



CNAS-GLOXX

**医学实验室 基于患者数据实时质量
控制方法建立及评估指南**

Medical laboratories-

**Guidance on the Development and Evaluation of
Patient-based Real-time Quality Control**

(征求意见稿)

中国合格评定国家认可委员会

前 言

本文件由中国合格评定国家认可委员会（CNAS）制定，是对 CNAS-CL02《医学实验室质量和能力认可准则》中“7.3.2 质量控制”要求涉及的医学实验室基于患者数据实时质量控制所做的具体解释和指导，也适用于即时检验（POCT）系统检测。供医学实验室和评审员参考使用。

本文件为首次发布。

医学实验室 基于患者数据实时质量控制方法

建立及评估指南

1 范围

本文件为医学实验室开展患者数据作为实时质量控制方法的建立、评估以及信息系统的应用提供指导，也适用于即时检验（POCT）系统检测。

本文件适用于申请认可或已获认可的医学实验室，也可供评审员在现场评审过程中参考使用。

2 规范性引用文件

下列文件对于本指南的应用是必不可少的。凡是注明日期的引用文件，仅该版本适用于本指南。凡是未注明日期的引用文件，其最新版本（包括所有的修改部分）适用于本指南。

GB/T 19000 质量管理体系—基础和术语

GB/T 20468 临床实验室定量测定室内质量控制指南

GB/T 29790 即时检验 质量和能力的要求

GB/T 36344 信息技术 数据质量评价指标

GB/T 41867 信息技术 人工智能 术语

GB/T 5271.31 信息技术 词汇 第 31 部分：人工智能 机器学习

3 术语和定义

GB/T 22576和ISO 15189界定的术语和定义适用于本文件。为方便使用列出以下术语和定义：

3.1 质量控制 quality control

质量管理的一部分，致力于满足质量要求。

[GB/T 19000-2016, 定义 3.3.7]

3.2 基于患者数据的实时质量控制 Patient-based real-time quality control; PBRTQC

使用患者样本实时计算的参数作为质量控制的一种形式。

3.3 数据 data

信息的可再解释的形式化表示，以适用于通信、解释或处理。

注：可以通过人工或自动手段处理数据。

[GB/T 33644-2018, 定义 2.1]

3.4 数据集 data set

具有一定主题，可以标识并可以被计算机化处理的数据集合。

[GB/T 33644-2018, 定义2.6]

3.5 被控变量 controlled variable

监测过程内被监测的研究对象。

3.6 运算程序 operation program

基于患者数据的实时质量控制方法所需的算法及其对应的参数。

注：在本指南中，术语“运算程序”的参数具有广泛的含义包括但不限于以下内容：

- a) 截断限 truncation limit; TL
- b) 控制限 control limit; CL
- c) 步长 block size

3.7 步长 block size

监测过程中固定长度的患者数量。

3.8 超参数 super parameter

算法过程中涉及的内部参数。

3.9 学习策略 learning strategy

学习技术的使用优先于其应用的一种规划。

[GB/T 5271.31-2006, 定义 31.01.05]

3.10 监督(式)学习 supervised learning

获得的知识的正确性通过来自外部知识源的反馈加以测试的学习策略。

[GB/T 5271.31-2006, 定义 31.03.08]

3.11 无监督学习 unsupervised learning

一种学习策略，它在于观察并分析不同的实体以及确定某些子集能分组到一定的类别里，而无需在获得的知识上通过来自外部知识源的反馈，以实现任何正确性测试。

[GB/T 5271.31-2006, 定义 31.01.09]

3.12 分类模型 classification model

〈机器学习〉一种对给定输入数据，输出其所属的一个或多个类别的机器学习模型。

[GB/T 41867-2022, 定义 3.2.6]

3.13 回归模型 regression model

以给定数值为输入，预期的输出为连续变量的机器学习模型。

[GB/T 41867-2022, 定义 3.2.8]

3.14 机器学习 machine learning

通过计算技术优化模型参数的过程，使模型的行为反映数据或经验。

[GB/T 41867-2022, 定义 3.2.10]

3.15 机器学习模型 machine learning model

一种基于输入数据或信息生成推理或预测的计算结构。

示例：如果一个单变量线性方程 ($y = \theta_0 + \theta_1 x$) 经由线性回归训练，则结果模型为 $y = 3 + 7x$ 。

注：一个机器学习模型是基于机器学习算法训练的结果。

[GB/T 41867-2022, 定义 3.2.11]

3.16 机器学习算法 machine learning algorithm

依据给定的准则, 根据数确定机器学习模型参数的算法。

示例：考虑求解一个单变量线性函数 $y = \theta_0 + \theta_1 x$, 其中 y 是输出或结果, x 是输入, θ_0 是截距($x = 0$ 时 y 的值), θ_1 是权重。在机器学习中, 确定线性函数的截距和权重的过程称为线性回归。

[GB/T 41867-2022, 定义 3.2.12]

3.17 数据标注 data labelling

给数据样本指定目标变量和赋值的过程。

[GB/T 41867-2022, 定义 3.2.29]

3.18 允许总误差 allowable total error, TEa

指分析过程需要达到的质量目标, 即单次测量或单次检测结果内不精密度(随机误差)和偏倚(系统误差)可允许的最大误差范围。

注 1: 在质量控制计划中, 假定无标本特异性影响, 因为标本特异性属方法整体性能的一部分, 非质量控制监测范围。

注 2: 有些出版物将允许总误差称为“ATE”。

[CLSI C24 4th ed, 术语]

3.19 系统测量误差 systematic measurement error

简称系统误差 (systematic error)

在重复测量中保持恒定不变或按可预见方式变化的测量误差的分量。

[JJF 1001-2011, 定义 5.4]

3.20 随机测量误差 random measurement error

简称随机误差 (random error)

在重复测量中按不可预见方式变化的测量误差的分量。

[JJF 1001-2011, 定义 5.6]

3.21 瞬态误差 transient error

随样本而来去或影响后续样本的误差。

注：在质量控制中，瞬态误差是发生在样本检测过程中的错误，通常是样本准备阶段及分析前产生的，如微小凝块、微小气泡、干扰物质等，这类错误由于在分析前、分析后缺乏识别标记，本质上是随机发生的。

3.22 特异性 specificity

在出现干扰现象（影响量）时，实验或检测程序能够正确地识别或定量确定某一实体物质的能力。

注：在质量控制中，是指当特定原因变异确实不存在时，质量控制系統能够指出特定原因变异不存在的概率。

[WS/T 415-2013, 术语 2.8]

3.23 敏感度 sensitivity

测量系统的示值变化除以相应的被测量值变化所得的商。

注 1: 在质量控制中，是指质量控制系統检测误差的能力。

注 2: 在定性实验中，是指实验方法获得阳性结果的能力。

[WS/T 415-2013, 术语 2.7]

3.24 分析中检查 IntraSpect Check

样本检测过程中对异常的传感器错误进行实时的监测。

4 概述

基于患者数据的实时质量控制方法是采用算法软件动态监测患者数据，实现对检测过程分析性能连续监控的质量控制方法。该类方法将针对仪器的质量

控制问题转化为针对数据的分析问题，按照算法原理可分为统计学方法和机器学习方法。其中，统计学方法是基于历史数据统计分析得出用于检测应用阶段的最优控制限；机器学习方法采用学习策略构建出用于检测应用阶段的机器学习模型。

