|  |  |
| --- | --- |
| ICS  |       |
| CCS  |

|  |
| --- |
| D:\000000部门项目\09标准化插件开发\程序源代码\StandardEditor_ShanDongKeXieYuan\团标首页面字母T.pngD:\000000部门项目\09标准化插件开发\程序源代码\StandardEditor_ShanDongKeXieYuan\团标首页面字母T后面的反斜杠.png CMA |

      |

中国计量协会团体标准

T/CMA JD XXX—2024

代谢组学质谱标准参考数据评估规范

Evaluation Specification for Metabolomics Mass Spectrometry Standard Reference Data

2024 - XX - XX发布

2024 - XX - XX实施

中 国 计 量 协 会  发布

目次

[前言 III](#_Toc206700401)

[1 范围 4](#_Toc206700402)

[2 规范性引用文件 4](#_Toc206700403)

[3 术语和定义 4](#_Toc206700404)

[3.1 代谢物 metabolites 4](#_Toc206700405)

[3.2 代谢组 metabolome 4](#_Toc206700406)

[3.3 代谢组学 metabolomics 4](#_Toc206700407)

[3.4 质谱法 mass spectrometry 4](#_Toc206700408)

[3.5 代谢谱分析 metabolic profiling 4](#_Toc206700409)

[3.6 基于质谱的代谢组学 MS-based metabolomics 4](#_Toc206700410)

[3.7 标准参考数据 standard reference data 5](#_Toc206700411)

[3.8 质谱标准参考数据 mass spectrometry standard reference data 5](#_Toc206700412)

[3.9 代谢组学质谱标准参考数据 metabolomics mass spectrometry standard reference data 5](#_Toc206700413)

[3.10 保留时间 retention time 5](#_Toc206700414)

[3.11 一级质谱 MS1 5](#_Toc206700415)

[3.12 二级质谱 MS2 5](#_Toc206700416)

[3.13 质量控制样本 quality control sample 5](#_Toc206700417)

[3.14 计量溯源性 metrological traceability 5](#_Toc206700418)

[3.15 测量不确定度 6](#_Toc206700419)

[3.16 严格评估原则 6](#_Toc206700420)

[4 概述 6](#_Toc206700421)

[5 评审内容 6](#_Toc206700422)

[5.1 数据来源鉴别 6](#_Toc206700423)

[5.2 评价 6](#_Toc206700424)

[5.3 准确性核验 7](#_Toc206700425)

[5.4 相关性 8](#_Toc206700426)

[6 评估流程 8](#_Toc206700427)

[6.1 核心评审角色 8](#_Toc206700428)

[6.2 评审总流程 8](#_Toc206700429)

[7 评估指标 11](#_Toc206700430)

[8 数据提交要求 11](#_Toc206700431)

[8.1 格式规范 12](#_Toc206700432)

[8.2 辅助材料 12](#_Toc206700433)

[9 评审程序 12](#_Toc206700434)

[9.1 专家资质 12](#_Toc206700435)

[9.2 双盲评审 12](#_Toc206700436)

[9.3 申诉机制 12](#_Toc206700437)

[10 结果判定 12](#_Toc206700438)

[10.1 通过 12](#_Toc206700439)

[10.2 限期整改 12](#_Toc206700440)

[10.3 不通过 12](#_Toc206700441)

[附录A （规范性） 代谢组学质谱数据评估表（含各指标评分细则） 13](#_Toc206700442)

[附录B （规范性） 申请资料清单 14](#_Toc206700443)

[附录C （规范性） 数据提交模板（MIxS XML、mzML 及 Excel 文件示例） 15](#_Toc206700444)

[C.1 MIxS XML 格式模板 15](#_Toc206700445)

[C.2 mzML 格式核心片段示例 15](#_Toc206700446)

[C.3 Excel 处理后数据模板 16](#_Toc206700447)

[附录D （规范性） 评审材料清单 17](#_Toc206700448)

前言

本文件按照GB/T 1.1—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》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由中国计量协会提出。

本文件由中国计量协会XXX归口。

本文件起草单位：

本文件主要起草人：

代谢组学质谱标准参考数据评估规范

* 1. 范围

本规范规定了代谢组学研究中使用质谱技术生成的标准参考数据的评估要求，适用于数据采集、处理、分析及报告的全流程。评估对象包括但不限于液相色谱-质谱联用（LC-MS）、气相色谱-质谱联用（GC-MS）等技术平台产生的代谢组学数据。

* 1. 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

本规范引用下列文件：

ISO 20391-1:2020 生物技术 生物样本库 第 1 部分：通用要求

MIxS（Minimum Information about a Metabolomics Experiment）标准

MSI（Metabolomics Standards Initiative）代谢物鉴定指南

JJF 1001-2011 通用计量术语及定义

GB/T 44827-2024/ISO 23118:2021 分子体外诊断检验 尿液、静脉血清和血浆代谢组学检验前过程的规范

ISO/IEC Guide 98-3:2008 测量不确定度 第三部分：测量不确定度表示指南（GUM:1995）

* 1. 术语和定义

下列术语适用于本规范。

* + 1. 代谢物 metabolites

宿主生物体、微生物群落的小分子代谢中间产物和/或终产物（≤2000Da），它们来自食物、饮料、药物或污染物。

[来源：GB/T 44827-2024/ISO 23118:2021]

