



中华人民共和国医药行业标准

XX/T XXXXX—XXXX

医疗器械真实世界数据 质量通用要求

Medical device real world data—General quality requirements

(点击此处添加与国际标准一致性程度的标识)

征求意见稿

本草案完成时间：2026.08.27

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

XXXX—XX—XX 发布

XXXX—XX—XX 实施

国家药品监督管理局 发布

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这些专利的责任。

本文件由国家药品监督管理局提出。

本文件由全国医疗器械临床评价标准化技术归口单位归口。

本文件起草单位：

本文件主要起草人：

医疗器械真实世界数据 质量通用要求

1 范围

本文件规定了医疗器械真实世界数据的质量通用要求，包括准备阶段质量、采集阶段质量和结果阶段质量三个维度共50个指标。

本文件适用于医疗器械临床评价真实世界数据。

2 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。

YY/T 2000 医疗器械真实世界数据 术语和定义

YY/T 1930 医疗器械临床评价 术语和定义

GB/T 36344 信息技术 数据质量评价指标

3 术语和定义

YY/T 2000-2026界定的以及下列术语和定义适用于本文件。

3.1 前瞻性数据 prospective data

在研究开始时确定相关人群，并于该时间点以后收集的暴露/治疗和结局等相关数据。

3.2 回顾性数据 retrospective data

在研究开始时，收集既往已经存在于医院电子病历等数据中的暴露/治疗和结局等相关数据。

3.3 数据质量 data quality

在指定条件下使用时，数据的特性满足明确的和隐含的要求的程度。

[来源：《信息技术 数据质量评价指标》，GB/T 36344—2018]

3.4 透明性 transparency

指数据收集方案、数据治理方案和治理过程是否公开，以及公开的程度。

[来源：《用于产生真实世界证据的真实世界数据指导原则(试行)》，三(三)2，有修订]

3.5 数据字典 data dictionary

数据收集过程中使用的共同定义框架。

[来源：《真实世界证据支持药物研发与审评的指导原则（试行）》，二(二)2，有修订]

4 真实世界数据质量评价指标

4.1 质量评价指标的分类和编码说明

根据真实世界数据采集方式的不同，数据质量评价指标分为前瞻性真实世界数据和回顾性真实世界数据质量评价两个部分。质量评价指标又进一步分为准备阶段质量指标、采集阶段质量指标和结果阶段质量指标三个维度。其中准备阶段质量指标关注策划与准备过程的数据质量；采集阶段质量指标关注在数据采集和治理过程中的数据质量；结果阶段质量指标关注已构建数据库的数据质量。

为了便于数据质量评价指标的识别和管理，本文件对各项评价指标赋予唯一性编码。该编码由一级指标代码、二级指标代码和三级指标代码共同构成，共6位数字。其中，前两位为一级指标代码，01代表前瞻性真实世界数据质量评价指标，02代表回顾性真实世界数据质量评价指标；中间两位为二级指标

代码，01代表准备阶段质量指标、02代表采集阶段质量指标、03代表结果阶段质量指标；最后两位为三级指标顺序码，编码范围从01至99。编码规则见图1。

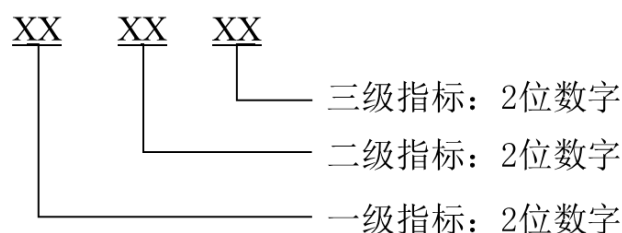


图1 编码规则

4.2 前瞻性真实世界数据质量评价指标

4.2.1 数据采集方案（编码：010101）

应提前制定数据采集方案及其配套的数据采集表，并对采集表中的变量进行解释说明。当数据采集方案和数据采集表通过伦理审查、方案注册或方案发表中的任意一项，且在数据采集开始前完成时，评价为“是”，否则为“否”。