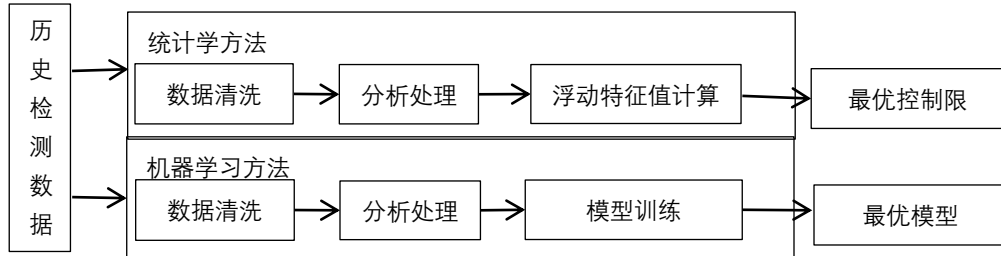


图 1 基于患者数据的质量控制算法构建流程

5 数据采集

5.1 数据选择

应选择相对稳定的患者人群的检测结果，排除分析系统改变以外的变异因素，包括但不限于：

- 1) 病、生理因素导致的个体生物学变异；
- 2) 待测物整体人群中的不同亚组, 如儿童、性别、孕产期等；
- 3) 患者类型的影响, 如一周内某天实验室检测透析室患者, 当天肌酐检测数量标本数增多, 导致该时段的平均肌酐浓度的检测值升高；
- 4) 对于某些测试，因临床解释不同导致的样本结果错误诊断和分类。

5.2 数据排除

对于同一检验项目，实验室存在如下情况数据时，应制定人工或自动手段清洗数据的规则，包括但不限于以下情况：

- 1) 检测的标本为非患者结果，如质控品结果及科研用途的标本结果；
- 2) 检测的标本类型不同，如肌酐项目的血清标本和尿液标本；
- 3) 多地点或场所使用的检测系统，如中心实验室、急诊实验室、发热门诊实验室；
- 4) 不符合运算程序要求。

注：基于患者数据的实时质量控制运算程序对数据要求不同，如delta check运算程序需要收集相同患者两次临近标本的检测结果。

5.3 采集方式

1) 实验室宜选择不少于一年患者历史数据，分年龄、性别、临床服务、医嘱申请科室、日期和时间绘制散点图，观察数据分布情况、单位时间内测试数量等，用于数据筛选；

2) 应保持患者数据的时间序列特征，并保证每个患者数据的可追溯性；

3) 患者的描述性信息可作为运算程序的数据参数，包括，患者唯一识别号、性别、年龄、申请科室、临床诊断、用药情况等。

6 数据处理

6.1 失控仿真^[1]

当真实场景中误差场景和类型覆盖范围不全时，实验室应充分评估可能存在的误差类型并进行数据仿真，如系统误差（包括比例误差和恒定误差）、随机误差等。

1) 系统误差按如下公式计算：

恒定误差：

$$x' = x + n \times TEa \times \bar{x} \dots \dots \dots (1)$$

注： x' 为引入恒定误差的检测结果； x 为质量在控的检验结果； n 为误差因子，取值为 $\pm 1/6$ 、 $\pm 1/4$ 、 $\pm 1/2$ 、 ± 1 、 ± 1.5 ； TEa 为项目的允许总误差； \bar{x} 为所有质量在控检验结果的平均值。

比例误差：

$$x' = x \times (1 + n \times TEa) \dots \dots \dots (2)$$

注： x' 为引入比例误差的检测结果； x 为质量在控的检验结果； n 为误差因子，取值为 -0.5 、 -0.48 、 -0.46 、 \dots 、 0.46 、 0.48 、 0.5 。

2) 随机误差按如下公式计算：

$$x' = x + e \dots \dots \dots (3)$$

注1： x' 为引入随机误差的检测结果； x 为质量在控的检验结果； e 为引入的噪音成分。

注2：随机误差分布类型可分为含高斯分布、 U 形分布、均匀分布三类。

高斯分布随机失控状态中的高斯分布误差采用高斯抽样方式生成，均值满足 ± 0.3 、 ± 0.5 、 0 倍 TEa ，标准差满足 $1/3$ 、 $1/2$ 、 1.5 、 3 倍的原检测结果的标准差；U形分布随机失控状态中的U形分布噪音采用高斯抽样方式生成，均值满足 ± 0.3 、 ± 0.5 、 0 倍 TEa ，标准差满足 $1/3$ 、 $1/2$ 、 1.5 、 3 倍的原检测结果的标准差；均匀分布随机失控状态中的均匀分布噪音是采用均匀抽样方式生成，上下限范围满足 ± 0.3 、 ± 0.5 、 0 倍 TEa 或 $1/3$ 、 $1/2$ 、 1.5 、 3 倍的原检测结果的标准差。

6.2 样本扩充

在建立方法阶段，当患者数据的数量不满足建立需求时，应通过数据衍生以扩充数据量。衍生时应基于数据特征进行仿造，加随机噪声的方式扩充患者数据集。

7 方法建立

实验室应根据患者数据分布特点及质量控制要求，进行方法的建立、优化及应用，并记录。

7.1 检测范围约束

实验室对患者数据清洗后，应约束数据的有效检测范围，如去除特殊科室、特殊时间段对应的人群样本。对所得患者数据进行过滤时，比例应控制在 5%以内。

7.2 数据拦截

实验室应对患者数据的离群值进行拦截处理。应根据不同方法设置截断限（truncation limit, TL）进行数据拦截，包括：

- 1) 统计学方法应至少比较两种拦截方式，如去除和替换离群值。

示例：采用Winsorization方法替换离群值。如公式（4）所示，其中 x 指处理前数值， x' 指处理后数值：

$$\begin{cases} x' = \begin{cases} LTL & \text{if } x < LTL \\ UTL & \text{if } x > UTL \end{cases}, & \text{替换} \\ x' = \begin{cases} NAN & \text{if } x < LTL \\ NAN & \text{if } x > UTL \end{cases}, & \text{去除} \end{cases} \dots\dots\dots (4)$$

2) 机器学习方法应至少比较两种聚类算法, 宜采用无监督(式)学习算法, 如孤立森林、K邻近等。

7.3 数据转态

应测试患者数据转正态的方式, 寻找符合检测项目的转态方法。

示例: Box-Cox转正态方式处理, 数据变换方式如公式(5)所示, 其中 x' 指处理前数值, x'' 指处理后数值:

$$x'' = \begin{cases} x' & , neat \\ \frac{x'^{\lambda}-1}{\lambda}, \lambda \neq 0 & , box - cox \\ \ln x', \lambda = 0 \end{cases} \dots\dots\dots (5)$$

7.4 样本建立

应根据不同算法的原理, 提取患者数据特征并转化为建立方法所需的质量控制样本。

7.4.1 统计学方法

提取患者数据特征值的算法应包括但不限于以下种类:

- 1) 固定数量患者数据的平均值(MA);
- 2) 在1)的基础上引入加权平均值(EWMA);
- 3) 固定数量患者数据的中位数(MM);
- 4) 固定数量患者数据的统计中位数(HD50);
- 5) 固定数量患者数据的统计标准差(MovSD);
- 6) 固定数量患者数据的超出临床决定限的特殊数据个数(MovSO)。

建立方法的过程应至少包含3种特征提取算法, 计算出的特征值相当于质量控制变量作为后续操作的样本素材。

7.4.2 机器学习方法

对患者数据应进行预处理以满足建立模型需求, 包括但不限于以下步骤: 数据标准化、按固定长度组织队列, 将处理后的数据做为后续操作的样本素材, 划分为训练集、验证集和测试集。

7.5 运算程序的建立

统计学方法程序建立应从样本中得出数值范围做为质量控制的控制限。机器学习方法程序建立应包括两种场景: 采用回归算法用于学习样本的控制限, 或者采用分类算法直接用于判断质量控制结果。实验室应对每个检验项目的运