* + 1. 代谢组 metabolome

在生物体或生物样品中发现的完整的代谢物集合。

[来源：GB/T 44827-2024/ISO 23118:2021]

* + 1. 代谢组学 metabolomics

对生物标本[如生物体、细胞、组织或生物体液]代谢组的全面分析。

[来源：GB/T 44827-2024/ISO 23118:2021]

* + 1. 质谱法 mass spectrometry

基于化合物的质荷比而对其进行分析的方法。

[来源：GB/T 44827-2024/ISO 23118:2021]

* + 1. 代谢谱分析 metabolic profiling

利用分析平台同时测量生物体系中的代谢物成分集合，这些代谢物可能通过所使用（或所选定的）技术来进行测量。

[来源：GB/T 44827-2024/ISO 23118:2021]

* + 1. 基于质谱的代谢组学 MS-based metabolomics

利用质谱法测量生物样品中的代谢物。

[来源：GB/T 44827-2024/ISO 23118:2021]

* + 1. 标准参考数据 standard reference data

由公认的权威机构发布的参考数据，包括软件、算法、模型、音视频、指纹等数字数据。

例 1：国际科学联合会科学技术数据委员会（ICSU CODATA）作为法规评定和发布的基本物理常量的值。

例 2：元素的相对原子质量值，也称原子重量值，由国际理论和应用化学联合会（IUPAC-CIAAW）在国际理论和应用化学联合会（IUPAC）全会上每两年评定一次并在《纯应用化学》和《物理化学参考数据》上发布

* + 1. 质谱标准参考数据 mass spectrometry standard reference data

通过标准化流程生成的、可用于方法验证或质量控制的质谱数据集合。

* + 1. 代谢组学质谱标准参考数据 metabolomics mass spectrometry standard reference data

通过标准化实验流程生成，包含代谢物特征信息（如保留时间、质荷比、碎片模式等）及定量结果，可用于质谱方法验证或数据质量控制的代谢组学数据。

* + 1. 保留时间 retention time

样品中某一组分从进入色谱系统到被检测器检出的时间间隔，单位为分钟（min）。

* + 1. 一级质谱 MS1

对离子源产生的母离子进行质量分析得到的质谱图，用于获取物质的精确分子量信息。

* + 1. 二级质谱 MS2

对选定的母离子进行碎裂后，对产生的碎片离子进行质量分析得到的质谱图，用于物质结构鉴定。

* + 1. 质量控制样本 quality control sample

将所有待测样本按等比例混合制备的标准品，用于监控实验过程中仪器稳定性及数据重现性。

* + 1. 计量溯源性 metrological traceability

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。

注:

1. 本定义中的参照对象可以是实际实现的测量单位的定义，或包括无序量测量单位的测量程序，或测量标准。
2. 计量溯源性要求建立校准等级序列。
3. 参照对象的技术规范必须包括在建立等级序列时所使用该参照对象的时间，以及关于该参照对象的任何计量信息，如在这个校准等级序列中进行第一次校准的时间。
4. 对于在测量模型中具有一个以上输入量的测量，每个输入量本身应该是经过计量溯源的，并且校准等级序列可形成一个分支结构或网络。为每个输入量建立计量溯源性所作的努力应与对测量结果的贡献相适应。
5. 测量结果的计量溯源性不能保证其测量不确定度满足给定的目的，也不能保证不发生错误。
6. 如果两个测量标准的比较用于检查，必要时用于对量值进行修正，以及对其中一个测量标准赋予测量不确定度时，测量标准间的比较可看作一种校准。
7. 两台测量标准之间的比较，如果用于对其中一台测量标准进行核查以及必要时修正量值并给出测量不确定度，则可视为一次校准。
8. 国际实验室认可合作组织(ILAC)认为确认计量溯源性的要素是向国际测量标准或国家测量标准的不间断的溯源链、文件规定的测量不确定度、文件规定的测量程序、认可的技术能力、向SI的计量溯源性以及校准间隔。
9. “溯源性”有时是指“计量溯源性”，有时也用于其他概念，诸如“样品可追溯性”、“文件可追溯性”或“仪器可追溯性”等，其含义是指某项目的历程(“轨迹”)。所以，当有产生混淆的风险时，最好使用全称“计量溯源性”。

[来源：JJF 1001-2011]

* + 1. 测量不确定度

表征合理赋予被测量值的分散性，与测量结果相联系的参数。

[来源：ISO/IEC Guide 98-3:2008）]

