ICS 11.120
CCS C 07

团　　体　　标

T/CRHA XXX—2024

淋巴瘤临床研究元数据模型及数据质量评估规范

Meta Data Model and Data Quality Evaluation Specification for Clinical Research of Lymphoma

（征求意见稿）

2024-XX-XX 发布　　2024-XX-XX 实施

中国研究型医院学会　发　布

引　　言

为利用真实世界研究的数据库和方法学研究来加速药品的上市，旨在为医疗机构淋巴瘤临床研究提供一套术语规范、定义明确、语义语境无歧义的标准，以提高临床研究的数据质量，实现临床研究患者相关临床数据在发布、交换等一致性和可比性，推动现有医疗机构临床研究在不同医疗机构、不同软件系统和不同区域之间进行交换和共享。

本标准基于已有的淋巴瘤研究标准数据集，构建淋巴瘤数据质量评估制度、原则和评测标准体系，为后续的临床研究和真实世界研究数据库的建立提供技术支持，从而推进行业规范与标准。

开展淋巴瘤临床研究数据集的数据质量评估的目的是构建淋巴瘤数据质量评估制度、原则和评测标准体系，为后续的临床研究和真实世界研究数据库的建立提供技术支持，从而推进行业规范与标准。本文件基于《国家卫健委淋巴瘤标准数据集（2022 版 V2.0）》规定的元数据模型构建标准研究数据库，从多家医疗机构收集研究数据，并对数据集内容质量进行评估评价，从疗效风险获益的角度向监管准入等部门提供更具说服力的证据，对于药品全生命周期管理具有重要的价值。

本文件受北京大学医学部生物统计系侯艳教授主持科技部国家重点研发计划项目(2021YFF0901401)《基于医学本体的监管数据建模与分析技术》(2021/12-2024/11)支持。

淋巴瘤临床研究数据集的数据质量评估规范

1　范围

本文件规定了基于淋巴瘤临床研究数据集构建数据库的术语与定义、数据质量评估对象、评估指标框架、评估指标说明、评估方法、评估流程、数据结果分析评级及质量报告。

本文件适用于开展淋巴瘤研究的规范数据集的创建，通过数据的集中收集、处理、清洗和质量评估评价，建立行业普遍认同的高质量淋巴瘤研究中心数据库。

2　规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅注日期的版本适用于本文件。凡是不注日期的引用文件，其最新版本（包括所有的修改版）适用于本文件。

GB/T 36344—2018 信息技术　数据质量评价指标

DB11/T XXXX—XXXX《恶性肿瘤临床研究数据集 第 4 部分：淋巴瘤》

3　术语及定义

下列术语和定义适用于本文件。

3.1

淋巴瘤临床研究数据集

淋巴瘤临床研究数据集是淋巴瘤真实世界研究领域统一、规范的数据元标准，该数据集以《国家卫健委淋巴瘤标准数据集（2022 版 V2.0）》《中国临床肿瘤学会（CSCO）淋巴瘤诊疗指南（2021 版）》《NCCN 临床实践指南：慢性淋巴细胞白血病/小淋巴细胞淋巴瘤（2021.V1）》及《NCCN 临床实践指南：巨球蛋白血症/淋巴浆细胞性淋巴瘤（2021.V1）》等为主要参考，结合医院特点自定义模块及数据元，从而建立高质量的专科领域治疗体系。

数据集包含患者人口学信息、就诊记录、现病史、既往史、个人史、家族史、体格检查、诊断信息、实验室检查、内镜检查、超声检查、影像学检查、病理及免疫组化、骨髓检查、药物治疗、手术治疗、放射治疗、疗效评价、随访、样本库等 20 个模块，共计 888 个变量。为方便数据提取，数据集针对每一个数据项进行详细的原始数据来源映射和规则匹配，为进一步提升淋巴瘤诊疗水平、造福患者打好数据基础。

3.2

淋巴瘤临床研究数据集的数据质量维度

针对淋巴瘤临床研究数据集进行科学评测的数据质量特性。

3.3

淋巴瘤临床研究数据集的数据质量评估

基于淋巴瘤临床研究数据集的数据质量维度对淋巴瘤临床研究数据反映业务真实情况且满足应用要求的程度进行诊断和量化评价的活动。

3.4

淋巴瘤临床研究数据的规范性

衡量淋巴瘤临床研究数据符合数据标准、数据模型、业务规则、元数据或权威参考数据的程度。

3.5

淋巴瘤临床研究数据的准确性

衡量淋巴瘤临床研究数据与客观事实的符合程度。

3.6

淋巴瘤临床研究数据的完整性

衡量淋巴瘤临床研究数据具备或保持应有信息的程度。

3.7

淋巴瘤临床研究数据的可用性

衡量淋巴瘤临床研究数据可供使用的时间。

3.8

淋巴瘤临床研究数据的一致性

衡量具通过不同途径获取淋巴瘤临床研究数据的一致程度。

3.9

淋巴瘤临床研究数据可访问性

衡量淋巴瘤临床研究数据能被访问的程度。

3.10

淋巴瘤临床研究数据时效性

衡量淋巴瘤临床研究数据产生和流转满足管理和使用时效要求的程度即从事实发生到数据入库可用的时间间隔。

4 评估对象

本文件以淋巴瘤临床研究数据集各子域包含数据元为评估对象，子域类别及数据元引用 DB11/T XXXX—XXXX《恶性肿瘤临床研究数据集 第 4 部分:淋巴瘤》数据集由 15 个子域组成，见表 1，数据元明细见附录 A。

表1 淋巴瘤临床研究数据集子域

域号	名称	定义
FA	研究项目基本情况	临床研究名称、方案版本号、临床研究中心等项目基本信息。
RZ	入组及知情同意	受试者临床研究入组筛查及知情同意相关信息。
RK	人口学资料	受试者年龄、性别、民族等基本人口社会学信息。