算程序进行参数设置，并定期评估调整。

7.5.1 控制限选取方法

实验室应根据被控变量的分布情况选择适宜的控制限，用于判断在控和失控状态。不同的方法控制限的性质和单位不同，如，统计学方法控制限的性质和单位取决于 7.4.1 提取特征对应的算法；机器学习的分类模型控制限的性质为预测概率。

其中，统计学方法的控制限计算方法包括但不限于：

1) 对称统计法：控制限下限 (lower control limit, LCL) 和控制限上限 (upper control limit, UCL) 设置公式为： $\bar{x}_p \pm n \times s_p$ ，其中 \bar{x}_p 、 s_p 为被控变量的均值和标准差， n 为倍数；

2) 所有日统计法：如适用，采用模拟多个实验日方式时，以所有实验日被控变量的 0.5% 和 99.5% 分位数做为控制限的下限和上限；

3) 日极值统计法：如适用，采用模拟多个实验日方式时，统计每个模拟实验日被控变量的最小值，在所有模拟实验日的最小值中取 5% 分位数做为控制限下限；同样的方式，取所有模拟实验日患者数据最大值的 95% 分位数做为控制限上限。

7.5.2 机器学习模型

如适用，应根据程序的目标进行相应的样本标注以完成监督（式）学习。若采用回归模型，应将 7.5.1 得出的控制限做为被预测值；若采用分类模型，应将 6.1 或 6.2 得出的失控数据和在控数据按 7.4.2 形成样本，将数据质量状态做为分类标签。

训练模型所用算法应包括但不限于：逻辑回归 (Logistic Regression)、决策树 (Decision Tree)、支持向量机 (Support Vector Machine, SVM)、K 最近邻 (K-Nearest Neighbors, K-NN)、随机森林 (Random Forest)、神经网络 (Neural Network)、贝叶斯分类器 (Naive Bayes Classifier)

7.6 运算程序的优化与验证

实验室应结合应用需求制定适宜的优化过程及参数设置规则，包括：

1) 如适用，应对各参数给出合理取值范围，在所有参数的组合下进行遍历实验，按照合理的寻优法则选出最佳参数。适用于或者针对参数设立目标函数，

如退火算法、遗传算法等优化算法选择最佳参数。

2) 采用机器学习方法时, 应根据选择算法特点进行模型调优。可调节的因素有: 步长长度、过滤数据比例, 分类算法的超参数, 调节过程方法包括但不限于网格搜索或经验法, 调节期间通过验证集观察, 验证方法包括但不限于交叉验证。

8 运算程序的评估内容和方法

8.1 评估内容

8.1.1 临床评估

8.1.1.1 评估指标

1) 混淆矩阵相关指标包括: 准确度 (ACC)、真阳率 (TPR)、真阴率 (TNR)、假阳率 (FPR)、假阴率 (FNR);

2) 临床验证的指标包括: 从引入误差到报警所经过的患者样本数量 (Nped), 所有测试天数中Nped平均数 (ANped)、所有测试天数中, Nped中位数 (MNped)、所有测试天数中, Nped95分位数 (Nped_95)、不稳定指标 (I)。

注: $I = \text{Nped的四分位数间距} / \text{MNped}$, I值表征一批数据的离散程度, 可反应所有测试天的NPed结果的稳定性, I值越大说明算法越不稳定。

8.1.1.2 具体方案

1) 实验室应根据检验项目的预期用途和性能要求, 制定适宜的质量控制合格标准, 评估的数据至少达到10000个。

2) 实验室制定评估标准时, 应参考制造商或研发者声明、相关标准、公开发表的临床应用指南和专家共识等。

3) 方法临床评估结果应满足误差检出概率达95%以上, 且假失控概率/假阳性率在1%以下, 且评估结果应结合质控图进行量化展示, 如误差检出图、柱状图等。

4) 采用机器学习方法时, 应增加对算法模型性能的评估。

8.1.2 计量学评估

采用计量学评估应至少包括以下方法一种:

1) 将基于患者数据的质量控制算法软件的计算结果溯源至数字量具;

2) 将基于患者数据的质量控制算法软件的计算结果溯源至算法量具。

通过上述方法，将算法模型的输出结果溯源至物理世界的实际量值，并通过一条具有测量不确定度、不间断的计量溯源链，最终溯源到国家基标准，实现输出结果准确性的计量学评价，并使同类算法软件的输出结果具备一定的可比性。

8.1.3 前瞻性临床评估

收集临床出现的质量失控的数据，应将数字孪生后进行质量控制方法性能评估，至少半年进行一次评估，孪生前质量失控数据应至少达到500条。

9 评估方法

9.1 实施前评估

实验室在正式实施自行开发/引进具有基于患者数据实时质量控制运算程序的软件之前，应对程序的功能、参数、规则进行评估，确保符合实验室的要求。评估的内容应包含但不限于：

- 1) 应对运算程序的所有功能、规则及参数进行评估，保证其性能符合实验室的要求；
- 2) 应逐一评估运算程序所涉及的检测项目和样品类型；
- 3) 应评估运算程序识别患者数据出现报警的具体位置；
- 4) 评估时间宜不少于3个月和/或患者数据不少于50000条。

9.2 必要时评估

在使用运算程序过程中，若发生仪器设备更新、运算程序参数变更、信息系统升级等可能影响其功能的改变都应对其进行评估，确保符合要求后方可继续使用。评估的内容、数据量可由实验室根据变更内容确定。

9.3 定期评估

由指定人员对运算程序进行复核。至少半年进行一次评估，评估数据量不少于 10000 份。

9.4 评估参数

9.4.1 假性失控概率 (FPR, false positive rate)

采用质量在控数据输入算法进行计算。

$$FPR = \frac{1}{n} \sum_{i=1}^n x_i \times 100\% \quad \dots\dots\dots (6)$$

式中：

n ——在控数据个数；

x_i ——第 i 个实际判定结果，取值为 0（判定在控状态）或者 1（判定失控状态）。

9.4.2 误差检出概率（TPR, true positive rate）

采用质量失控数据输入算法进行计算。

$$TPR = \frac{1}{n} \sum_{i=1}^n x_i \times 100\% \quad \dots\dots\dots (7)$$

式中：

n ——失控数据个数；

x_i ——第 i 个实际判定结果，取值为 0（判定在控状态）或者 1（判定失控状态）。

9.4.3 误差检出前影响的患者标本数

从误差引入到检出所影响的患者样本数（NPed）、NPed的均值（ANPed）、NPed的中位数（MNPed）、95分位数NPed（95NPed）的结果及不稳定指数（I）。

10 信息化功能要求

医学实验室可使用实验室信息系统或者中间件实施基于患者数据的实时质量控制。实验室可自行设置或采用已具备本方法程序的信息系统实施基于患者数据的实时质量控制。

10.1 实验室信息系统要求

实验室信息系统应具备信息收集、数据提取、数据导出等功能，至少包含以下信息：

- 1) 患者识别号；
- 2) 患者性别；
- 3) 患者出生日期/年龄（至少小数点后一位）
- 4) 实验室检测日期及时间（小时、分钟、秒）；
- 5) 运算程序的参数；
- 6) 实验室标本类型；
- 7) 实验室仪器检测非患者材料（外部QC、透析液、动物研究样本、研究样

本等)；

- 8) 标本收集条件(如, 定时收集的标本)；
- 9) 标本申请科室；
- 10) 临床申请内容/类型；
- 11) 实验室检测结果；
- 12) 使用的试剂批号；
- 13) 检测状态标签(如, 仪器、溶血、黄疸、血脂指数等)；
- 14) 初次检测或复测标本；
- 15) 检测项目的单位；
- 16) 患者历史结果。