* + 1. 严格评估原则

针对数值数据（如物质或物质系统的属性信息），需要保证数据的完整性（例如，明确测量不确定度和相关测量标准），检查数据的合理性（例如，比较数据与物理原理的一致性，对比不同的独立方法获得的数据），并评估数据的可用性（例如，包含元数据和完整记录的测量程序）；针对数字数据对象（如图像、视频），首先要确保该对象是基于物理原理、基础科学或广泛接受的数据收集标准操作程序获取的，并且要检查该对象已经过测试，或者其计算数据和实验数据已经过定量比较。

* 1. 概述

本规范旨在为代谢组学质谱数据的评估提供一套科学、严谨、可操作的技术要求和流程。其核心目标是评估一项代谢组学质谱数据是否能够作为质谱标准参考数据。本规范是在《标准参考数据评审技术规范》的总体框架和原则指导下 ，针对代谢组学质谱数据的特殊性而制定的专业领域实施细则。评估内容将遵循基础规范，围绕“严格评价”和“准确性核验”两个核心维度展开 。其中，“严格评价”具体化为对数据完整性、一致性、实用性及可靠性等标准属性的评审 ；“准确性核验”则具体化为对数据准确性、计量溯源性及重复性等计量属性的核验 。

本规范适用于指导数据生产者提交、整理数据，并为评审机构和技术专家对代谢组学质谱标准参考数据进行技术评审提供统一依据。

* 1. 评审内容

对代谢组学质谱标准参考数据的评估，应遵循《标准参考数据评审技术规范》所确立的基本原则，并结合领域特性进行适配。主要原则如下：

* + 1. 数据来源鉴别

这是对数据来源有效性的基本要求，旨在确保数据的“出身”清晰、合规。

来源清晰：数据生产者或所有者的信息应明确，并提供必要的资格证明文件，如CNAS测试证书、计量证书（CMA）等 。

标识唯一：数据应有关联的唯一标识，如数字资源标识（DOI）或机构内部统一编号，便于追溯和引用 。

* + 1. 评价

严格评价关注数据的标准属性，确保数据在科学上是可靠的、完整的、一致且可用的。

* + - 1. 完整性 (Completeness)

数据应包含元数据、原始数据和处理后的数据，信息充分，没有缺项。

元数据完整：应遵循代谢组学领域的最低信息标准（如MIxS），详细记录样本来源、前处理方法、仪器型号参数、实验日期等信息 。

测量程序完整：必须提供完整记录的测量程序，包括详细的SOP（标准操作程序），确保实验流程的可复现。

数据提交格式与内容完整性：

原始数据：采用 mzML 1.1.0 或 mzXML 3.2 格式，包含完整的质谱扫描事件（如扫描次数、扫描类型）及仪器参数（如离子源类型、喷雾电压、鞘气流量）；

处理后数据：采用 Excel 2007 及以上版本（.xlsx），字段包括：代谢物名称、CAS 号、化学式、保留时间（RT）、一级质谱精确质量（MS1）、二级质谱碎片信息（MS2）、定量值及单位。

元数据：采用 MIxS 格式 XML 文件，包含样本编号、基质类型（如血浆、尿液）、采集时间、前处理步骤（如提取试剂、衍生化条件）、仪器型号及序列号等；

辅助材料完整：需包含实验方案（明确样本采集标准、前处理流程、仪器操作参数等）、质量控制报告（含 QC 样本的总离子流图叠加图、PCA 得分图、代谢物峰面积 RSD 分布直方图等）、方法学验证报告（涵盖线性范围、检出限、定量限、回收率及精密度等参数）、溯源性证明（如标准物质证书编号及校准记录）。

* + - 1. 一致性

数据内部以及与其他公认数据或理论之间不应存在矛盾。

原理一致性：数据结果应与公认的物理化学及生物学原理一致 。例如，在电喷雾正离子模式下，观测到的主要离子加和形式（如[M+H]+, [M+Na]+）应与所用的流动相体系和样品基质相符。

方法一致性：若条件允许，可与其他独立分析方法获得的数据进行比对，结果应表现出良好的一致性 。

技术平台与处理一致性：

技术平台合规性：GC-MS 需提供衍生化试剂类型（如 BSTFA、MSTFA）及衍生化条件（温度、时间）；LC-MS 需明确色谱柱固定相类型（如 C18、HILIC）及流动相组成（如甲醇 - 水、乙腈 - 0.1% 甲酸）；CE-MS 需说明缓冲液成分（如甲酸铵、ammonium acetate）及 pH 值；

数据处理一致性：峰对齐算法（如 XCMS）的稳定性需满足保留时间偏差≤0.2 min，峰面积缺失值≤20%；批次效应校正需采用基于 QC 样本的 LOESS 校正或 ComBat 算法，校正后批次间代谢物信号中位数偏移≤10%；缺失值处理需遵循规则（缺失率＞70% 赋 0；20%-70% 采用随机森林或 kNN 插补；＜20% 按组别中位数填充）。