注：关键变量通常包括器械使用或暴露信息、对照（未暴露）信息、目标结局以及重要混杂变量。在制定研究方案时，需根据研究目的和器械品种等特点确定研究的关键变量。

4.2.2 隐私保护（编码：010102）

应在数据采集前制定个人隐私保护措施，明确采集范围、去标识化以及访问权限管理要求。评价方法为核查研究方案及相关记录，隐私保护措施符合要求且与实际流程一致时，评价为“是”，否则为“否”。

4.2.3 伦理审批（编码：010103）

研究实施前，研究方案应获得伦理批准，覆盖研究设计、数据采集方案、参与者知情同意等内容。评价方法为核查批准文件及相关记录，评价为“是”，否则为“否”。

4.2.4 数据库测试（编码：010104）

数据库应在数据采集前完成测试，确认数据采集表、访视模块、逻辑核查及数据导出等功能可正常使用。评价方法为核查测试记录或测试报告，测试已完成并通过时，评价为“是”，否则为“否”。

4.2.5 数据库质控（编码：010105）

应在研究启动前制定覆盖数据采集、录入、储存、传输和使用等环节的质量控制措施，并在研究启动前、研究中、锁库前和锁库后各阶段落实。评价方法为核查质控计划及实施记录，各阶段质量控制措施均已落实且符合要求时，评价为“是”，否则为“否”。

4.2.6 数据库安全保护（编码：010106）

应在数据库启用前制定数据安全保护措施，涵盖身份认证、分级授权、操作日志、传输与存储及备份恢复等，并在数据收集和贮存过程中实施。评价方法为检查系统配置、权限清单及相关记录，数据安全保护措施明确且与实际运行情况相符，评价为“是”，否则为“否”。

4.2.7 数据库锁定计划（编码：010107）

应在研究启动前制定数据库锁定程序，明确锁定条件、质疑处置、解锁条件及审批责任等。评价方法为核查锁库计划及锁库前后的执行记录，各阶段要求均满足时，评价为“是”，否则为“否”。

4.2.8 研究参与者知情同意（编码：010201）

应在数据采集前按伦理批准的要求，取得研究参与者或其监护人的有效知情同意并保留记录。评价方法为核查知情同意文件、签署主体、日期、及退出处理记录，均符合要求时，评价为“是”，否则为“否”。

4.2.9 数据采集完整性（编码：010202）

应按数据采集方案完成采集表及访视模块内容的采集。评价方法为核查截至评价日已到期应完成的采集任务清单与实际记录，报告数据采集任务完成率，计算方法为：已完成全部必需采集内容的到期任务数÷已到期应完成的任务总数×100%。

4.2.10 数据采集及时性（编码：010203）

应预设访视或测量时间窗（如器械植入后30±7天），记录事件和采集日期，并报告时间窗数据采集及时率，计算方法为：在规定时间内完成的采集任务数÷截至评价日采集时间窗已结束且应完成的任务总数×100%；

4.2.11 数据备份（编码：010204）

应在数据采集过程中对原始数据进行备份，明确备份的内容、频率及方式，备份数据与原始数据分开存放，并开展恢复测试以确认备份数据可用。评价方法为核查备份记录，备份覆盖全部原始数据且分开保存时，评价为“是”，否则为“否”。

4.2.12 数据质疑与编码（编码：010301）

数据采集完成后、锁库前，所有质疑应已处理并留痕；无法解决的残留项应按锁定程序说明原因并经审批。评价方法为核查质疑清单、编码记录，质疑均已处理（或残留项已说明并获批）且编码完整时，评价为“是”，否则为“否”。

4.2.13 关键变量值完整性（编码：010302）

数据库中的关键变量应完整，缺失情况应按各变量以及不同访视时点分别进行报告。评价方法为计算数据库中每个关键变量值缺失的百分比，计算方法为：某关键变量的缺失数量÷数据库中的样本量×100%。

4.2.14 数据准确性（编码：010303）

应根据数据采集方案中对关键变量的测量方法、匹配规则等进行核验，可采用随机抽样方式开展核验，应报告每个关键变量的准确率，计算方法为：某关键变量符合预设准确性标准的数量÷数据库中的样本量×100%。