10.2 使用要求

- 1) 应有简单易操作的用户界面, 方便操作人员选择、提取运算程序的参数；
- 2) 应具备数据变换的计算功能, 如自然对数、平方根、Box cox、幂变换, 以及各种比率的计算等。
- 3) 应具备一定统计分析功能, 允许用户计算简单的统计数据, 如平均值、中位数、百分位数及标准差等；
- 4) 应具备参数计算功能, 如数据截断、参与计算的患者标本数等；
- 5) 应具备数据模拟仿真功能；
- 6) 应提供几种可选不同工作原理算法, 如移动均值、指数加权移动均值、移动中位数和机器学习算法等；
- 7) 应有明确定义的实时警报形式, 并可在故障排除后进行确认和重置。在检查警报后, 应该能够重置算法并排除特定患者；
- 8) 应记录确认报警人的姓名和时间, 且用户可在质量控制图上进行报警注释, 并生成报告；
- 9) 应可随时进行质量控制运算而不触发失控报警；
- 10) 应具备数据可视化功能, 可进行质控图展示, 包括但不限于Levey-Jenney图、Z分数图、MNped曲线图(图2), 以及柱状图等。

示例: MNped曲线图

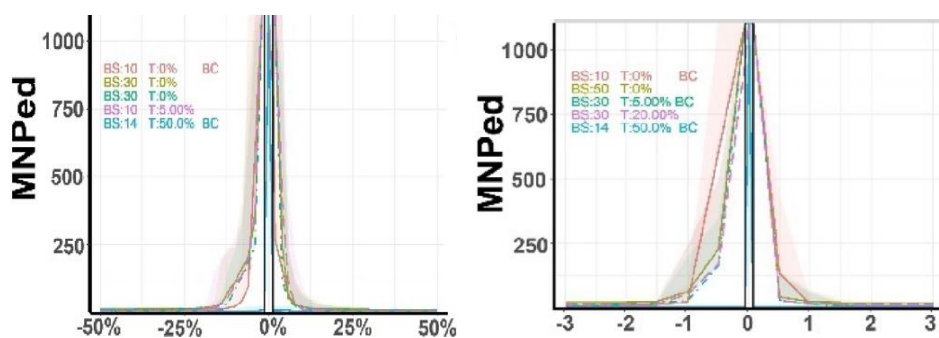


图2 MNped图

10.3 数据安全

实验室应保证数据存储、传输和备份的完整性：

- 1) 信息系统应保证实验室数据存储介质的存放、运输等物理安全；
- 2) 数据获取、传输和使用应符合安全及患者隐私的法规要求；
- 3) 如需将数据传输到第三方应用程序，应对数据集进行去标识处理。

10.4 设施和环境条件：

实验室宜选择在 Windows 或 Linux 下运行，运算工具包括但不限于：Python3.7.3 “numpy”，” pandas”，“sklearn” 等工具包。

10.5 患者数据采集设施：

在实施室间质量评价时，应具备以下功能：

- 1) 数据过滤：选择门诊患者数据，实验室可自定义步长、时间窗口长度、过滤周末及节假日患者数据功能。
- 2) 算法考量：宜采用统计学方法更为稳定的 MM 算法，可计算每日中位数或浮动中位数。
- 3) 分组原则：应按照专业、项目设计分组方法，如：血细胞分析项目可按仪器分组，生化项目可按方法学分组。对实验室数少或仪器数量少的组，可缺省组处理；或通过散点图区别显示方式，若实验室所使用的仪器有别于其他实验室，散点图中对该实验室数据颜色加以区分，该实验室数据仅用于监测,不参均值和控制限计算。
- 4) 控制限设置：应根据方法原则上遵循：基于患者数据的室间质量评价的控制限宜小于该项目的室间质评标准，而大于室内质控标准。可选择基于生物学变异的偏移标准设置控制限，两者取较大值。在实际操作中可根据项目的稳定性进行适当调整。

5) 提供信息：实验室可依据单个实验室、同实验室组或所有仪器组提供长期中位数、长效精密度、同组偏移或所有设备靶值等信息。

附录 A (资料性)

基于患者数据实时质量控制算法中英文对照

基于患者数据实时质量控制算法中英文对照见表 A.1。

表 A.1 基于患者数据实时质量控制算法中英文对照表

英文简称	英文全称	中文全称
AON	Average of normal	正态平均值法
MA	Moving average	移动均值法
EWMA	Exponentially weighted moving average	指数加权移动均值法
MM	Moving median	移动中位数法
HD50	Harrel-Davis 50 percentile estimator	HD 统计中数法
MovSD	Moving standard deviation	移动标准差法
MovSO	Moving sum of outliers	移动异常值法
MovSum	Moving sum of number of positive patient result	移动阳性患者结果总和
EAMM	Exponentially Adjusted Moving Mean	指数调整移动平均值法
MQ	Moving quartile	移动四分位数法
DC	Delta check	差值校验法
MP	Moving percentiles	移动百分位数法
RF	Random forest	随机森林
LR	Logistic Regression	逻辑回归
SVM	Support Vector Machine	支持向量机
KNN	K-Nearest Neighbor	k-最近邻
NBC	Naive Bayesian Classifier	朴素贝叶斯分类器

附录 B (资料性)

基于患者数据实时质量控制统计学方法建立示例

B.1 数据拦截与转态

通过医院实验室部实验室信息系统选取了 11 万血小板 (PLT)。首先对正常数据进行拦截, 拦截限定义为正常数据的阈值边界对应的分位数 t , $t = \{0, 0.01, 0.05, 0.2, 0.4\}$, 拦截限上限为: 数据的 $(1-t)$ 分位数; 拦截限下限为: 数据的 t 分位数。

拦截异常点的方式分替换与去除两种形式。数据拦截后, 进一步的, 转态方式分原型和 BOXCOX 转态两种形式, 原型为 neat 型, 指数数据按原级处理。

B.2 数据仿真

进一步的, 将样本分为 200 天, 每天 1150 个样本, 前 150 为正常样本, 后 1000 引入误差做为异常样本。对误差数据进行模拟, 将误差定义为比例误差、恒定误差随机误差。

B.3 遍历实验

步方式长按照 $\{10, 30, 50, 90, 110, 130, 150\}$ 的进行搜索, 所有实验操作数量见表 B.1。

表 B.1 遍历实验次数表

算法	过滤方式	过滤限	转正态方式	步长	控制限算法
ewma	2	5	2	7	3
ma	2	5	2	7	3
mm	2	5	-	7	3
hd50	-	-	2	7	3
movSD	2	5	2	7	3
movS0	-	-	-	7	3

进一步的, 将所有实验结果统计见表 B.2, 其中所有引入的误差因子做为表头, 排列组合形式做为首列, 统计实验结果的指标: 假阳性、真阳性、真阴性、假阴性、累计 ANPed (Nped 的均值)、累计 95NPed (Nped 的 95%分位数)、MNPed (Nped 的中位数)。进一步的, 在正常样本的均值数列下计算控制限。按照当前的控制限对每日的 1150 个检测样本进行实验, 其中 1000 个异常样本向前选取步长宽度的样本做为一个样本的均值计算基础, 当检测到超出控制限即算作报警, 当前对应的序号为 Nped, 如果 1000 个样本均为检出, 计做 1100。

