**检验医学定量检测项目基于实验室检测结果的**

**实时质量控制算法溯源方法（征求意见稿）**

**规范编制说明**

一、任务来源和背景

（一）任务来源

国家计量技术规范《患者白蛋白实验室检测数据机器学习质量控制算法溯源》、《患者谷草转氨酶实验室检测数据机器学习质量控制算法溯源》是由国家市场监督管理局提出并以市监计量函[2021]2062号批准的项目。该标准由全国法制计量管理计量技术委员会归口单位归口，中国计量科学研究院负责牵头起草。

本标准的制定工作由全国法制计量管理计量技术委员会归口单位提出，起草组由来自产学研医检审各个领域的成员单位组成，包括中国计量科学研究院、首都医科大学附属北京朝阳医院、内蒙古卫数数据科技有限公司、北京中计列伯技术交流有限公司、内蒙古财经大学、苏州天鸿盛捷医疗器械有限公司、内蒙古云想科技有限公司等相关企业全程参与讨论，起草组共同开展文献调研、国内外已上市产品调研、技术资料收集、技术讨论，完成标准征求意见稿。

在最初定题期间，初步摸索了数个高频检测且具有临床代表性项目，例如：谷丙转氨酶(ALT)、谷草转氨酶 (AST)、葡萄糖 (GLU)、白蛋白(ALB)和总蛋白(TP)5 个检测项目，一方面这些项目代表临床人群变异，即个体内生物学变异和个体间生物学变异，另一方面也反映了方法学变异，例如：设备间、试剂及定标品间差异。

但是经过一系列基于患者实验室检验检测数据的分析和挖掘，首先，我们发现相同检测项目的结果无显著差异，不会引入数据分布波动引起的平均值变化。其次，在起草 AST、ALB 两个标准过程中，逐渐发现临床上虽然检测项目间数据特点有不同，但是在算法溯源流程上样本数量、误差类型、频次等流程相同，特别是针对临床同类定量检测项目的工作流程是相同的。最后，基于患者数据的质量控制算法（Patient-based real-time quality control，PBRTQC）中除机器学习外，统计学算法也缺少溯源标准，因此考虑将本规范起草任务的目标放大，项目从个别项目扩展为了检验医学定量检测项目，即溯源的对象从机器学习算法扩展为所有基于实验室数据的质量控制算法。

最终决定将申报的《患者白蛋白实验室检测数据机器学习质量控制算法溯源》和《患者谷草转氨酶实验室检测数据机器学习质量控制算法溯源》两项国家计量技术规范，合并制定成为一项《检验医学定量检测项目基于实验室数据的实时质量控制算法溯源方法》规范。在附录A中，考虑到生物学变异性能要求，本规范根据不同的实验室检验项目，给出了优、中、低三个等级的方法性能评价水平；本规范附录B中给出了10类临床检验定量测定项目的指标范围，便于用户进一步识别、参考选择可靠的检测、检验、性能评价参数。

（二）主要起草人(以表格形式将内容明确)初拟

本规范主要起草人及任务分工见表1。

表1 主要起草人及任务分工

| 姓名 | 性别 | 职务/职称 | 工作单位 | 任务分工 |
| --- | --- | --- | --- | --- |
| 崔伟群 | 男 |  | 中国计量科学研究院 | 项目负责人、整体统筹协调 |
| 周睿 | 女 |  | 首都医科大学附属北京朝阳医院 | 主导标准制定 |
| 陈超 | 男 |  | 内蒙古卫数数据科技有限公司 | 主导标准制定 |
| 宋彪 | 男 |  | 内蒙古财经大学 | 参与标准起草 |
| 王哲 | 男 |  | 内蒙古卫数数据科技有限公司 | 参与标准起草 |
| 王亭亭 | 女 |  | 中国计量科学研究院 | 参与标准起草 |
| 郭婷 | 女 |  | 内蒙古工业大学 | 参与标准起草 |
|  |  |  |  | 资料收集、数据整理 |
|  |  |  |  | 资料收集、数据整理 |
|  |  |  |  | 资料收集、数据整理 |

二、制定规范的必要性和目的意义

**（一）质量控制算法的发展情况**

临床医学检验的目的是发出准确的检测报告，质量控制（Quality Control，QC，简称：质控）历来是临床检验管理中最重要的环节之一。传统质控临床应用中的技术问题包括：（1）质控品互换性差，导致假性预警；（2）质控品的浓度水平不足，无法涵盖检测阈值，不能有效监测某些具有临床意义的检测项目；（3）质控品的检测频次过低，导致对检测过程的监控多为以批为单位的回顾性分析，无法满足目前高通量检测技术和新型检测模式的要求；（4）无法监测分析前过程，而实验室约70%的差错发生在于分析前，传统质控不能对其进行监控；（5）目前高通量检测技术的广泛应用和临床医生对检测结果质量需求的提升，导致质控成本陡增。