4.2.15 数据格式标准化（编码：010304）

应确保数据库中变量的类型、长度、单位、日期格式及编码等均符合定义且标准化。评价方法为核查数据库中变量的格式与数据采集方案的一致性，数据格式与方案规定一致时，评价为“是”，否则为“否”。

4.2.16 数据唯一性（编码：010305）

受试者编号在数据库内唯一，并与源记录中的对应标识一致；访视及表单记录应通过受试者编号、访视标识、表单标识及必要的重复实例标识进行唯一识别。评价时，核查唯一性约束规则及相关记录清单，并与源记录核对。满足上述规则，且不存在未处理的非预期重复记录时，评价为“是”；否则评价为“否”。

4.2.17 数据一致性（编码：010306）

原始记录数据与最终数据库记录应保持一致、同一字段数据在不同数据库之间应保持一致。当数据库录入的数据和原始数据保持一致时、同一字段数据在不同数据库之间保持一致时评价为“是”，否则为“否”。

4.2.18 数据稽查记录完整性（编码：010307）

数据库应具备完整、不可关闭的稽查功能，能追溯数据的创建、修改、删除及修改原因，并可按要求导出。评价方法为核查数据库的稽查记录，稽查记录完整且可追溯至具体操作者时，评价为“是”，否则为“否”。

4.2.19 数据库解锁（编码：010308）

确需解锁的数据库，应经书面说明理由和审批，评估其对分析结果的影响，解锁与再锁定过程均应留痕。评价方法为核查解锁申请与审批记录、影响评估及再锁定记录，报告第一次锁库后的累计解锁次数。

4.2.20 可访问（编码：010309）

数据库应保证授权人员在研究开展期间、研究结束直至数据保存期限内，可按其权限查阅数据。当授权人员可按规定权限获取且访问留痕时，评价为“是”，否则为“否”。

4.2.21 可使用性（编码：010310）

数据应在规定的保存和使用期限内保持可读取、可解析，并配套版本说明，且能被授权人员或机构使用。评价方法为数据库满足预设使用要求时，评价为“是”，否则为“否”。

4.3 回顾性真实世界数据质量评价指标

4.3.1 数据收集方案（编码：020101）

应根据研究目的提前制定数据收集方案，明确数据来源、人群纳排标准、时间范围、拟收集变量及其定义。评价方法为核查方案、版本与批准日期，内容完整并按方案执行时，评价为“是”，否则为“否”。

4.3.2 隐私保护（编码：020102）

应提前制定个人隐私保护措施，对去标识化、访问权限、传输与存储保护及使用限制等做出明确规定。评价方法为核查保护措施及相关记录，当具有实施依据且与实际流程一致时，评价为“是”，否则为“否”。

4.3.3 伦理审批（编码：020103）

应在数据提取前完成所需伦理审查。评价方法为核查审查文件及其适用范围、版本和日期，已取得批准或有符合规定的免除依据时，评价为“是”，否则为“否”。

4.3.4 研究参与者知情同意（编码：020104）

应在数据采集前，取得研究参与者或其监护人的有效知情同意并保留相关记录，或者取得伦理委员会批准的知情同意豁免。评价方法为核查知情同意文件、签署主体、日期及退出处理记录是否符合要求，或是否已取得伦理委员会批准的知情同意豁免；符合要求或已获豁免的，评价为“是”，否则为“否”。

4.3.5 可及性（编码：020105）

应取得源数据库使用授权，并支持监管核查。评价方法为核查授权文件与数据使用协议，当有授权文件且符合要求时，评价为“是”，否则为“否”。

4.3.6 源数据库关键变量覆盖度（编码：020106）

应在数据提取前，将研究方案所需的关键变量与源数据库中的变量逐项比对。评价结果为描述关键变量的缺失率，以及其处置方式。应报告关键变量缺失率，计算方法为：缺失的关键变量数÷按方案应有关键变量数×100%。

4.3.7 源数据库关键变量值缺失率（编码：020107）

源数据库的关键变量值缺失比例不应过高，应报告源数据库的关键变量值缺失率，计算方法为：该变量缺失的数据量÷源数据库样本量×100%。

4.3.8 样本量（编码：020108）

经纳排标准筛选后，源数据库中的研究人群应满足研究方案中预设的样本量，评价方法为比较经纳排后源数据库中的研究人群与研究预设样本量，达到要求时评价为“是”，否则为“否”。