* + - 1. 可靠性

数据的可信度应通过外部比对或评价来证明。

经过比对验证：数据应经过有效的比对，如室内比对、实验室间比对或国际比对，以证明其稳定性和可靠性。

经过第三方评价：数据或其产生方法经过第三方权威机构（如NIST）的评价或测评者优先。

数据采集可靠性：

质谱参数合规：质谱分辨率（如 Orbitrap 需≥120,000）、质量精度（内标物误差≤5 ppm）及扫描模式（如 DDA 的 Top N 设置）需符合标准；

质量控制规范：序列分析中每 10 个样本插入 1 个 QC 样本，QC 样本占比≥10%；采用 Shewhart 控制图检测异常 QC 样本（如连续 2 个 QC 样本峰强度超 ±2SD）。

* + 1. 准确性核验

准确性核验关注数据的计量属性，确保数据的测量值精准，并能溯源至公认标准。

* + - 1. 准确性

数据应准确表示其所描述的代谢物的真实特征。

质量轴准确性：一级质谱的质量精度是关键指标，内标物误差应小于5 ppm，目标代谢物误差应小于10 ppm 。

鉴定准确性：代谢物鉴定需提供多维度证据，如精确质量、二级质谱碎片匹配、保留时间比对等 。对于鉴定结果，需报告假阳性发现率（FDR）等质量控制指标 。

代谢物鉴定准确性：一级质谱质量误差需≤10 ppm，离子加和形式需符合流动相体系；二级质谱匹配得分≥60 分（如 GNPS 库匹配），碎片离子覆盖率≥50%；保留时间与标准品或自建数据库偏差≤0.5 min，且通过至少 3 次独立进样验证；同分异构体需提供离子淌度质谱（IM-MS）的碰撞截面（CCS）值，与数据库偏差≤2%；未知代谢物注释需结合核心结构搜索和假设中性损失匹配。

定量分析准确性：内标回收率需在 80%-120% 范围内，线性范围 R²≥0.99；采用 PLS-DA 或 OPLS-DA 模型筛选差异代谢物时，VIP 值需≥1，且通过 200 次置换检验（Q² 截距≤0.05）。

* + - 1. 计量溯源性

定量数据的测量结果应能通过不间断的校准链与国家或国际基准联系起来。

可溯源至标准物质：对于定量数据，其量值应能溯源至有证标准物质（CRMs/SRMs），并提供溯源性证明 。

明确测量不确定度：必须评定并报告关键测量结果（如代谢物浓度）的测量不确定度。

* + - 1. 可重复性

在相同条件下重复测量，结果应表现出高度的一致性 。

数据重现性：通过连续进样质量控制样本（QC）进行评估，QC样本中大部分代谢物峰面积的相对标准偏差（RSD）应满足要求（例如LC-MS中≤30%）。

稳定性：考察数据在规定时间内的稳定性，确保其在有效期内质量可靠。

* + 1. 相关性

数据必须与其声称描述的现象、物体或物质特性有明确的关联 。

明确关联对象：数据应清晰地与特定的生物样本（如人血浆、尿液）、代谢物（已知化合物成分）及相关生物学现象建立关联 。

服务于特定应用：数据应能服务于方法验证、仪器校准、算法模型开发等明确的科学或计量应用。

* 1. 评估流程
		1. 核心评审角色

评审过程涉及的角色及其在代谢组学领域的特定职责如下 ：

申请人：代谢组学质谱数据的提供方或提交者 。

形式审查员：负责初步审查，专注于检查代谢组学数据提交的格式规范性、辅助材料的完整性 。

技术委员会：其下属的专业组是评审的核心技术机构 。对于此类数据，应设立一个包含代谢组学、质谱技术、计量学和生物信息学专家的核心评估组 。

指导委员会：对技术委员会的评审结论和资料进行验证与复核 。

管理办公室：作为最终的批准与发证机构，负责整个评审的组织与管理

评审过程中涉及的角色关系与各角色承担任务如图1所示。



图1 标准参考数据评估机构角色图

* + 1. 评审总流程

评审总流程如图2所示。申请人首先向评审机构提交数据及相关资料 。审查过程将依次通过形式审查、技术委员会专家评审、指导委员会验证评审，最终由管理办公室批准认证 。



图2 标准参考数据评审总流程图

* + - 1. 资料提交

申请人向标准参考数据评审机构提交相关资料，申请资料清单可参考附录B中表B.1（8 数据提交要求）

* + - 1. 形式审查员评审流程

形式审查员进行资料受理与数据清单审查流程如图3所示：审查清单包括：数据来源鉴别、数据类型、数据有效性、数据格式、可读性、证明文件资料、数据标识、数据文档资料、数据分类等，审查资料清单详细内容见附录D中表D.1：形式审查员资料形式审查清单。形式审查员收到申请人提交的资料后，首先进行受理该申请并进行形式审查，通过后将由形式审查员依据申请数据所属专业领域分配到技术委员会中的对应专业组，如质谱专业组、光学专业组、化学专业组等，由专业组专家进行技术专家评审。