**（二）本规范将为检验医学定量检测项目算法溯源标准化打下坚实基础**

鉴于市场需求和信息技术的快速发展，市场上涌现出名目繁多的PBRTQC产品，它们基于传统统计学算法或机器学习算法，检验程序的性能特征差异显著。此类算法在一定程度上存在着不协调、不系统，科学性、权威性、实用性、可靠性，缺少统一的、标准的计量手段。因此迫切需要对算法加强规范，对实验理论到实际落地的转化制定合理的评估手段，避免出现对优异结果的高估和对常庸结果的低估，防止将质量控制的意义跌落至过程控制，甚至随意控制。

本规范是参考国家、行业相关技术规范与标准以及国际相关行业专家共识，结合我国临床工作实际需求而制定，鉴于各种PBRTQC差异算法建立和实施的系统评价方法及客观量化评价指标的缺乏，提出一套高标准的规范或者规程去定量评价该类算法在临床应用的可靠程度的流程与计量方式，辅助该类算法的数据质量评价、建模指导、模型强化、泛化性分析。本规范的制定发布对提高实验室检测结果质量、保证检验结果的计量溯源性起到积极作用，最终做到检验结果可靠、一致、可比、有效。

**（三）计量规范是实验室检验检测安全的支撑条件**

与临床检测相关的误差大多发生在检验前、检验中阶段。其中，检验前的误差大多来自于不当的样本采集、运输或处理；检验中的误差大多来自于仪器，实验室通常使用内部质量控制程序来监测、保障和管理检验全过程的检测质量。

要做到检验结果可靠、一致、可比、有效，关键要保证检验结果的计量溯源性，基于患者数据的实时质量控制方法是采用算法软件监测患者数据动态，实现对检测过程分析性能连续监控的质量控制方法。本规范主要包括计量特性、数字量具特性、溯源方法及测量不确定度评定实例等内容，为实验室检验检测项目质量控制算法溯源工作提供技术依据。该规范的制定发布，加快推进各检测机构与生产企业实现测量数据准确一致，对规范实验室检验项目质控溯源方法及推进其标准化都发挥了积极作用。（内容不确定）

**（四）计量规范为健康中国提出的检验结果互认提供技术保证**

实现不同检测系统间结果的可比性是检验结果互认的基础。要实现检验结果可比性，检测系统的溯源性就很重要，如果都能溯源到国际单位制或者高等级的参考测量程序及参考物质，那么不同检测系统的结果从理论上就能够有可比性。计量溯源性是通过文件规定的不间断的校准链，测量结果与参照对象联系起来的特性，校准链中的每项校准均会引入测量不确定度。

本规范无论是其描述内容，还是其推出时机，均符合《“十四五”数字经济发展规划》的要求。通过对检验医学定量检测项目基于实验室数据的实时质量控制算法的治理，定量评价该类算法在临床应用的可靠程度，定量评价数据质量，辅助该类算法的建模指导、模型强化和泛化性分析。该规范作为通用性要求，具有广泛的适用性。本规范建设的算法溯标准参考数据，可保障以“数据、算法、算力”为核心要素的数字经济健康发展，保障算法计量准确可靠，支撑数字计量发展。

# 主要起草过程

本标准起草小组具体工作纪要如下：

（一）第一阶段（2022年1月到3月）

2022年1月，中国计量科学研究院接到《全国法制计量管理计量技术委员会立项函》通知后，与首都医科大学附属北京朝阳医院、内蒙古卫数数据科技有限公司联合成立标准编制小组，并召开网络会议，会议向各参与单位介绍了标准制修订管理规定、一般流程和时限，介绍了本标准提案的内容和立项历程，对各家的任务分工进行了梳理，明确了标准起草的流程和时间节点。

2022年2-3月，成员单位根据分工及进度计划，开展了资料收集工作，包括国内政策法规、管理办法、各级现行有效标准等相关技术资料。

（二）第二阶段（2022年4月到11月）

编制小组对收集到的国家、行业及先进发达国家检验医学定量检测项目基于实验室数据的实时质量控制算法溯源法标准制定资料进行细致、合理的分析，确定了标准的基本框架结构。

编制小组第一次推进会（2022年4-5月）：重新确定了标准的撰写范围和基本写作格式，按照GB/T1.1-2020标准化工作导则 第1部分要求撰写本标准。经过讨论研究，决定将申报的《患者白蛋白实验室检测数据机器学习质量控制算法溯源》和《患者谷草转氨酶实验室检测数据机器学习质量控制算法溯源》两项国家计量技术规范，合并制定成为一项《检验医学定量检测项目基于实验室数据的实时质量控制算法溯源方法》规范。

