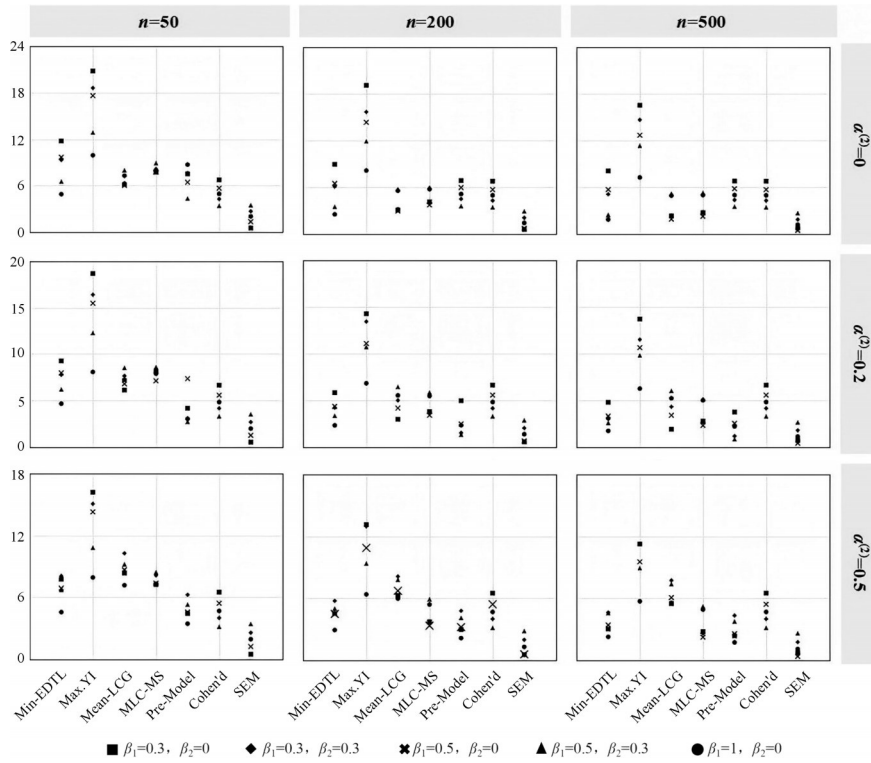


图5 相同参数组合下不同MCID方法RMSE值比较

Figure 5. Comparison of RMSE across different MCID estimation methods under identical parameter settings

注：Min-EDTL. 最佳临界点法；Max. YI. 约登指数最大法；Mean-LCG. 最小变化组平均值法；MLC-MS. 最小变化组与稳定组平均值差法；Pre-Model. 预测建模法。

表1 患者治疗前后得分统计

Table 1. Statistics of patient scores before and after treatment

维度	Cronbach's α t_1	t_1	t_2	差值	t	P
麻醉相关	0.718	6.56±1.98	4.83±1.96	1.73±1.42	21.579	<0.001
手术相关	0.768	6.77±2.14	4.62±1.86	2.15±1.50	25.389	<0.001

表2 患者在麻醉和手术相关的焦虑方面对PGRC回答

Table 2. Patient responses to the PGRC regarding anesthesia- and surgery-related anxiety

维度	非常差	差	稍差	无变化	稍好	好	非常好
麻醉相关	1 (0.32)	2 (0.64)	4 (1.28)	142 (45.22)	88 (28.03)	45 (14.33)	32 (10.19)
手术相关	2 (0.64)	4 (1.28)	7 (2.23)	133 (42.36)	84 (26.72)	49 (15.61)	35 (11.15)

表3 不同方法计算得到的MCID值

Table 3. MCID values calculated using different methods

计算方法	维度	MCID估计值	MCID参考值	Bias绝对值
预测建模法	麻醉相关	1.08	1.03	0.05
	手术相关	1.10	1.13	0.03
SEM法	麻醉相关	1.05	1.03	0.02
	手术相关	1.03	1.13	0.10