图3 形式审查员审查流程

* + - 1. 技术委员会评审流程

技术委员会评审流程如图4所示，由对应专业组专家进行技术专家评审，评审内容包括严格评价、准确性验证、相关性评审、数字数据评审等内容，评审内容详细列表查看附录D中表D.2：技术委员会评审检查清单，通过评审后生成评审报告并进行评审资料备案，通过后的备案资料将提交指导委员会。

 

图4 技术委员会评审流程

* + - 1. 指导委员会评审流程

指导委员会评审流程按图5所示，由指导委员会验证技术委员会提交的评审资料，并对申请数据的严格评价、准确性验证、相关性、数字数据等再次进行评审和核验，评审内容详细列表查看附录D中表D.3：指导委员会评审检查清单，通过指导委员会评审后的数据将提交到管理办公室，由管理办公室批准后认证为正式标准参考数据。



图5 指导委员会评审流程

* + - 1. 批准与发证

管理办公室根据指导委员会的评审结论，对申请数据进行最终审批 。

批准：对于通过所有评审环节的数据，批准其成为正式的标准参考数据 。

发证与注册：由管理办公室制作并颁发《标准参考数据证书》，规定统一编号，并为其注册数字对象唯一标识符 (DOI)。

具体判定标准见（10 结果判定）。

* 1. 评估指标
1. 评估指标

|  |  |  |
| --- | --- | --- |
| 序号 | 评估维度 | 合格标准 |
| 1 | 质谱分辨率 | LC-MS ≥120,000（m/z 200），GC-MS ≥5,000（FWHM） |
| 2 | 质量精度（MS1） | ≤10 ppm（内标物≤5 ppm） |
| 3 | 二级质谱匹配得分 | ≥60 分（总分 100），碎片离子覆盖率≥50% |
| 4 | 保留时间重复性 | QC 样本 RSD≤5%，与标准品偏差≤0.5 min |
| 5 | 数据完整性 | 元数据符合 MIxS 标准，原始数据与处理后数据一致率≥95% |
| 6 | 定量准确性 | 内标回收率 80%-120%，线性范围 R²≥0.99 |
| 7 | 统计模型可靠性 | OPLS-DA 模型 Q²≥0.4，置换检验 Q² 截距≤0.05 |
| 8 | QC样本稳定性​​ | ≥70%代谢物峰面积RSD≤20%（LC-MS） |
| 9 | 批次效应校正​​ | 校正后组内变异≤组间变异的15% |
| 10 | 假阳性控制​​ | FDR估计误差率≤5%（0-0.1区间） |
| 11 | 数据标准化​​ | 优先采用BRDG/MED法，而非TIC归一化 |

* 1. 数据提交要求
		1. 格式规范

原始数据：mzML 格式，包含完整的扫描事件及仪器参数（如喷雾电压、碰撞能量）。

处理后数据：Excel 表格，包含代谢物名称、RT、MS1、MS2 及定量值。

元数据：MIxS 格式 XML 文件，记录样本来源、前处理方法及仪器型号。

* + 1. 辅助材料

自建数据库（如 LUG、EMDB）需提供构建方法及验证数据（如标准品匹配率≥90%）。

AI 模型（如 CMSSP）需提交训练集、验证集及独立测试集的准确率（Top-1 命中率≥80%）。

* 1. 评审程序
		1. 专家资质

评审组需包含代谢组学、质谱技术及生物信息学领域的专业人员，且具备3年以上相关研究经验。

* + 1. 双盲评审

数据生产者信息与评审专家信息相互保密，确保评估结果客观公正。

* + 1. 申诉机制

对评估结果有异议时，数据生产者可在收到报告后15个工作日内提交申诉材料，由第三方机构进行复核。

* 1. 结果判定
		1. 通过

所有评估指标均符合 6.1 的要求，无 “不通过项”，且整改项已完成验证。

* + 1. 限期整改

存在≤3 项非关键指标未达标（如保留时间偏差 0.25 min、缺失值 22%），数据生产者需在 30 个工作日内完成整改并通过二次评审。

* + 1. 不通过

存在以下情况之一的，判定为不通过：

* + - 1. 关键指标未达标（如质量精度＞15 ppm、代谢物鉴定匹配得分＜50分）。
			2. ＞3 项指标未达标；
			3. 整改后仍不符合要求；
			4. 存在数据造假（如篡改质谱图、伪造标准品证书）。
1. （规范性）
代谢组学质谱数据评估表（含各指标评分细则）
	1. 代谢组学质谱数据评估得分表