4.3.9 代表性（编码：020109）

源数据库研究人群应与器械拟适用目标人群在关键变量特征上具有可比性。评价方法为比较源数据库关键变量特征并报告分布差异，达到预设且有依据的可比性要求时，评价为“是”，否则为“否”。

4.3.10 数据库安全保护（编码：020110）

应在源数据库启用前制定数据安全保护措施，涵盖身份认证、分级授权、操作日志、传输与存储及备份恢复等，并在数据收集和贮存过程中实施。评价方法为检查系统配置、权限清单及相关记录，数据安全保护措施明确且与实际运行情况相符，评价为“是”，否则为“否”。

4.3.11 数据治理计划（编码：020111）

应在数据治理实施前制定源数据库治理计划，明确数据提取与链接方法、清洗规则、数据结构化和标准化方法，以及版本管理等要求。评价方法为核查数据治理计划及执行记录，制定数据治理计划并按其执行时，评价为“是”，否则为“否”。

4.3.12 数据库锁定计划（编码：020112）

应在数据采集开始前制定数据库锁定程序，明确锁定条件、审批责任及解锁与再锁定流程。评价方法为核查锁库计划及执行记录，程序完整、已发生的锁库操作符合程序时，评价为“是”，否则为“否”。

4.3.13 数据备份（编码：020201）

应按预定频率和关键节点，对源数据库及治理过程数据进行备份，并验证恢复能力。评价方法为核查备份策略及恢复测试记录，备份按计划完成且恢复验证满足要求时，评价为“是”，否则为“否”。

4.3.14 数据库的可链接性（编码：020202）

当需要从多个源数据集或数据库中获取数据时，应预先规定数据链接规则，明确主库、链接变量、以及链接方式。评价方法为核查链接规则、链接变量的可用性 & 链接结果，报告链接成功率，计算方法为：按预设规则成功建立链接的记录数÷预先定义的应链接记录总数×100%。

4.3.15 数据库链接的正确性（编码：020203）

对于已开展的数据链接，应核验链接结果是否对应同一受试者或同一目标事件。评价方法为依据源记录或其他可靠参考信息进行抽样核验，报告链接正确率，计算方法为：抽样核验中确认链接正确的记录数÷抽样核验的已链接记录总数×100%。

4.3.16 数据编码（编码：020204）

数据库中术语应规范统一、完整，治理过程中应按预先规定的字典（注明版本）将需要编码的字段统一映射为标准编码。评价方法为核查编码字典及映射规则，术语统一、编码完整时，评价为“是”，否则为“否”。

4.3.17 矛盾数据处理（编码：020205）

应按数据治理方案处理矛盾数据，记录识别规则、来源优先级、判定依据及处理前后值。评价方法为交叉核对处理规则、问题清单及变更日志，规则清楚、处理方式具有实践依据时，评价为“是”，否则为“否”。

4.3.18 缺失值处理（编码：020206）

应按数据治理方案处理缺失值，说明删除、保留或插补的规则与理由；采用插补应记录方法、参数及处理代码。评价方法为核查方案描述、程序及处理记录，能区分原始值与插补值、可追溯处理过程时，评价为“是”，否则为“否”。

4.3.19 数据字典（编码：020301）

应建立与当前数据库版本一致的数据字典，包含变量的名称、标签、类型、单位、编码、缺失值含义以及衍生变量的生成规则。评价方法为逐项核对实际字段清单，信息完整且与实际一致时，评价为“是”，否则为“否”。

4.3.20 治理后数据库的关键变量值完整性（编码：020302）

治理后数据库的关键变量值缺失比例不应过高，应报告治理后数据库的各关键变量值缺失率，计算方法为：该变量缺失的数据量÷治理后数据库样本量×100%。

4.3.21 数据准确性（编码：020303）

数据库中数据应与源数据库记录一致，关键变量应符合预设的临床定义，数据转换及计算应正确。评价方法为抽样核对源数据库记录或可靠参考标准，报告核验准确率及错误处理情况，达到预设准确性标准时，评价为“是”，否则为“否”。