B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH		
控制方式	控制限	步长	控制限方法	cl_lower	cl_upper	假阳率	ANped_24	MNped_24	95NPed_24	95NPed_24	95NPed_24	95NPed_24	95NPed_24	95NPed_24	95NPed_24	95NPed_24	95NPed_24	95NPed_24	95NPed_24	95NPed_24	95NPed_24	95NPed_24	95NPed_24	95NPed_24	95NPed_24	95NPed_24	95NPed_24	95NPed_24	95NPed_24	95NPed_24	95NPed_24	95NPed_24	95NPed_24	
替换高群点	0	10	symaetric	11.99605	22.78917	0.04475	10	11	16	0.45455	16.4	21	25.8	0.85714	16	19	25.8	0.94737	1238.2	1119	2527.6	25.67261												
替换高群点	0	10	all FERTOC	13.1	22	0.10031	8.2	10	12.6	0.4	10	3	24.6	5.66667	9.8	3	24.4	5.33333	381.8	204	932.8	58.93109												
替换高群点	0	10	daily extremes	9.24	26.56	0.00999	146	137	294.6	0.27007	77.8	38	194.8	2.37143	61.8	32	178.8	0.378	13628.6	15656	18648.8	15.94239												
替换高群点	0	30	symaetric	14.26483	20.52046	0.04462	16.8	14	23.2	0.5	21.8	25	29.8	0.48	21	23	28.8	0.47826	1953.2	1545	3705.6	16.75407												
替换高群点	0	30	all FERTOC	14.9	20.03333	0.09677	14.8	11	22.6	0.54545	20.8	23	28.8	0.47826	20	23	26.8	0.3913	937.2	787	1688.8	15.13483												
替换高群点	0	30	daily extremes	11.06667	22.44	0.00159	77.6	62	133.6	0.74194	40	41	54	0.29268	38.2	40	50.6	0.35	11219.6	12548	16188.8	15.23513												
替换高群点	0	50	symaetric	14.96742	19.81779	0.04565	24.6	26	34.8	0.42308	25.4	29	30.8	0.06897	24	27	30.8	0.22222	3048.6	3273	4883.2	9.50078												
替换高群点	0	50	all FERTOC	15.44	19.42	0.09795	17.4	18	25	0.72222	23	27	30	0.33333	22.4	26	29.8	0.34615	1455.8	1334	2275	11.25269												
替换高群点	0	50	daily extremes	13.916	21.588	0.00155	42.4	41	49.2	0.14634	79.4	91	110.4	0.51649	53	49	64.6	0.28571	9310.6	11183	13607.6	10.34569												
替换高群点	0	90	symaetric	15.87837	19.20591	0.04883	35.4	39	48	0.23977	33.2	31	43.4	0.22881	31	30	42.2	0.33333	3814.8	3201	5734.8	10.95538												
替换高群点	0	90	all FERTOC	15.92222	18.91111	0.0985	29	33	43.2	0.51515	26.2	25	37.4	0.44	25	25	34.4	0.32	1961.2	1805	3536.8	12.7607												
替换高群点	0	90	daily extremes	14.99778	19.92444	0.0067	50.6	48	60.4	0.25	56	57	68.6	0.21053	52.2	51	63.2	0.21569	6979.2	7282	9818.2	8.2184												
替换高群点	0	110	symaetric	15.78314	19.02234	0.04689	33.6	40	61.4	0.525	48.6	51	61.2	0.29412	42.6	45	50	0.31111	3876.2	4165	5892.2	9.36355												
替换高群点	0	110	all FERTOC	16.98182	18.78455	0.09946	34.4	33	57.8	0.42484	30.2	30	45.2	0.5	27	28	41.2	0.10714	2319.4	2366	4894.4	14.36979												
替换高群点	0	110	daily extremes	15.28394	20.00545	0.00507	50.4	54	68.2	0.2963	68.6	71	89.8	0.3662	66.4	69	85.6	0.37681	8234.6	9008	10722.4	7.14676												
替换高群点	0	130	symaetric	15.90551	18.88013	0.04693	44.2	43	67	0.67442	53.4	60	64.2	0.25	49.6	50	60	0.28	4425.2	4307	6791.8	12.95475												
替换高群点	0	130	all FERTOC	18.19231	18.03846	0.09644	31.8	37	45.8	0.18191	21.8	38	46.8	0.44444	28.8	31	42.6	0.28696	2739.6	3023	4788	16.3609												
替换高群点	0	130	daily extremes	15.59231	19.91846	0.00755	55.2	52	78.6	0.46154	78.4	83	98.2	0.38554	75.2	77	94.6	0.33766	8173.6	8520	10555.4	8.13043												
替换高群点	0	150	symaetric	16.01505	18.7708	0.04907	45	52	65.8	0.42308	52.4	48	68	0.375	51	48	65	0.33333	3904.2	4128	6764.8	11.11326												
替换高群点	0	150	all FERTOC	16.29333	18.54	0.09994	36.6	38	59	0.31579	22.2	28	30	0.35714	21.6	28	35.8	0.28571	2794	2860	4994.8	15.95473												

表 B. 2 实验结果统计表

B. 4 结果寻优

在结果寻优步骤中，综合统计结果表的各项指标，按照一定优先级选择最佳的一种结果作为该网格搜索实验的最优解。

- 1) 将假阳列排序，保留假阳率<5%的部分；
- 2) 在 1) 的条件下，保留识别率（真阳率）≥90%的部分；
- 3) 将累计 MNped 排序，选择最小值，筛选结果不唯一时，进一步地，比较 ANPed 和 95NPed 选最小值。

原始数据寻优结果如图 B. 1 和图 B. 2 所示：

```
[{'项目': 'TP'}, {'算法': 'ma'}, {'TL_lower': 63.75}, {'TL_upper': 88.95}, {'转
正态方式': 'neat'}, {'转态λ': '-'}, {'滑动长度': 130}, {'控制限方法': '对称S
D'}, {'CL_lower': 73.57049}, {'CL_upper': 79.13615}]
```

图 B. 1 代码寻优结果输出截图

由此确定质控步长是 130；控制下限为 73.57049；控制上限为 79.13615。

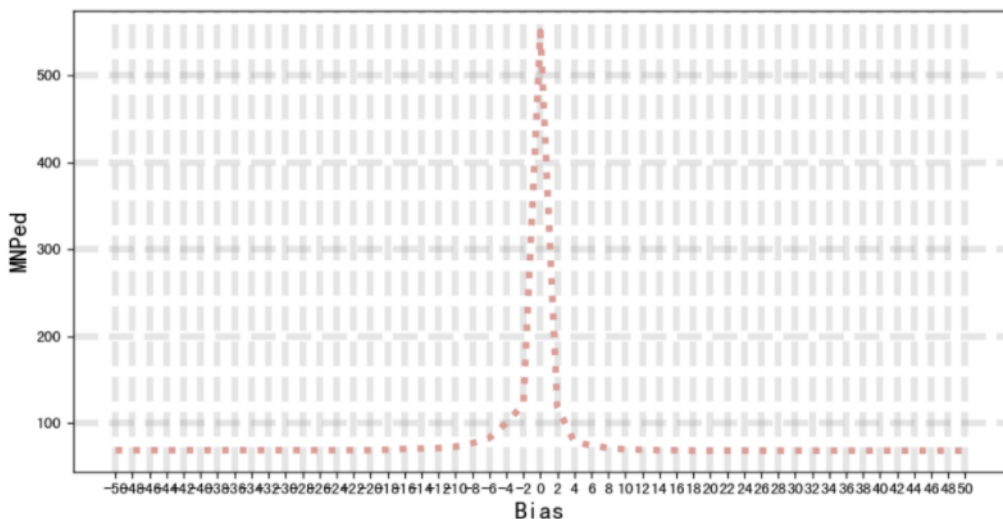


图 B. 2 MNped 曲线

附录 C (资料性)