模法中，麻醉和手术两个维度的 Bias 绝对值分别为 0.05 和 0.03；SEM 法中，麻醉和手术两个维度的 Bias 绝对值分别为 0.02 和 0.10。

5 讨论

本研究通过模拟比较了多种 MCID 计算方法在不同临床场景下的统计性能，结果显示，在大样本且治疗无效的情境下，最小变化组平均值法、最小变化组与稳定组平均值差法以及 SEM 方法的 RMSE 相对较低；在大样本且治疗改善或有效的情境下，预测建模法与 SEM 方法的表现更佳；在小样本场景中，无论治疗是否有效，SEM 方法均表现出较好的稳健性。SEM 法主要依赖于量表的信度和标准差，与外部锚点无关，因此受样本量和改善率的影响相对较小，表现出跨场景的稳定性。预测建模法在样本量充足且存在显著分数变化时，能够通过 logistic 回归有效区分有或无改善的个体，因此在大样本有效场景中更具优势。而均值类锚定法依托“稍微变好”或“稳定”患者群体的分布特征，在治疗无效的情境下与真实 MCID 值更为接近，因此其主要评价指标 RMSE、偏倚值和方差均表现较好。

本研究的模拟结果与既往方法学研究相互印证并有所拓展。Terluin 等^[24]通过模拟研究发现，在变化分数方差相等且数据分布满足一定条件时，预测建模法与 ROC 分析法估计的最小重要变化值相比，前者具有更高的精度，表现为更窄的置信区间。本研究在大样本治疗有效场景中同样观察到预测建模法的优异表现，其 RMSE 中位数仅为 2.51，优于多数锚定法，验证了该方法在样本量充足且存在真实变化时的优势。Terluin 等^[25]进一步发现改善患者比例偏离 50% 时，最小重要变化估计会出现系统性偏差，并提出了基于预测模型的校正公式。本研究中预测建模法在治疗有效场景下的稳定表现，间接支持了该方法通过纳入变化分数与锚定关系进行校正的合理性。

与上述研究不同的是，本研究引入 MPC 概念，通过设置 β_1 和 β_2 参数控制回忆偏倚的程度。结果显示，在 β_1 和 β_2 较高的情境下，即回忆受当前状态影响较大时，锚定法的 RMSE 普遍升高，而 SEM 法仍保持稳健。这从模拟角度印证了回忆偏倚对锚定法估计效度的影响机制。此外，本研究发现 SEM 法在各类场景下均表现稳健，尤其在

小样本条件下优势明显。这与 Wyrwich 等^[26]基于经典测量理论提出的“SEM 法可作为个体重要变化判断依据”的观点一致，也为分布法在样本受限研究中的应用提供了实证支持。

本研究聚焦于 MCID 方法的统计性能，而不仅仅是报告估计值的差异。已有文献多从理论或经验角度探讨锚定法与分布法的优缺点，而较少通过大规模模拟系统性比较方法间的准确性与可靠性。本研究结果在一定程度上验证了 SEM 方法在文献中被认为具有“普适性”的观点，同时也提示预测建模法在特定场景下具有潜在优势。值得注意的是，预测建模法在理论上存在一定的争议，例如其估计逻辑可能被质疑为依赖于总体改善比例，但本研究的模拟结果显示，在大样本且治疗有效时，其 RMSE 依然优于其他方法。总的来说，临床研究中大样本治疗改善或有效时，SEM 法与预测建模法可作为首选，以提高 MCID 估计的准确性和可靠性；在小样本研究中，SEM 方法因其稳健性更值得推荐；在治疗无效的大样本情境下，均值类锚定法仍然具有实用价值。

附件见《药物流行病学杂志》官网附录 (<https://ywlbx.whuzhmedj.com/futureApi/storage/appendix/202512063.zip>)