4.3.22 数据质疑与编码（编码：020304）

数据治理完成后、锁库前，所有质疑应已处理并留痕；无法解决的残留项应按锁定程序说明原因并经审批。评价方法为核查质疑清单和编码记录，质疑均已处理（或残留项已说明并获批）且编码完整时，评价为“是”，否则为“否”。

4.3.23 数据格式标准化（编码：020305）

应确保数据库中变量的类型、长度、单位、日期格式及编码等均符合定义且标准化。评价方法为核查数据库中变量的格式与数据收集方案的一致性，数据格式与方案规定一致时，评价为“是”，否则为“否”。

4.3.24 数据唯一性（编码：020306）

受试者编号在库内唯一，并与源数据库记录中的对应标识一致；访视及表单记录应通过受试者编号、访视标识、表单标识及必要的重复实例标识进行唯一识别。评价时，核查唯一性约束规则及相关记录清单，并与源数据库记录核对。满足上述规则，且不存在未处理的非预期重复记录时，评价为“是”，否则评价为“否”。

4.3.25 数据一致性（编码：020307）

源数据库与治理后数据库记录保持一致、同一字段数据在不同数据库之间应保持一致，当治理后的数据库记录与源数据库记录保持一致、同一字段数据在不同数据库之间保持一致时评价为“是”，否则评价为“否”。

4.3.26 数据治理可溯源性（编码：020308）

从源数据库到治理后数据库的各处理环节均应有记录，可追溯到数据治理中的具体方法、代码、时间及治理前后值，并可复现治理结果。评价方法为核查治理日志及相关程序，治理结果可复现时，评价为“是”，否则为“否”。

4.3.27 数据库解锁（编码：020309）

确需解锁的源数据库和治理后数据库，应经书面说明理由和审批，评估其对分析结果的影响，解锁与再锁定过程均应留痕。评价方法为核查解锁申请与审批记录、影响评估及再锁定记录，报告第一次锁库后的累计解锁次数。

4.3.28 可访问（编码：020310）

源数据库和治理后数据库应保证授权人员在研究开展期间、研究结束直至数据保存期限内，可按其权限查询和查阅数据。当授权人员可按规定权限获取且访问留痕时，评价为“是”，否则为“否”。

4.3.29 可用性（编码：020311）

源数据库和治理后数据库应在规定的保存和使用期限内保持可读取、可解析，并配套版本说明，且能被授权人员或机构使用。评价方法为数据库满足预设使用要求时，评价为“是”，否则为“否”。

5 质量评价指标的使用

在以医疗器械临床评价为目的的真实世界数据库构建完成后，包括前瞻性真实世界数据和回顾性真实世界数据，在数据质量评价过程中，可以使用前瞻性真实世界数据质量评价表（附录A）或回顾性真实世界数据质量评价表（附录B）对数据质量进行评价，评价流程见图2。