JW	个人家族及既往疾病情况	受试者个人史、社会史及既往史信息。
TC	一般状况及体格检查	受试者筛查期及访视期体格检查信息。
ZD	淋巴瘤诊断	受试者恶性肿瘤诊断及分期信息，其中恶性肿瘤指符合临床试验入选标准、为该研究主要治疗目标的肿瘤，不包括受试者既往患有的或合并诊断的恶性肿瘤。
ZL	淋巴瘤既往治疗史	受试者临床研究入组前恶性肿瘤的相关治疗信息。
SY	干预措施	受试者参与临床研究期间试验性治疗相关信息。
HB	合并用药	受试者参与临床临床研究期间合并用药相关信息。
JY	试验室检验及标本采集	受试者临床研究筛选及入组期间生物样本采集和实验室检验信息。
JC	医学检查	受试者参与临床研究期间影像学、内镜及功能检查（如心电图、肺功能等）大型仪器设备检查信息。
PX	疗效评价	受试者临床试验期间疗效评价信息。
FZ	不良反应	受试者临床试验期间不良反应信息。
YH	随访及预后	受试者随访及长期预后信息，包括临床试验出组后的后续治疗信息。

5　数据质量评估指标框架

本指标框架主要参考《GB/T 36344—2018 信息技术　数据质量评价指标》中“3 数据治理评价指标框架”为指导，结合淋巴瘤临床研究数据集的特点，制定本规范的指标框架。

数据质量评估指标框架如图 1 所示。评估指标包括以下内容

①内容质量：包括数据规范性、准确性、完整性和可用性；

②过程质量：包括数据处理过程的处理效果、一致性；

③效用质量：包括数据可访问性、时效性；

④运维质量：包括数据管理员、操作员的专业程度以及数据的可维护性。

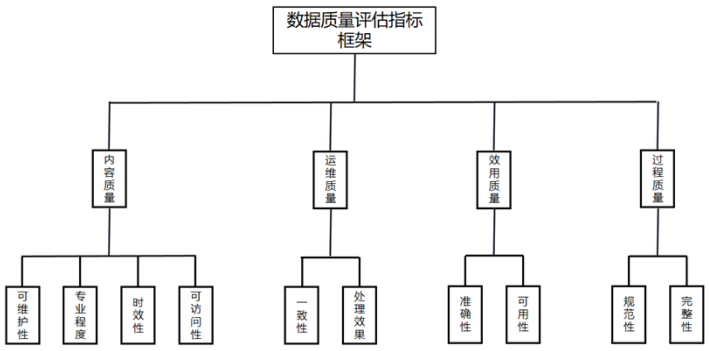


图1　质量评估指标框架图

6　数据质量评估指标说明

具体说明见表 2。

表1　数据质量评估指标说明

一级指标	二级指标	三级指标	指标描述
内容质量	规范性	命名规范性	命名规范性 数据库、数据集、数据元的命名方式符合相关命名规范的情况，即命名与标准数据集相同的变量在全部变量中的比例。
		元数据规范性	元数据描述符合相关规范的情况，即变量内容与标准数据集一致的比例。
		参考数据规范性	参考数据符合既有格式及规范的情况即参考数据格式与标准数据集要求一致的比例。（按需，如无参考数据则按满分值计算）
		数据权限规范性	是否基于相关法律法规制定数据安全权限的规范性文件。
		敏感字段脱敏占比	已脱敏字段占全部需要脱敏字段的比例（按需，如不需要数据脱敏则按满分值计算）。
	准确性	数据格式合规性	数据格式(数据类型、数据范围、数据长度、精度、编码等)是否满足预期要求，如姓名、性别、出生日期、诊断日期、国籍、民族等。具体体现为满足格式要求的数据集字段数占总数据集的字段总数的比例。
		数据重复率	数据记录中的重复数据、冗余数据的比例。
		数据唯一率	数据记录中唯一且正确数据的比例。
		脏数据出现率	错误或不准确的数据通常被称为“脏数据”，脏数据出现率低于阈值的数据集的比例。
	完整性	数据必填字段空值率	对于元数据模型规定的必填字段，存在值为空的数据记录比例。
		数据记录空值率	数据记录中存在空值的比例。
		数据记录模块缺失率	存在数据记录未覆盖到的模块（数据块/表）范围的记录比例。
	可用性	数据集有效可用时间占比	数据集可以被有效访问的年小时总数占比，即有效运行小时数/年总小时数。
		数据集维护时间占比	数据集处于人为计划性维护而无法被访问的年小时总数占比，即维护小时数/年总小时数。
		数据集失效时间占比	数据集因非维护性原因而无法被访问的年小时总数占比，即失效小时数/年总小时数。
		数据集可靠性评价	数据具备容灾、备份恢复能力的指标评价。主要从容灾能力，备份能力和恢复响应时效机制等三方面评价。
		数据合格率	校验后符合数据定义和规则的数据与总体数据的比例。

过程质量	处理效果		
		清洗保留率	清洗后的数据量占清洗前数据总量的比例。
		数据审核层级	较为完善的数据审核包括录入自查、上级审查和部门核查三级，根据实际采用的数据审核层级进行评价。
	一致性	相同数据一致性	同一数据在不同位置存储或被不同应用或用户使用时，数据的一致情况，即存储逻辑一致性。
		关联数据一致性	关联数据之间的逻辑关系正确和完整的数据的比例。
效用质量	可访问性	数据字段可访问率	可访问的数据字段的数量与总数的比例。
		数据记录可访问率	可访问的数据记录的数量与数据记录总数的比例。
		数据接口有效性	可访问接口正确性比例（如无外部接口调用，则按满分值计算）。
	时效性	时段数据正确性	日期范围的记录或者频率分布符合业务需求的情况。
		时点数据正确性	特定时点的记录数、频率分布或延迟时间符合业务需求的情况。
		数据时序正确性	数据时序的正确性，即有时序要求的变量，符合时序的数据占总数据的比例。（如无时序性数据，则按满分值计算）
运维质量	专业程度	数据库管理员专业程度	数据库管理员拥有最高权限，实现所有的设置和操作，对于数据库进行定时优化，对其他使用用户进行分配管控和清理，其专业性直接影响数据库的各项性能。通过执业资格、专业业务熟练度、合作态度评价等来判定数据库管理员的专业程度。
		审核人员专业程度	数据库审核人员通过其专业性来判定录入数据是否满足入库条件，直接影响数据内容质量。通过工作经验来判定其专业程度。
		数据录入人员专业程度	数据录入人员根据原始资料将数据录入数据库，其业务熟练度影响数据的真实性、准确性和完整性。通过工作经验来判定其专业程度。
	可维护性	维护难易程度	维护人员理解、改正、改动和改进系统的难易程度。易理解、易测试性、易修改。
		贯标难易程度	可以很好的适应对于标准数据集的版本变化，及时调整数据库结构。