|  |  |  |  |
| --- | --- | --- | --- |
| 评估维度 | 评估指标 | 评分标准（总分 100 分） | 得分 |
|
| 质谱性能（20 分） | 质谱分辨率 | LC-MS≥120,000（m/z 200）得 10 分；GC-MS≥5,000（FWHM）得 10 分，未达标按比例扣分 | 　 |
| 质量精度（MS1） | ≤10 ppm（内标物≤5 ppm）得 10 分，每超 1 ppm 扣 2 分，扣完为止 | 　 |
| 数据质量（30 分） | 二级质谱匹配 | 匹配得分≥60 分且碎片覆盖率≥50% 得 10 分，一项不达标扣 5 分 | 　 |
| 保留时间重复性 | QC 样本 RSD≤5% 且与标准品偏差≤0.5 min 得 8 分，一项不达标扣 4 分 | 　 |
| 数据完整性 | 元数据符合 MIxS 标准且原始与处理后数据一致率≥95% 得 7 分，一项不达标扣 3.5 分 | 　 |
| QC 样本稳定性 | ≥70% 代谢物峰面积 RSD≤20%（LC-MS）得 5 分，每降低 5% 扣 1 分 | 　 |
| 定量与统计（25 分） | 定量准确性 | 内标回收率 80%-120% 且线性范围 R²≥0.99 得 10 分，一项不达标扣 5 分 | 　 |
| 统计模型可靠性 | OPLS-DA 模型 Q²≥0.4 且置换检验 Q² 截距≤0.05 得 8 分，一项不达标扣 4 分 | 　 |
| 批次效应校正 | 校正后组内变异≤组间变异 15% 得 7 分，每超 5% 扣 2 分 | 　 |
| 数据处理（15 分） | 假阳性控制 | FDR 估计误差率≤5%（0-0.1 区间）得 8 分，每超 1% 扣 1.6 分 | 　 |
| 数据标准化 | 采用 BRDG/MED 法得 7 分，采用 TIC 归一化得 3 分，未标准化得 0 分 | 　 |
| 综合判定 | 合格线 80 分，≥80 分得 “通过”；60-79 分得 “限期整改”；＜60 分得 “不通过” | / | 　 |

1. （规范性）
申请资料清单

表B.1 申请资料清单表

|  |  |  |
| --- | --- | --- |
| 数据类型 | 格式要求 | 具体内容说明 |
| 原始数据 | mzML 格式 | 包含完整的扫描事件及仪器参数（如喷雾电压、碰撞能量） |
| 处理后数据 | Excel 表格 | 包含代谢物名称、保留时间（RT）、一级质谱精确质量（MS1）、二级质谱碎片信息（MS2）及定量值 |
| 元数据 | MIxS 格式 XML 文件 | 记录样本来源、前处理方法及仪器型号 |
| 辅助材料-自建数据库 | - | 需提供构建方法及验证数据（如标准品匹配率≥90%） |
| 辅助材料-AI模型 | - | 需提交训练集、验证集及独立测试集的准确率（Top-1 命中率≥80%） |

1. （规范性）
数据提交模板（MIxS XML、mzML 及 Excel 文件示例）
	1. MIxS XML 格式模板

<mixs xmlns="http://example.org/mixs" version="1.0">

 <sample>

 <sample\_id>Sample-2025-001</sample\_id>

 <sample\_type>血浆</sample\_type>

 <collection\_date>2025-08-01</collection\_date>

 <matrix\_type>人血浆</matrix\_type>

 <preparation\_steps>

 <step>4℃，12000rpm离心10min，取上清</step>

 <step>加入乙腈（1:3，v/v）沉淀蛋白，-20℃静置30min</step>

 </preparation\_steps>

 <instrument>

 <model>Thermo Q Exactive HF</model>

 <serial\_number>SN123456</serial\_number>

 <ionization\_mode>ESI+</ionization\_mode>

 <scan\_range>m/z 50-1200</scan\_range>

 </instrument>

 <chromatography>

 <column>Thermo Hypersil Gold C18（2.1×150mm，1.9μm）</column>

 <mobile\_phase\_a>0.1%甲酸水溶液</mobile\_phase\_a>

 <mobile\_phase\_b>0.1%甲酸乙腈溶液</mobile\_phase\_b>

 <gradient>0-2min 5%B；2-15min 5%-95%B；15-18min 95%B</gradient>

 </chromatography>

 </sample>

</mixs>

* 1. mzML 格式核心片段示例

<mzML xmlns="http://psi.hupo.org/ms/mzml" version="1.1.0">

 <run id="run1" start\_time="2025-08-01T09:00:00">

 <scanList>

 <scan id="scan1" scan\_number="1" retention\_time="0.500min">

 <cvParam cvRef="MS" accession="MS:1000514" name="positive scan"/>

 <cvParam cvRef="MS" accession="MS:1000529" name="centroid spectrum"/>

 <binaryDataArrayList>

 <binaryDataArray type="m/z">

 <data>50.0123, 100.0456, 150.0789</data>

 </binaryDataArray>

 <binaryDataArray type="intensity">

 <data>1000.0, 5000.0, 3000.0</data>

 </binaryDataArray>

 </binaryDataArrayList>

 </scan>

 </scanList>

 </run>

</mzML>

* 1. Excel 处理后数据模板

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 代谢物名称 | CAS 号 | 化学式 | 保留时间（min） | MS1（m/z） | MS2 碎片（m/z: 强度） | 定量值 | 单位 | 鉴定级别 | FDR（%） |
| 葡萄糖 | 50-99-7 | C6H12O6 | 8.25 | 181.0765 | 89.023:1000, 121.034:800 | 5.2 | mmol/L | 1 级 | 0.02 |
| 乳酸 | 50-21-5 | C3H6O3 | 3.12 | 89.0234 | 43.005:600, 61.012:900 | 2.8 | mmol/L | 1 级 | 0.01 |