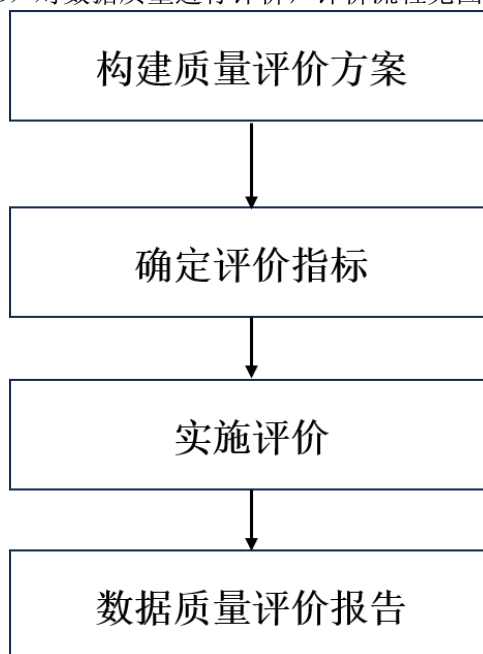


图2 真实世界数据质量评价过程

附录 A
前瞻性真实世界数据质量评价表

指标编号	指标名称	指标描述	评价结果
010101	数据采集方案	应提前制定数据采集方案及其配套的数据采集表，并对采集表中的变量进行解释说明	☐ 是 ☐ 否
010102	隐私保护	应在数据采集前制定个人隐私保护措施，明确采集范围、去标识化及访问权限管理要求	☐ 是 ☐ 否
010103	伦理审批	研究实施前，研究方案应获得伦理批准，覆盖研究设计、数据采集方案、参与者知情同意等内容	☐ 是 ☐ 否
010104	数据库测试	数据库应在数据采集前完成测试，确认采集表、访视模块、逻辑核查及数据导出等功能可正常使用	☐ 是 ☐ 否
010105	数据库质控	应在研究启动前制定覆盖数据采集、录入、储存、传输和使用等质量控制措施，并在各阶段落实	☐ 是 ☐ 否
010106	数据库安全保护	应在数据库启用前制定数据安全保护措施，并在数据收集和贮存过程中实施	☐ 是 ☐ 否
010107	数据库锁定计划	应在研究启动前制定数据库锁定程序，明确锁定条件、质疑处置、解锁条件及审批责任等	☐ 是 ☐ 否
010201	研究参与者知情同意	应在数据采集前按伦理批准的要求，取得研究参与者或其监护人的有效知情同意并保留记录	☐ 是 ☐ 否
010202	数据采集完整性	应按数据采集方案完成采集表及访视模块的内容	☐ 是 ☐ 否
010203	数据采集及时性	应预设访视或测量时间窗，并记录事件和采集日期，报告数据及时采集的比例	_____ %
010204	数据备份	应在数据采集过程中对原始数据进行备份，明确备份的内容、频率及方式，并开展恢复测试	☐ 是 ☐ 否
010301	数据质疑与编码	数据采集完成后、锁库前，所有质疑应已处理并留痕；残留项应按锁定程序说明原因并经审批	☐ 是 ☐ 否
010302	关键变量值完整性	数据库中的关键变量应完整，缺失情况应按各变量以及不同访视时点分别进行报告	_____ %
010303	数据准确性	应根据数据采集方案中对关键变量的测量方法、匹配规则等进行核验，报告准确率	_____ %
010304	数据格式标准化	应确保数据库中变量的类型、长度、单位、日期格式及编码等均符合定义且标准化	☐ 是 ☐ 否

010305	数据唯一性	受试者编号在库内唯一，并与源记录中的对应标识一致；访视及表单记录进行唯一标识	☐ 是 ☐ 否
010306	数据一致性	原始记录数据与最终数据库记录应保持一致、同一字段数据在不同数据库之间应保持一致	☐ 是 ☐ 否
010307	数据稽查记录完整性	数据库应具备完整、不可关闭的稽查功能，能追溯数据的创建、修改、删除及修改原因，并可导出	☐ 是 ☐ 否
010308	数据库解锁	确需解锁的数据库，应经书面说明理由和审批，评估其对分析结果的影响，解锁与再锁定过程均应留痕，报告第一次锁库后的累计解锁次数	_____次
010309	可访问	数据库应保证授权人员在研究开展期间、研究结束直至数据保存期限内，可查询和查阅数据	☐ 是 ☐ 否
010310	可使用性	数据应在规定的保存和使用期限内保持可读取、可解析、可使用，并配套版本说明	☐ 是 ☐ 否

附录 B
回顾性真实世界数据质量评价表

指标编号	指标名称	指标描述	评价结果
020101	数据收集方案	应根据研究目的制定数据收集方案，明确数据来源、纳排标准、时间范围、拟收集变量及其定义	☐ 是 ☐ 否
020102	隐私保护	应提前制定个人隐私保护措施，对去标识化、访问权限、传输与存储保护及使用限制等明确规定	☐ 是 ☐ 否
020103	伦理审批	应在数据提取前完成所需伦理审查	☐ 是 ☐ 否
020104	研究参与者知情同意	应在数据采集前，取得研究参与者或其监护人的有效知情同意并保留相关记录，或者取得伦理委员会批准的知情同意豁免	☐ 是 ☐ 否
020105	可及性	应取得源数据库使用授权，并支持监管核查	☐ 是 ☐ 否
020106	源数据库关键变量覆盖度	应在数据提取前，将研究方案所需的关键变量与源数据库中的变量逐项比对，报告缺失率	_____%
020107	源数据库关键变	源数据库的关键变量值缺失比例不应过高，应报	_____%