7 数据质量评估流程

淋巴瘤临床研究数据集质量评估应参照本文件中规定的流程执行。

第一步：内容质量评估，评估淋巴瘤临床研究数据集子域及数据元是否规范、完整、准确、可用，依据评分标准（见表 3）按照评价指标内容质量部分对子域和数据元逐项评分；

第二步：过程质量评估，评估淋巴瘤临床研究数据集数据元一致性和处理效果，依据评分标准（见表 3）按照评价指标过程质量部分对子域和数据元逐项评分；

第三步：效用质量评估，评估淋巴瘤临床研究数据集的可访问性和时效性，依据评分标准（见表 3）按照评价指标效用质量部分逐项评分；

第四步：运维质量评估，评估淋巴瘤临床研究数据集运维及贯标的难易程度，依据评分标准（见表 3）对评价指标运维质量部分逐项评分；

第五步：汇总数据集各部分质量评估分值，参照质量级别评定表（见表 4），判定数据集质量级别

8 数据结果分析评级及质量报告

8.1 数据结果分析评级

评估后要对评估结果进行分析：

1）对评估目标与结果进行对比分析，确定是否达到评估指标；具体说明见表 3。

表2 数据质量评估指标评分表

一级指标	二级指标	三级指标	分值	评分标准
内容质量	规范性	命名规范性	4	0 分：<60% 1 分：60%-70% 2 分：70%-80% 3 分：80%-90% 4 分：>90%
		元数据规范性	4	0 分：<60% 1 分：60%-70% 2 分：70%-80% 3 分：80%-90% 4 分：>90%
		参考数据规范性	3	0 分：<70% 1 分：70%-80% 2 分：80%-90% 3 分：>90%
		数据权限规范性	3	有验证文件 0.5 分，文件完善 1 分 有应急预案 0.5 分，文件完善 1 分 有 SOP0.5 分，文件完善 1 分
		敏感字段脱敏占比	3	0 分：<70%

				1 分：70%-80% 2 分：80%-90% 3 分：>90%
	准确性	数据格式合规性	4	0 分：<60% 1 分：60%-70% 2 分：70%-80% 3 分：80%-90% 4 分：>90%
		数据重复率	3	0 分：>15% 1 分：10%-15% 2 分：5%-10% 3 分：<5%
		数据唯一率	4	0 分：<80% 1 分：80%-85% 2 分：85%-90% 3 分：90%-95% 4 分：>95%
		脏数据出现率	3	0 分：<70% 1 分：70%-80% 2 分：80%-90% 3 分：>90%
	完整性	数据必填字段空值率	3	0 分：>30% 1 分：20%-30% 2 分：10%-20% 3 分：<10%
		数据记录空值率	3	0 分：>30% 1 分：20%-30% 2 分：10%-20% 3 分：<10%

		数据记录模块缺失率	3	0 分：>30% 1 分：20%-30% 2 分：10%-20% 3 分：<10%
	可用性	数据集有效可用时间占比	3	0 分：<70% 1 分：70%-80% 2 分：80%-90% 3 分：>90%
		数据集维护时间占比	3	0 分：>30% 1 分：20%-30% 2 分：10%-20% 3 分：<10%
		数据集失效时间占比	2	0 分：>20% 1 分：10%-20% 2 分：<10%
		数据集可靠性评价	2	有灾备方案 0.5 分，文件完善 1 分 12h 以上内恢复响应 0.5 分，12h 以内 1 分
过程质量	处理效果	数据合格率	5	0 分：<50% 1 分：50%-60% 2 分：60%-70% 3 分：70%-80% 4 分：80%-90% 5 分：>90%
		清洗保留率	5	0 分：<50% 1 分：50%-60% 2 分：60%-70% 3 分：70%-80% 4 分：80%-90% 5 分：>90%

		数据审核层级	3	有自查 1 分 有上级审查 1 分 有部门核查 1 分
	一致性	相同数据一致性	4	0 分：<60% 1 分：60%-70% 2 分：70%-80% 3 分：80%-90% 4 分：>90%
		关联数据一致性	3	0 分：<70% 1 分：70%-80% 2 分：80%-90% 3 分：>90%
效用质量	可访问性	数据字段可访问率	3	0 分：<70% 1 分：70%-80% 2 分：80%-90% 3 分：>90%
		数据记录可访问率	3	0 分：<70% 1 分：70%-80% 2 分：80%-90% 3 分：>90%
		数据接口有效性	3	0 分：<70% 1 分：70%-80% 2 分：80%-90% 3 分：>90%
	时效性	时段数据正确性	4	记录部分符合 1 分 记录完全符合 2 分 频率部分符合 1 分 频率完全符合 2 分
		时点数据正确性	4	记录符合 1 分