编制小组第二次推进会（2022年6-8月）：确定了标准的基本框架结构。包括本标准整体内容框架和各章节内容框架结构。

编制小组第三次推进会（2022年9-10月）：针对标准内容进行探讨、修改，主要完成前言、术语定义和概述等内容的修改完善。

2022年10-11月，编制小组同时对北京市临床检验中心和内蒙古自治区综合疾病预防控制中心进行了实地调研，结合临床实际，起草编制了《检验医学定量检测项目基于实验室数据的实时质量控制算法溯源法》国家标准初稿。

编制小组邀请有关单位和人员召开讨论会，对《检验医学定量检测项目基于实验室数据的实时质量控制算法溯源法》国家标准初稿进行研讨，会议对标准初稿进行了集体审阅，对成员单位提出的问题进行了讨论，明确了细化修改的方向。会后编制小组按照有关意见建议进行修改完善后，形成了《检验医学定量检测项目基于实验室数据的实时质量控制算法溯源法》国家标准征求意见稿。

（三）第三阶段（2022年12月-2023年2月）

2022年12月16日，编制小组组织召开线上标准专家咨询会，邀请产业领域、行业领域及相关企业的专家参会研讨，并经首都医科大学附属北京朝阳医院、首都医科大学附属北京地坛医院、内蒙古医科大学第二附属医院、内蒙古国际蒙医院试点验证。随后向北京市临床检验中心、内蒙古自治区综合疾病预防控制中心等多家相关单位征求意见。编制小组根据反馈的意见建议对本文件的征求意见稿进行了补充完善，形成了《检验医学定量检测项目基于实验室数据的实时质量控制算法溯源法》送审稿。

1. 制定规范的原则和依据

（一）编制原则（内容不确定，需改）

统一性原则。本规范对于评价不同类别的基于实验室检测结果的实时质量控制算法，将计量特性、数字量具特性、溯源方法、定量评价数据质量和测量不确定度进行了统一。

可及性原则。本规范在对检验医学定量检测项目基于实验室数据的实时质量控制进行调研基础上，结合国内各相关标准内容确定本标准条款、技术参数等对内容通用化处理，在数据和技术条件上均具有可及性。

适用性原则。本规范提出的技术指标，以实验室数据的实时质量控制算法为基础制定，覆盖了全部现有业务数据，满足临床检验医学日常管理及应用。

规范性原则：本技术规范严格按照JJF 1071-2010《国家计量校准规范编写规则》 、JJF 1001 《通用计量术语及定义技术》以及GB/T1.1-2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》给出的规则进行起草，文本格式规范。

（二）制定标准的依据

本规范在制定过程中，梳理了国内外监管法规要求和公布的技术验证方式，借鉴了CLSI C24 4th ed《Statistical Quality Control for Quantitative Measurement Procedures: Principles and Definitions》等国际测试的新思想，借鉴了WS/T 641《临床检验定量测定室内质量控制》等规范文件，调研了学术界对相关指标的定义和计算方法。标准的架构借鉴了JJF 1182 《计量器具软件测评指南》，基于实验室患者检测出数据，围绕实时质量控制机器学习算法软件的特殊场景对计量测试方法和测评指标进行描述。

（三）内容涵盖

本规范包括7个章节，分别为：

1. 范围；
2. 规范性引用文件；
3. 术语和定义；
4. 概述；
5. 计量特性；
6. 数字量具特性；
7. 溯源方法。

此外，本规范包括三个资料性附录，为标准的实施提供参考信息。其中，附录A是对相关医学概念的补充说明，用于帮助测试人员理解算法应用的场景和医学名词。附录B和附录C是对正文条款进行补充说明，用于帮助测试人员明确算法测试指标的含义，提高标准的可实施性。

正文内容考虑了算法性能测试的组织实施、指标定义与具体计算；附录的内容作为补充举例和思路的解释。在本标准制定的过程中，未涉及有争议项目。

五、主要条款的说明，主要技术指标、参数、试验验证的论述

1、术语和定义

1.1、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，术语 3.3.7]

## 1.2、失控状态 out-of-control condition

某过程或过程中的某部分操作处于非稳定状态。

注: 对于定量测量程序，失控状态通常表现为测量程序偏离稳定均值或者稳定精密度增大。

[CLSI C24 4th ed]