	量值缺失率	告源数据库的关键变量值缺失率	
020108	样本量	经纳排标准筛选后，源数据库中的研究人群应满足研究方案中预设的样本量	☐ 是 ☐ 否
020109	代表性	源数据库研究人群应与器械拟适用目标人群在关键变量特征上具有可比性	☐ 是 ☐ 否
020110	数据库安全保护	应在源数据库启用前制定数据安全保护措施，并在数据收集和贮存过程中实施	☐ 是 ☐ 否
020111	数据治理计划	应在数据治理实施前制定源数据库治理计划，明确数据提取与链接方法、清洗规则、数据结构化和标准化方法，以及版本管理等要求	☐ 是 ☐ 否
020112	数据库锁定计划	应在数据采集开始前制定数据库锁定程序，明确锁定条件、审批责任及解锁与再锁定流程	☐ 是 ☐ 否
020201	数据备份	应按预定频率和关键节点，对源数据库及治理过程数据进行备份，并验证恢复能力	☐ 是 ☐ 否
020202	数据库的可链接性	应预先规定数据链接规则，明确主库、链接变量、以及链接方式，报告链接成功率	_____ %
020203	数据库链接的正确性	应核验链接结果是否对应同一受试者或同一目标事件，报告链接正确率	_____ %
020204	数据编码	数据库中术语应规范统一、完整，治理过程中应按数据字典将字段统一映射为标准编码	☐ 是 ☐ 否
020205	矛盾数据处理	应按数据治理方案处理矛盾数据，记录识别规则、来源优先级、判定依据及处理前后值	☐ 是 ☐ 否
020206	缺失值处理	应按数据治理方案处理缺失值，说明删除、保留或插补的规则与理由	☐ 是 ☐ 否
020301	数据字典	应建立与当前数据库版本一致的数据字典，包含变量的名称、标签、类型、单位、编码、缺失值含义以及衍生变量的生成规则	☐ 是 ☐ 否
020302	治理后数据库的关键变量值完整性	应报告治理后数据库的各关键变量值缺失率	_____ %
020303	数据准确性	数据库中数据应与源记录一致，关键变量应符合预设的临床定义，数据转换及计算应正确	☐ 是 ☐ 否

020304	数据质疑与编码	数据治理完成后、锁库前，所有质疑已处理并留痕	☐ 是 ☐ 否
020305	数据格式标准化	应确保数据库中变量的类型、长度、单位、日期格式及编码等均符合定义且标准化	☐ 是 ☐ 否
020306	数据唯一性	受试者编号在库内唯一，并与源记录中的对应标识一致；访视及表单记录进行唯一标识	☐ 是 ☐ 否
020307	数据一致性	源数据库与治理后数据库记录保持一致、同一字段数据在不同数据库之间应保持一致	☐ 是 ☐ 否
020308	数据治理可溯源性	从源数据库到治理后数据库的各处理环节均应有记录并可复现治理结果	☐ 是 ☐ 否
020309	数据库解锁	确需解锁的源数据库和治理后数据库，应经书面说明理由和审批，报告第一次锁库后的累计解锁次数	_____次
020310	可访问	源数据库和治理后数据库应保证授权人员在研究开展期间、研究结束直至数据保存期限内，可按其权限查询和查阅数据	☐ 是 ☐ 否
020311	可使用性	源数据库和治理后数据库应在规定的保存和使用期限内可读取、可解析、可使用，并配套版本说明	☐ 是 ☐ 否