利益冲突声明：作者声明本研究不存在任何经济或非经济利益冲突。

参考文献

- 1 Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference[J]. *Control Clin Trials*, 1989, 10(4): 407-415. DOI: 10.1016/0197-2456(89)90005-6.
- 2 Ameer B. Patient-reported outcomes: listening for what is most important in clinical care and patient-focused drug development [J]. *J Clin Pharmacol*, 2021, 61(7): 845-847. DOI: 10.1002/jcph.1867.
- 3 Crawford LS, Mateczak GJ, Moore EM, et al. Patient-centered drug development and the Learning Health System[J]. *Learn Health Syst*, 2017, 1(3): e10027. DOI: 10.1002/lrh2.10027.
- 4 King MT. A point of minimal important difference (MID): a critique of terminology and methods[J]. *Expert Rev Pharmacoecon Outcomes Res*, 2011, 11(2): 171-184. DOI: 10.1586/erp.11.9.
- 5 Beaton DE, Boers M, Wells GA. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research[J]. *Curr Opin Rheumatol*, 2002, 14(2): 109-114. DOI: 10.1097/00002281-200203000-00006.

- 6 Terwee CB, Peipert E JD, Champman R, et al. Minimal important change (MIC): a conceptual clarification and systematic review of MIC estimates of PROMIS measures[J]. *Qual Life Res*, 2021, 30(10): 2729–2754. DOI: [10.1007/s11136-021-02925-y](https://doi.org/10.1007/s11136-021-02925-y).
- 7 Arciero V, Delos Santos S, Koshy L, et al. Assessment of food and drug administration- and european medicines agency- approved systemic oncology therapies and clinically meaningful improvements in quality of life: a systematic review[J]. *JAMA Netw Open*, 2021, 4(2): e2033004. DOI: [10.1001/jamanetworkopen.2020.33004](https://doi.org/10.1001/jamanetworkopen.2020.33004).
- 8 Mcglothlin AE, Lewis RJ. Minimal clinically important difference: defining what really matters to patients[J]. *JAMA*, 2014, 312(13): 1342–1343. DOI: [10.1001/jama.2014.13128](https://doi.org/10.1001/jama.2014.13128).
- 9 Zhang HY, Xi X, Huang Y. The anchor design of anchor-based method to determine the minimal clinically important difference: a systematic review[J]. *Health Qual Life Outcomes*, 2023, 21(1): 74. DOI: [10.1186/s12955-023-02157-3](https://doi.org/10.1186/s12955-023-02157-3).
- 10 Devji T, Carrasco-Labra A, Qasim A, et al. Credibility of anchor-based minimal important differences for patient-reported outcomes: instrument development and reliability study[J]. *BMJ*, 2020, 369: m1714. DOI: [10.1136/bmj.m1714](https://doi.org/10.1136/bmj.m1714).
- 11 Kamper SJ, Maher CG, Mackay G. Global rating of change scales: a review of strengths and weaknesses and considerations for design[J]. *J Man Manip Ther*, 2009, 17(3): 163–670. DOI: [10.1179/jmt.2009.17.3.163](https://doi.org/10.1179/jmt.2009.17.3.163).
- 12 Copay AG, Subach BR, Glassman SD, et al. Understanding the minimum clinically important difference: a review of concepts and methods[J]. *Spine J*, 2007, 7(5): 541–546. DOI: [10.1016/j.spinee.2007.01.008](https://doi.org/10.1016/j.spinee.2007.01.008).
- 13 Revicki D, Hays RD, Cella D, et al. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes[J]. *J Clin Epidemiol*, 2008, 61(2): 102–109. DOI: [10.1016/j.jclinepi.2007.03.012](https://doi.org/10.1016/j.jclinepi.2007.03.012).
- 14 Vanier A, Leroy M, Hardouin J B. Toward a rigorous assessment of the statistical performances of methods to estimate the minimal important difference of patient-reported outcomes: a protocol for a large-scale simulation study[J]. *Methods*, 2022, 204: 396–409. DOI: [10.1016/j.ymeth.2022.02.006](https://doi.org/10.1016/j.ymeth.2022.02.006).