基于患者数据实时质量控制机器学习技术方法建立示例

C.1 数据准备

以 ALB 为例，针对专向科室、人群类别。按照数据采集、数据处理、算法选型、训练调优、模型评估步骤进行。

采集 1 095 000 个经过临床验证的在控患者连续在控报告。同时，收集相关患者信息，包括年龄、性别、科室、样本类型、测量时间点。只采集了血清样本。所有包括患者数据均处于对照组质量控制状态。历史数据要求。

由于原始样本均认定为无偏下的数据，按照公式仿真偏差数据： $x'=x \times (1+n)$ 。其中 x 表示包含人工误差的患者数据， x 是原始患者数据， \bar{x} 表示数据集的平均值。 σ 表示数据集的标准偏差。 n 是用于控制所引起的误差大小的因子， $n=-50\%$ ， -48% ， -46% ，..... 46% ， 48% ， 50% 。

C.2 特征工程设计

控制 $<20\%$ 分析误差的阈值进行数据过滤，以消除 $\geq 20\%$ 分析误差的荒谬或极值。该滤波算法是针对模型训练中采用的两种分类算法而定制的。

其中数据过滤利用预设聚类模型，根据异常控制比例对历史检测数据进行聚类处理，以确定聚类结果，控制比在 5%–50%范围中遍历；根据聚类结果对预设的初始分类模型进行训练，以确定质控分类模型；通过质控分类模型对测试数据进行误差识别测试，并根据测试结果对应的假阳率判断是否满足循环结束条件，假阳率用于表征被误判数据占测试数据的比例，5%以内算达标，被误判数据包含在误差识别测试的被剔除数据集中；若否，则根据预设调整规则调整异常控制比例的取值，并重新根据调整后的异常控制比例对历史检测数据进行聚类处理，重新训练质控分类模型，直至假阳率满足循环结束条件。

C.3 预训练

将无偏和加偏分别按相同的步长组队，一个步长内的一批数据为一个整体，形成针对机器学习分类模型的样本，无偏数据下的样本标记为“0”类，加偏数据下的样本标记为“1”类。对样本进行训练集、验证集、测试集划分，用训练集进行模型训练。

根据临床情况和数据分布，这些误差被分为两类： $<20\%$ 和 $\geq 20\%$ 的分析误差。训练组和实验组的比率为 8：2。

以 RF 算法为例，针对 5.4%临界比例误差水平下的 AST 分析物，给出调优结果。关键参数的最佳组合为：树数 200，最大深度 200，详见表 C.1。此外，数据块的大小也被定义为 10。

表 C.1 5.4%比例误差分析物 RF 算法优化结果

树数	最大深度	训练精度	测试精度
100	100	0.94	0.93
200	100	0.96	0.95
200	200	0.97	0.96
300	100	0.97	0.96
300	200	0.99	0.98
400	100	0.98	0.97
400	200	0.98	0.97
500	100	0.98	0.97
500	200	0.97	0.96

C.4 模型训练

对于 $<20\%$ 的系统误差和随机误差，由于数据分布的特点，对分割模型进行了训练。将 $<20\%$ 系统误差训练模型设置为“20、16、12、12、8、4、异常 1”，随机误差“10、12、8、4、1”。(模型优化)

模型优化的目的是提高模型对当前数据集的预测能力。如果检测 $\geq 20\%$ 的误差，预设算法本身的关键参数，如树的最大深度和树的数量(n 树)，调整块大小和值阈值的参数。如果检测 $<20\%$ 错误是预设块大小和值阈值，则调整树最大深度和 n 个树上的参数。

以 5.4%临界比例误差水平的 AST 分析物为例，在原始数据预处理后，以 10 步从 50 到 500 测量树最大深度和 n 树，并搜索了两个参数的最佳组合。分别计算了每个组合下的训练集和测试集的准确性。此外，用步骤 1 测量 5 到 100 的另一个参数。

可调节的因素有：步长长度、过滤数据比例，分类算法的超参数，调节过程方法包括但不限于网格搜索或经验法，调节期间通过验证集观察，验证方法包括但不限于交叉验证。

C.5 模型评估

表 C.2 RF 算法不同类型关键误差的预测得分

错误类型	项目	假阳 (%)	AUC	精确性	灵敏度	特异性	不稳定指标
比例	ALT	1.89	0.9889	0.9417	0.9357	0.9811	0.29

C.6 判断规则

为了提高临床场景中预测的准确性，设计了两级输出窗口。设计一级窗口输出模型预测值，然后基于统计逻辑的二级窗口，以降低一级窗口输出值的误率。二级窗口的逻辑设置为只有当大于预定义的截止值时，二级窗口中的异常值才确定为异常。二级窗口可以修改并输出最终正确的预测结果。

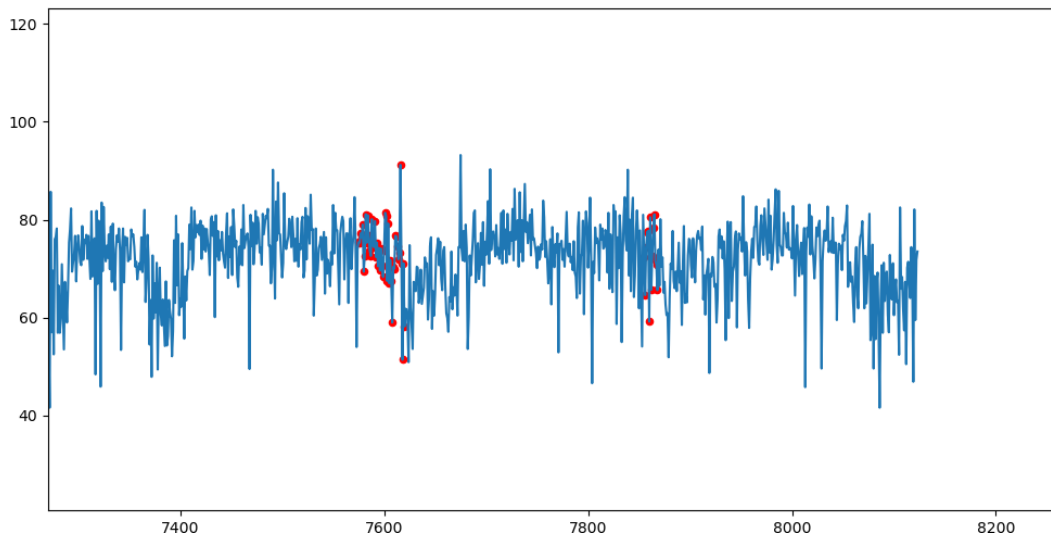


图 C.1 检测验证实验

附录 D

(资料性)

即时检验血气分析基于患者数据实时质量控制方法示例

D.1 基于患者数据的实时质量控制检查

以自动化智能质量管理模式为例，建立以患者样本为质控的处理系统机制，在样本检测读数阶段，自动化监测影响检测结果质量的瞬态错误。检查模型如下：

- 使用患者样本作为质控；
- 在样本检测期间对每个检测项目进行测定和分析；
- 根据相关的算法处理，自动检测任何异常响应；
- 若发生失控，应拒绝报告受错误影响的参数，并重新测定。

D.2 检查示例

以血气检验中 K+项目为例，在样本检测期间，检查收集一定时间内的样本的读数，通过异常传感器响应模式斜率形状和系数来监测整个检测过程中持续的样本完整性质量检查，以保证患者结果的准确度。样本检查流程如下图 D.1 所示，PBRTQC 检查如图 D.2 所示。