参考文献

- [1] WAND Y, WANG R Y. Anchoring Data Quality Dimensions in Ontological Foundations[J]. Commun. ACM, 1996, 39(11): 86-95.
- [2] Duke-Margolis Center for health policy. Characterizing RWD Quality and Relevancy for Regulatory Purposes [EB/OL]. [2023-11-10]. <https://healthpolicy.duke.edu/publications/characterizing-rwd-quality-and-relevancy-regulatory-purposes-0>.
- [3] WANG S V, SCHNEEWEISS S, BERGER M L, et al. Reporting to Improve Reproducibility and Facilitate Validity Assessment for Healthcare Database Studies V1.0[J]. Value Health, 2017, 20(8): 1009-1022.
- [4] WEISKOPF N G, WENG C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research[J]. J Am Med Inform Assoc, 2013, 20(1): 144-151.
- [5] BLACKETER C, DEFALCO F J, RYAN P B, et al. Increasing trust in real-world evidence through evaluation of observational data quality[J]. Journal of the American Medical Informatics Association: JAMIA, 2021, 28(10): 2251-2257.
- [6] VERMA A A, PASRICHA S V, JUNG H Y, et al. Assessing the quality of clinical and administrative data extracted from hospitals: the General Medicine Inpatient Initiative (GEMINI) experience[J]. J Am Med Inform Assoc, 2021, 28(3): 578-587.

- [7] NAIK S, VOONG S, BAMFORD M, et al. Assessment of the Nursing Quality Indicators for Reporting and Evaluation (NQuIRE) database using a data quality index[J]. J Am Med Inform Assoc, 2020,27(5): 776-782.
- [8] BIAN J, LYU T, LOIACONO A, et al. Assessing the practice of data quality evaluation in a national clinical data research network through a systematic scoping review in the era of real-world data[J]. J Am Med Inform Assoc, 2020,27(12): 1999-2010.
- [9] RAJAN N S, GOURIPEDDI R, MO P, et al. Towards a content agnostic computable knowledge repository for data quality assessment[J]. Comput Methods Programs Biomed, 2019,177: 193-201.
- [10] BUCHHEIT R B, GARRETT J H, MCNEIL S, et al. Automated Procedure to Assess Civil Infrastructure Data Quality: Method and Validation[J]. Journal of Infrastructure Systems, 2005,11(3): 180-189.
- [11] REIMER A P, MILINOVICH A, MADIGAN E A. Data quality assessment framework to assess electronic medical record data for use in research[J]. International journal of medical informatics, 2016,90: 40-47.
- [12] EVANS P. Scaling and assessment of data quality[J]. Acta crystallographica. Section D, Biological crystallography, 2006,62(Pt 1): 72-82.
- [13] LIAW S, GUO J G N, ANSARI S, et al. Quality assessment of real-world data repositories across the data life cycle: A literature review[J]. Journal of the American Medical Informatics Association: JAMIA, 2021,28(7): 1591-1599.
- [14] BLACK N, PAYNE M. Directory of clinical databases: improving and promoting their use[J]. Qual Saf Health Care, 2003,12(5): 348-352.
- [15] KAHN M G, RAEBEL M A, GLANZ J M, et al. A pragmatic framework for single-site and multisite data quality assessment in electronic health record-based clinical research[J]. Med Care, 2012,50 Suppl(0): S21-S29.
- [16] COOK J A, COLLINS G S. The rise of big clinical databases[J]. Br J Surg, 2015,102(2): e93-e101.
- [17] World Health Organization. Data quality assurance (DQA)[EB/OL]. [2023-12-05]. <https://www.who.int/data/data-collection-tools/health-service-data/data-quality-assurance-dqa>.
- [18] 王雯, 高培, 吴晶, 宣建伟, 贺小宁, 胡明, 李洪超, 窦丰满, 于川, 闫盈盈, 孙鑫, 代表中国真实世界数据与研究联盟 (ChinaREAL). 构建基于既有健康医疗数据的研究型数据库技术规范. 中国循证医学杂志, 2019, 19(7): 763-770.
- [19] 谭婧, 彭晓霞, 舒啸尘, 王丽, 黎国威, 张玲, 毛琛, 郭新峰, 孙鑫, 代表中国真实世界数据与研究联盟 (ChinaREAL). 患者登记数据库构建技术规范. 中国循证医学杂志, 2019, 19(7): 771-778.
- [20] Heesen P, Schelling G, Birbaumer M, et al. Real-World-Time Data and RCT Synergy: Advancing Personalized Medicine and Sarcoma Care through Digital Innovation. *Cancers (BaseI)*. 2024;16(14):2516.