## 1.3、在控状态 in-control status of QC

某过程或过程中的某部分操作处于非失控状态的稳定状态。

## 1.4、控制限 control limits

控制图上用于判断过程是否处于统计控制状态的界限。

注：控制限分上控制限（upper control limit）和下控制限（lower control limit）,当过程处于统计控制状态时，所考虑的统计量将以很高的概率位于上/下控制限之间。

[GB/T 3358.2-1993，过程量度术语5.39]

## 1.5、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（QC）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QC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]

## 1.6、不精密度 imprecision

检验结果精密度指标，指同一实验室用同种方法在多次独立检验中分析同一样品所得结果的离散程度。

注1：

不精密度有多种，如“批内”、“批间”、“室内”、“室间”不精密度等，计量学领域分重复性、中间、再现性不精密度等。本标准中的不精密度指室内（或中间）不精密度。

注2：

对于临床实验室，不精密度一般来自室内质控数据；在方法评价中，本标准不精密度包括批内和批间不精密度。

注3：

不精密度用标准差（SD）或变异系数（CV）表示。

[WS/T 403-2012，术语2.2]

## 1.7、特异性 specificity

在出现干扰现象（影响量）时，试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，质量控制系统能够指出特定原因变异不存在的概率。

## [WS/T 415-2013，术语2.8]

## 1.8、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，是指质量控制系统检测误差的能力。

注2:

在定性试验中， 是指试验方法获得阳性结果的能力。

[WS/T 415-2013，术语2.7]

**2、主要内容**

检验医学定量检测项目基于实验室检测结果的实时质量控制算法的计量特性包括：误差检出概率、假性失控概率、错误报告患者检测结果数量、算法评价值四个评价指标。

## 2.1、假性失控概率

假性失控概率（PFR）表征被测试对象的特异性，反映算法对于预期在控状态的误判概率，依据公式1：

（1）

式中：

——预期在控状态的检测结果个数；

——预期在控状态的第个检测结果的实际判定结果，取值为0（判定在控状态）或者1（判定失控状态），且。

## 2.2、误差检出概率

误差检出概率（PED）表征被测试对象的敏感度，反映算法对于预期失控状态的识别概率，依据公式2：

（2）

式中：

——预期失控状态的检测结果个数；

——预期失控状态的第个检测结果的实际判定结果，取值为0（判定在控状态）或者1（判定失控状态），且。

对于初始开机时核查，应满足：*PED*>95%，*PFR*<5%；对于过程核查，应满足：*PED*>90%，*PFR*<0.5%。该取值来自行业标准。

## 2.3、错误报告患者检测结果数量

错误报告患者检测结果数量（median number of patient results affected before error detection, MNped）表征被测试对象的识别速度，反映算法对预期失控状态的识别中所延误的结果数量。通过多个Nped的中位数计算所得，其中Nped定义为：在某次质量控制检测中，以引入失控状态数据的位置为起点，以算法初次识别为失控状态的位置为终点，统计两点之间的结果个数。每次质量控制检测均采用相同的检验结果数量。

2.4、算法评价值

算法评价值（）表征被测试对象的准确性和可靠性，反映算法对于误差检出概率、假性失控概率、错误报告患者检测结果数量的综合特性，如公式3：

（3）

式中：

，——由专家经验评估的权重系数，默认取值为2/3，1/3；

——预期在控状态的检测结果个数；

——预期失控状态的检测结果个数；

——预期在控状态的第个检测结果的实际判定结果，取值为0（判定在控状态）或者1（判定失控状态），且；

——预期失控状态的第个检测结果的实际判定结果，取值为0（判定在控状态）或者1（判定失控状态），且；

——每次质控检测的检验结果数量。

**六、与有关的现行法律、法规和强制性国家标准的关系**

本标准符合相关国家标准的相关规定，能与现行有关的法律、法规和国家标准规范相衔接，与其他现行法律、法规和强制性国标、行标没有冲突。

**七、重大分歧意见的处理经过和依据**

未出现重大意见分歧。

**八、采用国际标准或者国外先进标准的，说明采标程度，以及国内外同类标准的对比情况**

本标准未采用国际标准或国外先进标准。本标准不属于国际标准转化，国外目前暂无同类标准。

**九、作为推荐性标准或者强制性标准的建议及其理由**

由于基于患者实验室数据医学定量检测项目算法溯源的预期用途和应用场景较为稳定成熟，本规范的技术内容有较强的参考性，建议作为推荐性国家标准。

**十、贯彻标准的要求和措施建议**

建议本标准的过渡期为1年，本归口单位拟在标准发布后实施前进行宣贯。

**十一、其他应予说明的事项**

暂无。

《基于实验室数据的实时质量控制算法溯源方法》标准起草小组

2023年10月19日