1. （规范性）
评审材料清单

表D.1 形式审查员资料形式审查清单

|  |  |  |
| --- | --- | --- |
| 序号 | 审查项 | 审查内容 |
| 1 | 数据来源鉴别 | 1.资源标识（GB、GJB、JJG、JJF、DOI、ISBN、ORCID）2.资格证书（计量证书、CNAS 测试证书、其他证书） |
| 2 | 数据类型 | 文本、数字、日期等，并检查数据类型正确 |
| 3 | 数据有效性 | 数据的有效性，即数据是否符合特定的标准、规则或约束；数据的小数位数或有效数字位数符合规定 |
| 4 | 数据格式 | 检查数据格式，确保符合指定标准；数据完整，是否缺少任何必要字段或信息；数据不超出规定的长度范围 |
| 5 | 可读性 | 评估数据是否易读，是否有良好的格式和清晰的标记；命名规范，命名符合规定且具有一致性；数据根据规定的字段进行正确排序；数据的取值范围符合规定 |
| 6 | 证明文件资料 | 证明数据正确测量方法的专家评审意见书、相关证明文件、或者国标文件、JJF 等认定了数据准确性、不确定度的证明 |
| 7 | 数据标识 | 检查数据是否有独特的标识符，以便跟踪和识别 |
| 8 | 数据文档资料 | 审查数据的相关文档，包括数据定义、数据字典 |
| 9 | 数据分类 | 数据是否根据正确的分类标准进行分类（是否有对应领域的技术专家委员会） |

表D.2 技术委员会评定检查清单

|  |  |  |  |
| --- | --- | --- | --- |
| 序号 | 评审类别 | 评审内容 | 具体评审规则 |
| 1 | 严格评价 (标准性) | 完整性 | • 数据应包含元数据、原始数据和处理后的数据，信息充分，没有缺项 。• 元数据：应遵循代谢组学领域的最低信息标准（如MIxS），详细记录样本来源、前处理方法、仪器型号参数、实验日期等 。• 测量程序：必须提供完整记录的SOP（标准操作程序），确保实验流程的可复现性 。• 辅助材料：应提供实验方案、质量控制报告、方法学验证报告及溯源性证明等 。 |
| 一致性 | • 原理一致性：数据结果应与公认的物理化学及生物学原理一致（例如，电喷雾正离子模式下的主要离子加和形式应与流动相体系相符） 。• 方法一致性：与其他独立分析方法获得的数据比对，结果应表现出良好的一致性 。• 数据处理一致性：- 峰对齐：保留时间偏差 ≤ 0.2 min；峰面积缺失值 ≤ 20% 。- 批次效应校正：采用基于QC样本的LOESS校正或ComBat算法，校正后批次间信号中位数偏移 ≤ 10% 。- 缺失值处理：需遵循明确规则（例如，缺失率 < 20% 时按组别中位数填充） 。 |
| 可靠性 | • 比对验证：数据应经过有效的比对，如室内、实验室间或国际比对 。• 第三方评价：数据或其产生方法经过第三方权威机构（如NIST）的评价或测评者优先 。• 质量控制：序列分析中QC样本占比需 ≥ 10%；采用Shewhart控制图等方法检测异常QC样本 。 |
| 2 | 准确性验证 (计量性) | 准确性 | • 质量轴准确性：内标物误差 < 5 ppm；目标代谢物误差 < 10 ppm；Orbitrap等高分辨质谱仪分辨率需 ≥ 120,000 。• 鉴定准确性：- 一级质谱：质量误差 ≤ 10 ppm 。- 二级质谱：与公共库（如GNPS）的匹配得分 ≥ 60分；碎片离子覆盖率 ≥ 50% 。- 保留时间：与标准品或数据库的偏差 ≤ 0.5 min 。- 需报告假阳性发现率（FDR） 。• 定量准确性：内标回收率需在 80%-120% 范围内；线性范围 R² ≥ 0.99；模型筛选差异代谢物时VIP值需 ≥ 1并通过置换检验 。 |
| 溯源性 | • 溯源至标准物质：定量数据的量值应能溯源至有证标准物质（CRMs/SRMs），并提供溯源性证明 。• 报告不确定度：必须评定并报告关键测量结果（如代谢物浓度）的测量不确定度 。 |
| 重复性 | • 数据重现性：QC样本中大部分代谢物峰面积的相对标准偏差（RSD）应满足要求（例如LC-MS中 ≤ 30%） 。• 稳定性：考察数据在规定时间内的稳定性，确保其在有效期内质量可靠 。 |
| 3 | 相关性 | 与现象、物体或物质特性的相关性 | • 数据应清晰地与特定的生物样本（如人血浆、尿液）及相关生物学现象建立关联 。 |
| 与已知化合物成分或结构系统的相关性 | • 数据应清晰地与特定的代谢物（已知化合物成分）建立关联 。• 数据应能服务于方法验证、仪器校准、算法模型开发等明确的科学或计量应用 。 |
| 4 | 其他内容 | 数字数据 | • 原始数据：采用 mzML 1.1.0 或 mzXML 3.2 格式，包含完整的扫描事件及仪器参数 。• 处理后数据：采用 Excel（.xlsx），字段包括：代谢物名称、CAS号、化学式、RT、MS1、MS2、定量值及单位 。• 元数据：采用 MIxS 格式的 XML 文件，包含样本、前处理、仪器等详细信息 。 |