				频率符合 1 分 延迟时间符合 1 分 三项均符合 1 分
		数据时序正确性	3	0 分：<70% 1 分：70%-80% 2 分：80%-90% 3 分：>90%
运维质量	专业程度	数据库管理员专业程度	3	有 GCP 证书 1 分 有经验 1 分 合作态度良好 1 分
		审核人员专业程度	2	有内审经验 1 分 有外审经验 1 分
		数据录入人员专业程度	1	0 分：无经验 1 分：有经验
	可维护性	维护难易程度	2	可测试 1 分 可修改 1 分
		贯标难易程度	2	0 分：无版本控制 1 分：有版本控制，不能升级 2 分：有版本控制，可升级

2）对评估的方案的有效性进行分析，确认是不是合适等。根据评估结果确定对象的质量评价，根据评价结果鉴定质量级别。确定评价对象的质量级别是建立在相应的质量分级方案基础上的是判断数据质量成熟度的重要依据。质量级别评定标准见表 4。

表1　数据质量级别评定表

序号	**评价结果**	**质量级别**
1	Q≥85	高质量权威数据集
2	85>Q≥70	可用数据集
3	Q<70	参考数据集

8.2　数据结果质量报告

数据质量评估结果和评价报告是所有数据质量评价项目及其评测结果的合集。

在完整的数据质量评价结果和报告中，包括全部上述内容。此外，在数据质量评价报告中对此进行的评价过程的操作做出完整的记录，包括存在的质量级别的内容确定等。

附 录 A
（资料性）
淋巴瘤临床研究数据元

淋巴瘤临床研究数据集各个子域数据元见表 A.1～A.14。

表A.1　研究项目基本情况

内部编码	数据元名称	定义	数据类型	表示格式	允许值
CA.02.FA.00.0001	研究方案名称	研究方案中文名称	S1	AN.200	
CA.02.FA.00.0002	研究方案编号	研究方案文书的唯一识别号	S1	AN..50	
CA.02.FA.00.0003	研究方案版本日期	研究方案版本当日的公元纪年日期	D	D8	
CA.02.FA.00.0004	知情同意书版本日期	知情同意书版本当日的公元纪年日期	D	D8	
CA.02.FA.00.0005	申办单位名称	申办单位的名称	S1	AN..100	
CA.02.FA.00.0006	研究单位名称	研究单位的名称	S1	AN..100	
CA.02.FA.00.0007	研究单位编号	参与研究单位的唯一标识号	S1	AN..50	
CA.02.FA.00.0008	牵头研究者	在多中心临床研究中，负责整个临床研究的研究者，即牵头单位或组长单位的主要研究者姓名	S1	AN..10	
CA.02.FA.00.0008	研究单位主要研究者	研究单位负责临床研究的主要研究者姓名	S1	AN..10	
CA.02.FA.00.0009	临床研究状态	临床研究所处的阶段	S3	N1	见 表 B.1
CA.02.FA.00.0010	研究开始日期	临床研究启动会召开当日的公元纪年日期	D	D8	

CA.02.FA.00.0011	研究结束日期	对最后一例受试者完成全部研究相关数据采集当日的公元纪年日期	D	D8	

表A.1　入组及知情同意数据元专用属性

内部编码	数据元名称	定义	数据类型	表示格式	允许值
CA.02.RZ.00.0001	受试者标识	研究中分配给受试者以辨识其身份的唯一编号	S1	AN..20	
CA.02.RZ.00.0002	知情同意书签署日期	受试者签署知情同意书的日期	D	D8	
CA.02.RZ.00.0003	受试者筛选日期	受试者接受入组筛选的日期	D	D8	
CA.02.RZ.00.0004	受试者入组状态	受试者经筛选后是否成功入组	L	T/F	
CA.02.RZ.00.0005	受试者入组失败原因	受试者经筛选后无法入组的原因	S1	AN..50	
CA.02.RZ.00.0006	受试者是否符合所有入排标准	受试者是否符合所有入组标准且不符合任何一条排除标准	L	T/F	
CA.02.RZ.00.0007	受试者不符合入排标准的编号	如果受试者不符合的入排标准的具体编号	S1	AN..10	
CA.02.RZ.00.0008	受试者是否接受随机分组	受试者是否接受随机分组	S2	N1	1：是； 2：否； 9：不适用
CA.02.RZ.00.0009	受试者接受随机分组时间	受试者接受随机分组的时间	D	D8	

CA.02.RZ.00.0010	受试者随机号	受试者随机号	S1	AN..20	
CA.02.RZ.00.0011	受试者所分配的研究组	受试者所分配进入的研究组名称	S1	AN..20	

表A.2　人口学资料数据元专用属性

内部编码	数据元名称	定义	数据类型	表示格式	允许值
CA.02.RK.01.0001	出生日期	受试者出生当日的公元纪年日期	D	D8	
CA.02.RK.01.0002	年龄	受试者从出生当日公元纪年日起到签署知情同意书当日止生存的时间长度。年龄满 1 周岁者，以实足年龄的相应整数为准；不足 1 岁的，以实足天数为准	N	N3	备注：0-365
CA.02.RK.01.0003	年龄计量单位	年龄的计量单位。年龄满 1 周岁者，选择年为计量单位；不满 1 岁者，以天为计量单位	S2	N1	0：年 1：天
CA.02.RK.02.0001	性别	受试者生理性别在特定编码体系中的代码	S3	N1	GB/T2261.1-2003
CA.02.RK.03.0002	国籍	受试者所属国籍在特定编码体系中的代码	S3	AN3	GB/T 2659-2000
CA.02.RK.03.0003	民族	受试者所属民族类别在特定编码体系中的代码	S3	N2	GB/T3304-1991
CA.02.RK.03.0004	其他民族或种族	受试者所属民族或人种类别未在特定编码体系范围内时，记录的民族或种族信息	S1	AN..20	

CA.02.RK.04.0001	婚姻状况	受试者当前婚姻状况的代码	S3	N2	GB/T2261.2-2003
CA.02.RK.04.0002	孕次	受试者曾经妊娠次数的累计值	S2	N1	
CA.02.RK.04.0003	产次	受试者曾经生产次数的累计值	S2	N1	