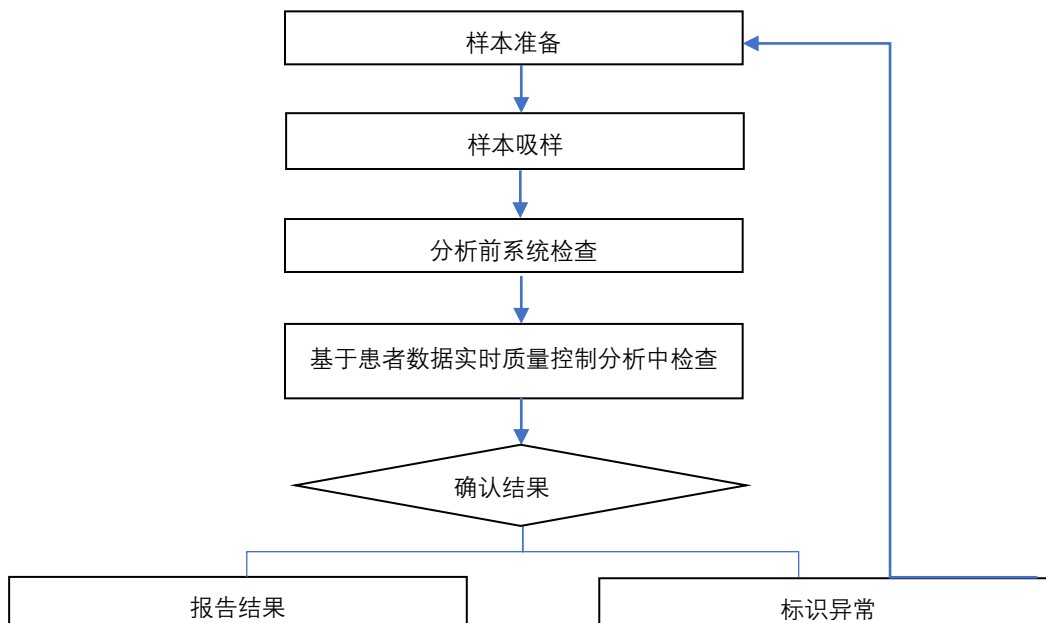
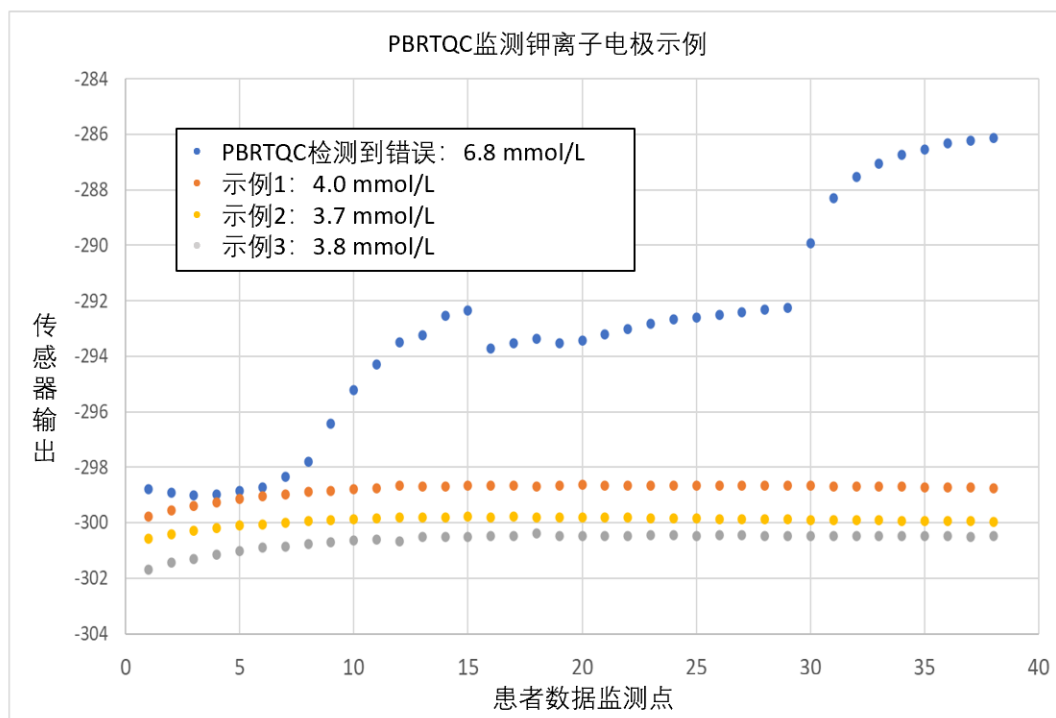


图 D.1 样本检查流程图



图D. 2 PBRTQC检查示例图

附录 E

(资料性)

基于患者数据实时质量控制方法评价可视化示例

E.1 异常数据模拟

对模拟异常数据的检测是评价质量控制方法效能的一个很重要的过程，此过程需要在方法的假阳率已满足要求的前提下进行。针对此过程有三种主要的展示图用于递进地表示方法的效能，包括 no_NPed 图, ANPed 图和 PED 图。

E.2 方法评价可视化

以红细胞项目质量控制为例展示三种 PBRTQC 评价可视化。

在模拟测试中，针对不同大小的偏差分别生成 10 个虚拟天，每天引入 1000 个连续的模拟异常数据，质量控制方法需要在每个虚拟天中以一定数量为单位依次检测样本，判断机器是否出现异常。若正确识别异常触发报警，则将异常发生点到首次报警点所经过的样本数记为当天的 NPed 值；将 1000 个异常数量中所有正确的报警次数的占比记作当天的 TPR；若 1000 个异常数据均未识别，则将当天的 NPed 值记为 1100；对于每个偏差，统计 10 天中 NPed 为 1100 的数目记作该偏差的 no_NPed。对模型测试结果的展示有如下三级别：（绿线代表引入误差水平为 6%，红线代表引入误差水平 12%，需要根据不同项目进行对应设置最小可觉误差线，本图 6%仅用于参考。）

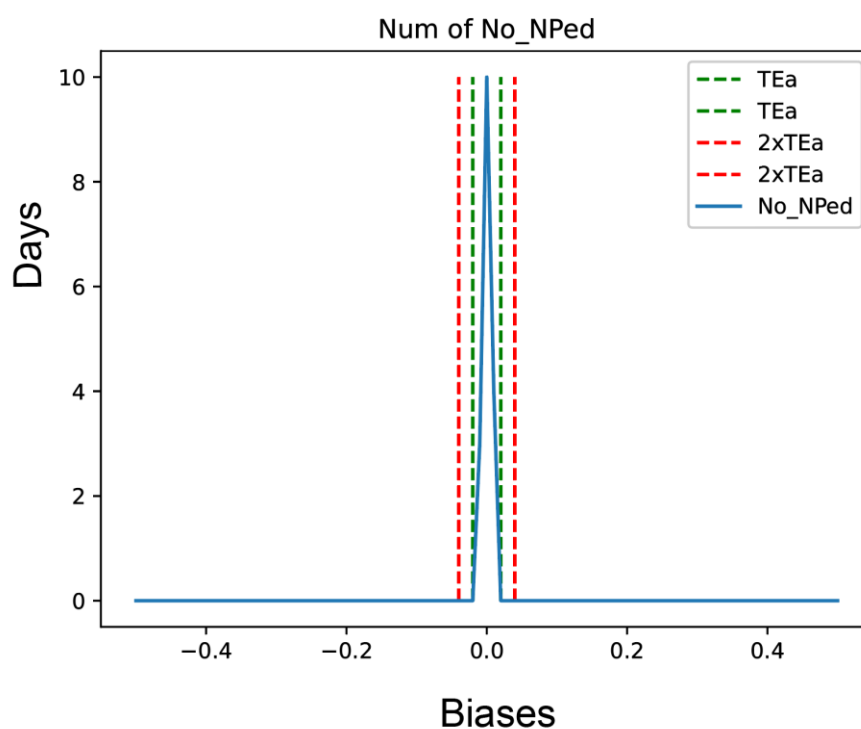


图 E.1 no_NPed 示例图

第一级别：首先通过观察 no_NPed 图，该图展示各偏差下未正确识别到异常的虚拟天数。横轴表示不同大小的偏差因子，越往两边说明加入的噪声比例越大。这个指标图表明，除了 0 噪声附近指标值不为 0 以外，其他比例大小的噪声对应的指标值均为 0，说明模型对各种比例，甚至很小比例的噪声，在设定的虚拟天内都能检测出异常。这个指标表明了模型的基础性能，既在机器发生异常的大多数虚拟天中模型最终都能够发现机器异常，模型应该尽可能将峰值约束在可觉误差线内。