表D.3 指导委员会评定检查清单

|  |  |  |  |
| --- | --- | --- | --- |
| 序号 | 评审类别 | 评审内容 | 具体评审规则 |
| 1 | 评审资料 | 技术委员会提交的评审资料 | • 技术委员会提交的评审资料需要有评审结果，评审过程的记录，以及评审证明 |
| 2 | 严格评价 (标准性) | 完整性 | • 数据应包含元数据、原始数据和处理后的数据，信息充分，没有缺项 。• 元数据：应遵循代谢组学领域的最低信息标准（如MIxS），详细记录样本来源、前处理方法、仪器型号参数、实验日期等 。• 测量程序：必须提供完整记录的SOP（标准操作程序），确保实验流程的可复现性 。• 辅助材料：应提供实验方案、质量控制报告、方法学验证报告及溯源性证明等 。 |
| 一致性 | • 原理一致性：数据结果应与公认的物理化学及生物学原理一致（例如，电喷雾正离子模式下的主要离子加和形式应与流动相体系相符） 。• 方法一致性：与其他独立分析方法获得的数据比对，结果应表现出良好的一致性 。• 数据处理一致性：- 峰对齐：保留时间偏差 ≤ 0.2 min；峰面积缺失值 ≤ 20% 。- 批次效应校正：采用基于QC样本的LOESS校正或ComBat算法，校正后批次间信号中位数偏移 ≤ 10% 。- 缺失值处理：需遵循明确规则（例如，缺失率 < 20% 时按组别中位数填充） 。 |
| 可靠性 | • 比对验证：数据应经过有效的比对，如室内、实验室间或国际比对 。• 第三方评价：数据或其产生方法经过第三方权威机构（如NIST）的评价或测评者优先 。• 质量控制：序列分析中QC样本占比需 ≥ 10%；采用Shewhart控制图等方法检测异常QC样本 。 |
| 3 | 准确性验证 (计量性) | 准确性 | • 质量轴准确性：内标物误差 < 5 ppm；目标代谢物误差 < 10 ppm；Orbitrap等高分辨质谱仪分辨率需 ≥ 120,000 。• 鉴定准确性：- 一级质谱：质量误差 ≤ 10 ppm 。- 二级质谱：与公共库（如GNPS）的匹配得分 ≥ 60分；碎片离子覆盖率 ≥ 50% 。- 保留时间：与标准品或数据库的偏差 ≤ 0.5 min 。- 需报告假阳性发现率（FDR） 。• 定量准确性：内标回收率需在 80%-120% 范围内；线性范围 R² ≥ 0.99；模型筛选差异代谢物时VIP值需 ≥ 1并通过置换检验 。 |
| 溯源性 | • 溯源至标准物质：定量数据的量值应能溯源至有证标准物质（CRMs/SRMs），并提供溯源性证明 。• 报告不确定度：必须评定并报告关键测量结果（如代谢物浓度）的测量不确定度 。 |
| 重复性 | • 数据重现性：QC样本中大部分代谢物峰面积的相对标准偏差（RSD）应满足要求（例如LC-MS中 ≤ 30%） 。• 稳定性：考察数据在规定时间内的稳定性，确保其在有效期内质量可靠 。 |
| 4 | 相关性 | 与现象、物体或物质特性的相关性 | • 数据应清晰地与特定的生物样本（如人血浆、尿液）及相关生物学现象建立关联 。 |
| 与已知化合物成分或结构系统的相关性 | • 数据应清晰地与特定的代谢物（已知化合物成分）建立关联 。• 数据应能服务于方法验证、仪器校准、算法模型开发等明确的科学或计量应用 。 |
| 5 | 其他内容 | 数字数据 | • 原始数据：采用 mzML 1.1.0 或 mzXML 3.2 格式，包含完整的扫描事件及仪器参数 。• 处理后数据：采用 Excel（.xlsx），字段包括：代谢物名称、CAS号、化学式、RT、MS1、MS2、定量值及单位 。• 元数据：采用 MIxS 格式的 XML 文件，包含样本、前处理、仪器等详细信息 。 |

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_