表A.3　个人家族及既往疾病情况数据元专用属性

内部编码	数据元名称	定义	数据类型	表示格式	允许值
CA.02.JW.01.0001	吸烟状况代码	受试者过去和现在的吸烟情况代码	S2	N1	WS 365-2011
CA.02.JW.01.0002	开始吸烟年龄（岁）	受试者开始吸第一支烟的周岁年龄，计量单位为岁	N	N..2	
CA.02.JW.01.0003	日吸烟量（支）	平均每天的吸烟量，计量单位为支	N	N..3	
CA.02.JW.01.0004	戒烟年龄（岁）	受试者层吸烟但已成功戒烟时的周岁年龄，计量单位为岁	N	N..2	
CA.02.JW.01.0005	饮酒频率代码	受试者饮酒的频率代码	S3	N1	Ws364.5 卫生信息数据元值域代码第5部分：健康危险因素 CV03.00.104 饮酒频率代码表
CA.02.JW.01.0006	开始饮酒年龄（岁）	受试者第一次饮酒时的周岁年龄，计量单位为岁	N	N..2	

CA.02.JW.01.0007	日饮酒量（两）	受试者平均每天的饮酒量相当于白酒量，计量单位为两	N	N..3	
CA.02.JW.01.0008	戒酒标志	标识受试者曾饮酒者是否成功戒酒	L	T/F	
CA.02.JW.01.0009	戒酒年龄（岁）	受试者曾曾饮酒但现已戒酒者成功戒酒时的周岁年龄 ，计量单位为岁	N	N..2	
CA.02.JW.01.0010	饮酒种类代码	受试者饮酒种类代码	S3	N1	WS364.5 卫生信息数据元值域代码第 5 部分：健康危险因素 CV03,011.105 饮酒种类代码表
CA.02.JW.02.0001	肿瘤家族史标志	标识受试者家族成员中是否有肿瘤患者	L	T/F	
CA.02.JW.02.0002	患肿瘤亲属与受试者关系代码	患肿瘤家族成员于受试者的家庭和社会关系所属类别在特定编码体系中的代码	S3	N2	GB/T 4761
CA.02.JW.02.0003	肿瘤家族史瘤别	家族成员所患肿瘤的国际疾病分类代码	S3	AN..5	ICD-10
CA.02.JW.03.0001	药物过敏史标志	标识受试者既往有无药物过敏经历	L	T/F	
CA.02.JW.03.0002	过敏史	受试者既往发生过敏情况的详细描述	S1	AN..100	

CA.02.JW.03.0003	既往或伴随疾病标志	受试者是否存在既往或伴随疾病	L	T/F	
CA.02.JW.03.0004	既往或伴随疾病名称	受试者存在的既往或伴随疾病名称	S3	AN..5	ICD-10
CA.02.JW.03.0005	既往或伴随疾病编号	受试者存在的既往或伴随疾病顺序号	N	N..2	
CA.02.JW.03.0006	既往或伴随疾病开始时间	受试者存在的既往或伴随疾病发病时间	D	D8	
CA.02.JW.03.0007	既往或伴随疾病目前状态	受试者既往或伴随疾病当前的情况	S2	N1	0：已治愈； 1：仍存在；
CA.02.JW.03.0008	既往或伴随疾病治愈时间	受试者既往或当前伴随疾病的治愈时间	D	D8	
CA.02.JW.03.0008	疫苗接种情况	受试者既往或当前伴随疾病的治愈时间	S1	AN..100	

表A.4　一般状况及体格检查数据元专用属性

内部编码	数据元名称	定义	数据类型	表示格式	允许值
CA.02.TC.00.0001	体格检查状态	受试者是否进行了体格检查	L	T/F	
CA.02.TC.00.0002	体格检查日期	受试者接受体格检查的日期	D	D8	
CA.02.TC.00.0003	体格检查时间	受试者接受体格检查的时间	T	T6	
CA.02.TC.01.0001	ECOG 评分	肿瘤患者体力状况 ECOG 评分	S3	N1	见表B.2

CA.02.TC.01.0002	B 症状	患者是否有 B 症状	L	T/F	
CA.02.TC.02.0001	体重(kg)	个体体重的测量值，计量单位为 kg	N	N3..5,1	
CA.02.TC.02.0002	身高(cm)	个体身高的测量值，计量单位为 cm	N	N4..5,1	
CA.02.TC.02.0003	收缩压(mmHg)	个体收缩压的测量值，计量单位为 mmHg	N	N2..3	
CA.02.TC.02.0004	舒张压(mmHg)	个体舒张压的测量值，计量单位为 mmHg	N	N2..3	
CA.02.TC.02.0005	呼吸频率(次/min)	个体每分钟呼吸次数的测量值，计量单位为次/min	N	N..3	
CA.02.TC.02.0006	脉率(次/min)	个体每分钟脉搏的次数测量值,计量单位为次/min	N	N2..3	
CA.02.TC.02.0007	体温(℃)	个体体温的测量值,计量单位为℃	N	N4,1	
CA.02.TC.03.0001	体格检查部位	受试者体格检查具体部位	S3	N1	见表B.3
CA.02.TC.03.0002	体格检查结果	受试者体格检查结果	S2	N1	0：正常; 1：异常; 9：未查;
CA.02.TC.03.0003	淋巴结侵犯区域	对发生淋巴瘤肿大的位置描述	S1	A..30	见表B.4，B.5,B.6
CA.02.TC.03.0004	淋巴结外侵犯区域	对发生淋巴瘤肿大的位置描述（淋巴结外）	N	N..3,1	见表B.7