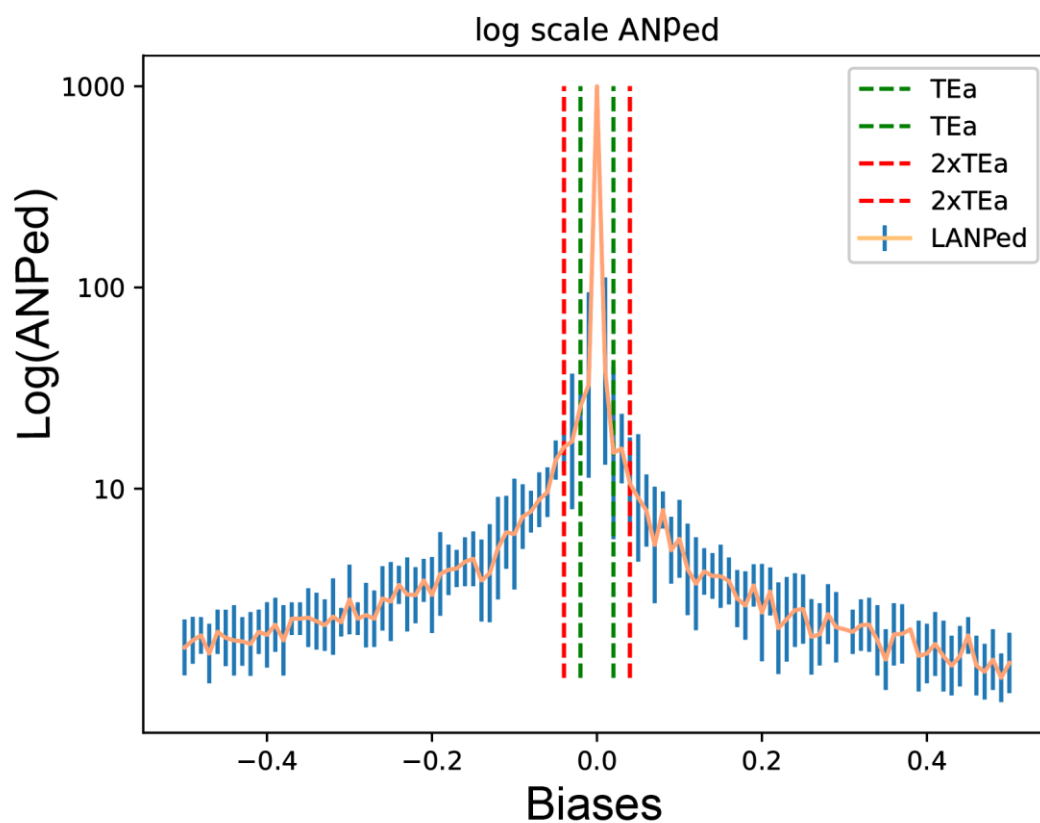


图 E.2 ANPED 示例图

第二级别：ANPed 则是计算每个偏差下所有虚拟天的 Nped 值的平均值，横轴表示不同大小的偏差因子。从图中可以看出除了 0 比例未检出异常外，其他噪声水平对应的 Anped 值均处于较低的水平，说明模型对各个噪声水平均能快速检测出异常。这个值越小说明模型灵敏性越高，能更快速准确的检测出异常，波峰越集中于中间，拐点越低且位于误差水平线越内部说明表现效果越好。ANPed 曲线无法展现每个偏差下所有虚拟日的 NPed 的离散程度，需进一步结合不稳定指标 I 值评估其稳定性， $I = \text{NPed 的四分位数间距} / \text{MNPed}$ 。

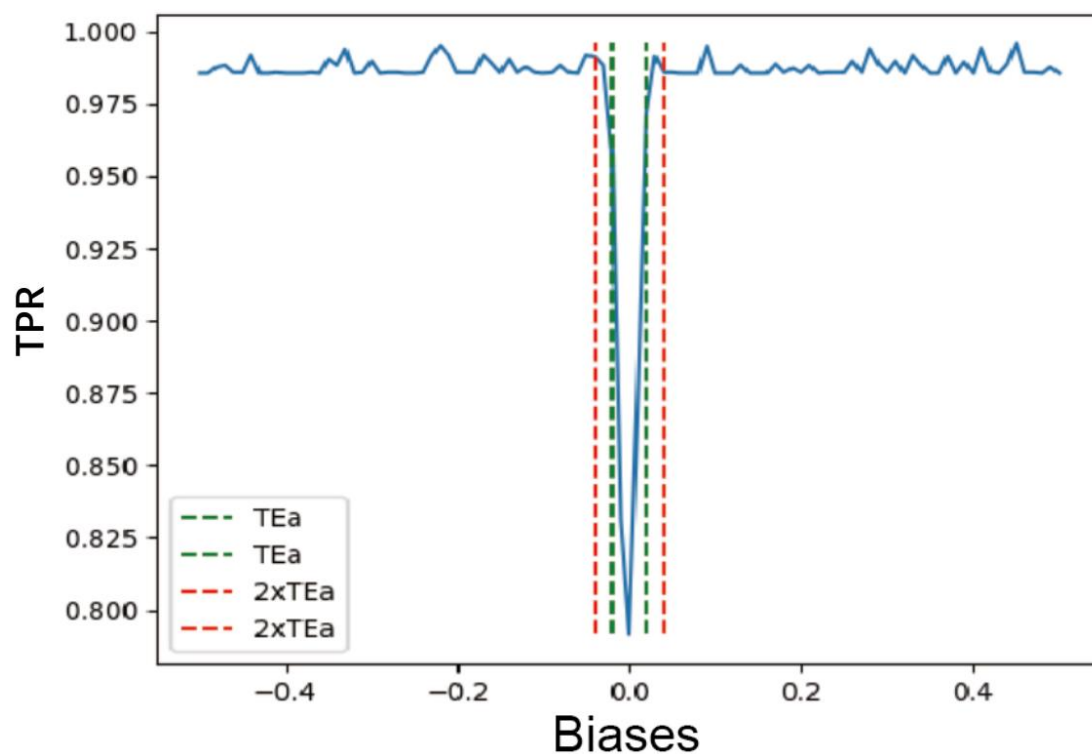


图 E.3 TPR 示例图

第三级别：由于 ANped 指标只表征报警时效性，所以最后还需结合 TPR 曲线评估方法的误差检出准度，从图中可以看出，除了 0 噪声水平附近明显下降以外，在两边其他各噪声水平下误差检出率均在 95%以上，说明模型对于各类样本在大量数据的测试下有很高的准度。TPR 曲线是方法对所有数据展现出来的宏观表现，波谷越窄（忽略 0 点处数据），两侧越高，说明方法效果越好。

参考文献

- [1] Duan X , Wang B , Zhu J , et al. Regression-Adjusted Real-Time Quality Control[J]. Clinical Chemistry, 2021.