- 15 The WHOQOL Group. Development of the World Health Organization WHOQOL-BREF quality of life assessment[J]. *Psychol Med*, 1998, 28(3): 551–558. DOI: [10.1017/s0033291798006667](https://doi.org/10.1017/s0033291798006667).
- 16 Wang D, Willis DR, Yih Y. The pneumonia severity index: assessment and comparison to popular machine learning classifiers[J]. *Int J Med Inform*, 2022, 163: 104778. DOI: [10.1016/j.ijmedinf.2022.104778](https://doi.org/10.1016/j.ijmedinf.2022.104778).
- 17 Francis CY, Morris J, Whorwell PJ. The irritable bowel severity scoring system: a simple method of monitoring irritable bowel syndrome and its progress[J]. *Aliment Pharmacol Ther*, 1997, 11(2): 395–402. DOI: [10.1046/j.1365-2036.1997.142318000.x](https://doi.org/10.1046/j.1365-2036.1997.142318000.x).
- 18 Rector TS, Cohn JN. Assessment of patient outcome with the Minnesota Living with Heart Failure questionnaire: reliability and validity during a randomized, double-blind, placebo-controlled trial of pimobendan[J]. *Am Heart J*, 1992, 124(4): 1017–1025. DOI: [10.1016/0002-8703\(92\)90986-6](https://doi.org/10.1016/0002-8703(92)90986-6).
- 19 Cohen J. A power primer[J]. *Psychol Bull*, 1992, 112(1): 155–159. DOI: [10.1037//0033-2909.112.1.155](https://doi.org/10.1037//0033-2909.112.1.155).
- 20 Altman DG, Bland JM. Standard deviations and standard errors [J]. *BMJ*, 2005, 331(7521): 903. DOI: [10.1136/bmj.331.7521.903](https://doi.org/10.1136/bmj.331.7521.903).
- 21 Karpecki PM, Nichols KK, Sheppard JD. Addressing excessive evaporation: an unmet need in dry eye disease[J]. *Am J Manag Care*, 2023, 29(13 Suppl): S239–S247. DOI: [10.37765/ajmc.2023.89448](https://doi.org/10.37765/ajmc.2023.89448).
- 22 Moerman N, Vandam FS, Muller MJ, et al. The amsterdam preoperative anxiety and information scale (APAIS)[J]. *Anesth Analg*, 1996, 82(3): 445–451. DOI: [10.1097/0000539-199603000-00002](https://doi.org/10.1097/0000539-199603000-00002).
- 23 Salzmann S, Kikker L, Tosberg E, et al. Impact of a personalized intervention on preoperative anxiety and determination of the minimal clinically important difference in anxiety levels: a randomized clinical trial[J]. *Anesthesiology*, 2025, 142(4): 680–691. DOI: [10.1097/ah.0000000000005351](https://doi.org/10.1097/ah.0000000000005351).
- 24 Terluin B, Eekhout I, Terwee CB, et al. Minimal important change (MIC) based on a predictive modeling approach was more precise than MIC based on ROC analysis[J]. *J Clin Epidemiol*, 2015, 68(12): 1388–1396. DOI: [10.1016/j.jclinepi.2015.03.015](https://doi.org/10.1016/j.jclinepi.2015.03.015).
- 25 Terluin B, Eekhout I, Terwee CB. The anchor-based minimal important change, based on receiver operating characteristic analysis or predictive modeling, may need to be adjusted for the proportion of improved patients[J]. *J Clin Epidemiol*, 2017, 83: 90–100. DOI: [10.1016/j.jclinepi.2016.12.015](https://doi.org/10.1016/j.jclinepi.2016.12.015).
- 26 Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life[J]. *J Clin Epidemiol*, 1999, 52(9): 861–873. DOI: [10.1016/s0895-4356\(99\)00071-2](https://doi.org/10.1016/s0895-4356(99)00071-2).

收稿日期: 2025年12月13日 修回日期: 2026年03月31日
 本文编辑: 杨燕 沈静怡