表A.5　淋巴瘤诊断数据元专用属性

内部编码	数据元名称	定义	数据类型	表示格式	允许值
CA.02.ZD.01.0001	恶性肿瘤诊断	本次临床研究针对的受试者恶性肿瘤诊断	S3	AN..5	ICD10
CA.02.ZD.01.0002	恶性肿瘤诊断具体描述	本次临床研究所诊治的受试者恶性肿瘤诊断，在ICD10中未找到具体描述者	S1	AN..20	
CA.02.ZD.01.0003	恶性肿瘤侧位	对于成对器官，恶性肿瘤原发于左侧或右侧的位置代码	S3	N1	
CA.02.ZD.01.0004	恶性肿瘤诊断依据代码	肿瘤诊断依据在特定编码体系中的代码	S3	N2	WS 364.10 卫生信息数据源值域代码 第10部分：医学诊断 CV05.01.027 肿瘤诊断依据代码表
CA.02.ZD.01.0005	恶性肿瘤首次诊断日期	受试者首次诊断为临床研究所针对的恶性肿瘤的日期	D	D8	
CA.02.ZD.02.0001	恶性肿瘤病理检测	恶性肿瘤是否行病理活检	L	T/F	
CA.02.ZD.02.0002	病理活检部位	恶性肿瘤病理活检部位	S3	N1	见表B.8
CA.02.ZD.02.0003	具体病理活检部位	恶性肿瘤病理活检具体部位，详细到器官	S1	N..2	见表B.4，B.5,B.6,B.7

CA.02.ZD.02.0004	免疫组化结果	恶性肿瘤免疫组化检测结果，仅在临床研究需要时记录特定检测结果	S1	AN..100	
CA.02.ZD.02.0005	基因检测状态	是否行淋巴瘤基因检测	L	T/F	
CA.02.ZD.02.0006	基因检测取样部位	受试者行淋巴瘤基因检测的取样部位	S1	AN..10	
CA.02.ZD.02.0007	基因检测时间	基因检测的采样时间	D	D8	
CA.02.ZD.02.0008	基因检测方法	基因检测技术方法	S1	AN..10	
CA.02.ZD.02.0009	基因检测结果描述	恶性肿瘤基因检测结果，仅在临床研究需要时记录特定检测结果	S1	AN..100	
CA.02.ZD.02.0010	恶性肿瘤病理分级	恶性肿瘤组织学分级	S3	N1	
CA.02.ZD.02.0011	其他恶性肿瘤病理分级	恶性肿瘤病理分级选择 9 时，具体描述肿瘤病理分级情况	S1	AN..10	
CA.02.ZD.03.0001	淋巴瘤 2014 版 Lugano 分期	Lugano2014 标准中的 Lugano 分期	S3	N1	见表 B.9
CA.02.ZD.03.0002	Rai 分期	是慢淋临床分期标准中的一种	S3	N1	见表 B.10
CA.02.ZD.03.0003	Binet 分期	目前慢淋采用较多的临床分期法	S2	N1	见表 B.11
CA.02.ZD.03.0004	EORTC-TNMB 分期	常用于判断皮肤蕈样霉菌病/塞扎里（Sezary）综合征临床分期	S3	N1	见表 B.12
CA.02.ZD.04.0001		受试者当前是否存在肿瘤原发灶或局部复发病灶	L	T/F	

	恶性肿瘤当前是否存在原发病变或局部复发				
CA.02.ZD.04.0002	恶性肿瘤最近一次进展或复发日期	受试者参加临床试验前，最近一次肿瘤进展或复发日期。对于初治肿瘤，该日期等同于肿瘤首次诊断日期	D	D8	
CA.02.ZD.04.0003	IPI 评分	指根据患者的年龄、一般状况评分、临床分期，以及淋巴结外受侵的部位数目，以及乳酸脱氢酶是否正常来进行的一种评分方式	N	N1	见表 B.13
CA.02.ZD.04.0004	IPS 评分	指晚期霍奇金淋巴瘤评分	S2	A..10	见表 B.14
CA.02.ZD.04.0005	CLL-IPI 评分	指慢性淋巴细胞白血病国际预后指数评分总分	N	N1	见表 B.15
CA.02.ZD.04.0006	FLIPI 评分	即滤泡性淋巴瘤国际预后指数，适用于评估侵袭性淋巴瘤的预后	N	N1	见表 B.16
CA.02.ZD.04.0007	FLIPI-2 评分	指新的滤泡性淋巴瘤国际预后指数	N	N1	见表 B.17
CA.02.ZD.04.0008	MIPI 评分	指滤泡淋巴瘤国际预后指数	N	N1	见表 B.18
CA.02.ZD.04.0009	NRI 评分	指 ENKTCL 淋巴瘤改进版风险指数	N	N1	见表 B.19
CA.02.ZD.04.0010	NCCN-IPI 评分	指 NCCN 指南推荐指数评分	N	N1	见表 B.20

CA.02.ZD.04.001 1	CNS-IPI 评分	指继发中枢侵犯风险评分	N	N1	见表 B.21

表A.6淋巴瘤既往治疗数据元专用属性

内部编码	数据元名称	定义	数据类型	表示格式	允许值
CA.02.ZL.01.0001	既往肿瘤手术治疗状态	受试者入组前是否接受恶性肿瘤手术治疗	L	T/F	
CA.02.ZL.01.0002	手术名称	既往因恶性肿瘤接受手术的名称	S1	AN..80	
CA.02.ZL.01.0003	手术/操作日期	既往因恶性肿瘤接受手术的日期	D	D8	
CA.02.ZL.01.0004	手术/操作目的	既往因恶性肿瘤接受手术的目的	S2	N1	1：根治； 2：姑息； 3：探查；
CA.02.ZL.01.0005	手术/操作部位	既往因恶性肿瘤接受手术的具体部位	S3	N..2	
CA.02.ZL.01.0006	手术其他部位的具体描述	手术/操作部位选择“其他”时，具体说明手术/操作部位	S1	AN..20	
CA.02.ZL.02.0001	既往肿瘤放射治疗状态	受试者入组前是否接受恶性肿瘤放射治疗治疗	L	T/F	
CA.02.ZL.02.0002	放射治疗目的	既往因恶性肿瘤接受放射治疗的目的	S2	N1	见表 B.22
CA.02.ZL.02.0003	放射治疗其他目的具体描述	放射治疗目的选择“其他”时，具体说明放射治疗目的	S1	AN..50	
CA.00.ZL.02.0004	放射治疗射线	放射治疗所用的射线性质	S3	N1	见表 B.23
CA.00.ZL.02.0005	放射治疗射线能量	放射治疗所用的射线能量	N	N..2	
CA.00.ZL.02.0006	放射治疗射线能量单位	放疗所用的射线能量单位	S1	AN..5	
CA.00.ZL.02.0007	放射治疗技术	放疗所用的治疗技术	S3	N1	见表 B.24

CA.00.ZL.02.0008	放射治疗开始日期	既往因恶性肿瘤接受放射治疗的首次治疗日期	D	D8	
CA.00.ZL.02.0009	放射治疗结束日期	既往因恶性肿瘤接受放射治疗的末次治疗日期	D	D9	
CA.00.ZL.02.0010	放射治疗部位	既往因恶性肿瘤接受放射治疗的具体部位	S3	N..2	
CA.00.ZL.02.0011	放射治疗其他部位的具体描述	放射治疗部位选择“其他”时，具体说明放射治疗部位	S1	AN..20	
CA.00.ZL.02.0012	放射治疗单次剂量	单次放射治疗剂量	N	N..4,1	
CA.00.ZL.02.0013	放射治疗次数	放射治疗分次数	N	N..2	
CA.00.ZL.02.0014	放射治疗疗效	放射治疗疗效评价	S3	N1	见表 B.25
CA.02.ZL.03.0001	既往肿瘤系统治疗状态	受试者入组前是否接受恶性肿瘤全身系统治疗	L	T/F	
CA.02.ZL.03.0002	系统治疗目的	既往因恶性肿瘤接受特定系统治疗的目的	S2	N1	1：新辅助治疗； 2：辅助治疗; 3：晚期
CA.02.ZL.03.0003	系统治疗线数	因淋巴瘤接受特定系统治疗的线数	N	N..2	系统治疗线数（CA.02.ZL.03.0003-0006）按照具体治疗情况增加
CA.02.ZL.03.0004	系统治疗方案	特定系统治疗方案简述	S1	AN..100	
CA.02.ZL.03.0005	系统治疗周期数	特定系统治疗使用周期数	N	N..2	

CA.02.ZL.03.0006	系统治疗最佳疗效	特定系统治疗的最佳疗效	S3	N1	见表 B.26
CA.02.ZL.03.0007	中止系统治疗原因	中止特定系统治疗的原因	S3	N1	见表 B.27
CA.02.ZL.03.0008	中止系统治疗的其他具体原因	中止系统治疗的原因选其他时，具体说明治疗中止原因	S1	AN..50	
CA.02.ZL.03.0009	系统治疗后疾病进展时间	特定系统中治疗过程中或治疗后疾病进展时间	D	D8	
CA.02.ZL.03.0010	系统治疗药物名称	系统治疗特定方案的每个具体药物名称	S1	AN..50	
CA.02.ZL.03.0011	系统治疗药物类型	系统治疗特定方案的每个具体药物的类型	S3	N1	见表 B.28
CA.02.ZL.03.0012	系统治疗药物其他类型	系统治疗药物类型选择其他时，具体说明药物类型	S1	AN..20	
CA.02.ZL.03.0013	系统治疗药物剂量	系统治疗特定方案的每个具体药物的剂量	N	N..7,2	
CA.02.ZL.03.0014	药物剂量单位	系统治疗特定方案的每个具体药物的剂量单位	S3	N2	见表 B.28
CA.02.ZL.03.0015	其他药物剂量单位	药物剂量单位选择其他时，具体说明药物剂量单位	S1	AN..10	
CA.02.ZL.03.0016	给药方式	系统治疗特定方案的每个具体药物的给药方式在特定编码体系中的代码	S3	AN..3	WS364.10 卫生信息数据元值域代码 第 12 部分：计划与干预

					CV06.00.102 用药途径代码表
CA.02.ZL.03.0017	给药频率	系统治疗特定方案的每个具体药物的给药频率	S1	AN..20	
CA.02.ZL.03.0018	首次给药时间	系统治疗特定方案的每个具体药物的首次给药时间	D	D8	
CA.02.ZL.03.0019	末次给药时间	系统治疗特定方案的每个具体药物的末次给药时间	D	D8	
CA.02.ZL.04.0001	既往肿瘤其他治疗状态	受试者入组前是否接受恶性肿瘤除手术、放疗及药物治疗以外的治疗	L	T/F	
CA.02.ZL.01.0002	其他治疗名称	既往因恶性肿瘤接受其他治疗的名称	S1	AN..80	
CA.02.ZL.01.0003	其他治疗日期	既往因恶性肿瘤接受其他治疗的日期	D	D8	
CA.02.ZL.01.0004	其他治疗具体情况	既往因恶性肿瘤接受其他治疗的具体情况	S1	AN..200	

表A.7　干预措施数据元专用属性

内部编码	数据元名称	定义	数据类型	表示格式	允许值
CA.02.SY.01.0001	手术干预措施状态	受试者试验干预措施中是否包含手术治疗	L	T/F	
CA.02.SY.01.0002	手术名称	手术干预措施名称	S1	AN..80	

CA.02.SY.01.0003	手术日期	手术干预措施日期	D	D8	

CA.02.SY.02.0001	放疗干预措施状态	受试者实验干预措施中是否包含放射治疗	L	T/F	
CA.00.SY.02.0002	试验性放疗射线性质	放射治疗所用的射线性质	S3	N1	见表 B.23
CA.00.SY.02.0003	放疗射线能量	放射治疗所用的射线能量	N	N..2	
CA.00.SY.02.0004	放疗射线能量单位	放疗所用的射线能量单位	S1	AN..5	
CA.00.SY.02.0005	试验性放疗技术	放疗所用的治疗技术	S3	N1	见表 B.25
CA.00.SY.02.0006	试验性放疗开始日期	放疗干预措施开始日期	D	D8	
CA.00.SY.02.0007	试验性放疗结束日期	放疗干预措施结束日期	D	D9	
CA.00.SY.02.0008	试验性放疗部位	放疗干预措施照射部位	S1	AN..20	
CA.00.SY.02.0009	放射治疗单次剂量	单次放射治疗剂量	N	N..4,1	
CA.00.SY.02.0010	放射治疗次数	放射治疗分次数	N	N..2	
CA.02.SY.03.0001	试验性药物治疗状态	受试者试验干预措施中是否包含药物治疗	L	T/F	
CA.02.SY.03.0002	试验性治疗药物名称	试验药物名称	S1	AN..50	
CA.02.SY.03.0003	试验药物实际剂量	试验药物实际用量	N	N..7,2	
CA.02.SY.03.0004	试验药物剂量单位	试验治疗药物的剂量单位	S3	N2	见表 B.28

CA.02.SY.03.0005	其他试验药物剂量单位	药物剂量单位选择其他时，具体说明药物剂量单位	S1	AN..10	
CA.02.SY.03.0006	试验药物给药方式	试验药物的给药方式在特定编码体系中的代码	S3	AN..3	WS364.10 卫生信息数据元值域代码 第12部分：计划与干预 CV06.00.102 用药途径代码表
CA.02.SY.03.0007	试验药物给药频率	试验药物的给药频率	S1	AN..20	
CA.02.SY.03.0008	首次给药日期	试验药物的首次给药日期	D	D8	
CA.02.SY.03.0009	首次给药时间	试验药物首次给药时间	T	T6	
CA.02.SY.03.0010	末次给药日期	试验药物的末次给药日期	D	D8	
CA.02.SY.03.0011	末次给药时间	试验药物末次给药时间	T	T6	
CA.02.SY.03.0012	造血干细胞移植日期	进行造血干细胞移植的公元纪年日期描述	D	D10	
CA.02.SY.03.0013	造血干细胞来源	造血干细胞的来源	S2	A..10	
CA.02.SY.03.0014	造血干细胞来源器官	造血干细胞来源的具体器官名称	S2	A..10	
CA.02.SY.03.0015	造血干细胞术前动员策略	术前动员策略的详细描述	S1	A..30	

CA.02.SY.03.0016	造血干细胞术前动员药物	术前动员使用药物的通用名称	S1	A..30	
CA.02.SY.03.0017	造血干细胞术前动员药物剂量	术前动员药物每日使用的剂量	N	N..3,1	
CA.02.SY.03.0018	造血干细胞术前动员开始时间	术前动员药物开始使用的公元纪年日期描述	D	D10	
CA.02.SY.03.0019	造血干细胞术前动员结束时间	术前动员药物结束使用的公元纪年日期描述	D	D10	
CA.02.SY.03.0020	造血干细胞采集目标剂量	预计采集造血干细胞的剂量描述	N	N..3,1	
CA.02.SY.03.0021	造血干细胞采集目标日期	预计采集造血干细胞的日期	D	D10	
CA.02.SY.03.0022	造血干细胞采集天数	指本次采集造血干细胞操作的总天数	N	N..3,1	
CA.02.SY.03.0023	造血干细胞分得单个核细胞数	本次采集造血干细胞单个核细胞数量	N	N..3,1	
CA.02.SY.03.0024	造血干细胞分得CD34细胞数	本次采集造血干细胞CD34细胞数量	N	N..3,1	
CA.02.SY.03.0025	造血干细胞移植前评效	移植前疗效评估	N	N..3,1	
CA.02.SY.03.0026	造血干细胞预处理日期	进行预处理的公元纪年日期描述	D	D10	
CA.02.SY.03.0027	造血干细胞预处理方案选择	预处理方案的详细描述	S1	A..30	
CA.02.SY.03.0028	造血干细胞预处理方案药物	预处理方案使用的药物通用名称	S1	A..30	
CA.02.SY.03.0029	造血干细胞预处理方案药物剂量	预处理方案药物每次使用的剂量	N	N..3,1	

CA.02.SY.03.0030	造血干细胞回输干细胞日期	粒细胞植入的公元纪年日期和时间的描述	D	D10	
CA.02.SY.03.0031	造血干细胞回输CD34细胞数目	血小板植入的公元纪年日期和时间的描述	D	D10	
CA.02.SY.03.0032	造血干细胞回输单个核细胞数目	移植后治疗情况的详细描述	S1	A..30	
CA.02.SY.03.0033	开始时间	患者开始进行CAR-T治疗时的公元纪年日期和时间的完整描述	D	D10	
CA.02.SY.03.0034	结束时间	患者结束CAR-T治疗时的公元纪年日期和时间的完整描述	D	D10	
CA.02.SY.03.0035	CAR类型	指使用的抗原类型	S2	A..10	
CA.02.SY.03.0036	预处理方案	患者CART治疗的预处理方案的描述，如：Flu/Cy	S2	A..10	
CA.02.SY.03.0037	预处理药物	患者CART治疗的预处理药物的药品名称	S1	A..30	
CA.02.SY.03.0038	剂量	患者CART治疗使用药物的剂量	N	N..3,1	
CA.02.SY.03.0039	CD4：CD8相对比例	CD4：CD8相对比例的实际检测值	N	N..3,2	
CA.02.SY.03.0040	细胞用量	CART治疗的细胞用量统计	N	N..3	
CA.02.SY.03.0041	采集时间	采集CART样本的公元纪年日期和时间的完整描述	D	D10	
CA.02.SY.03.0042	回输时间	回输CART样本细胞的公元纪年日期和时间的完整描述	D	D10	

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：

https://d.book118.com/977120123162006100