

队列研究基本信息数据元规范

Data element specifications of fundamental information
for cohort study

(点击此处添加与国际标准一致性程度的标识)

(征求意见稿)

XXXX – XX – XX 发布

XXXX – XX – XX 实施

目 次

前言 II

1 范围3

2 规范性引用文件 3

3 术语和定义 3

4 数据元属性 4

5 数据元5

6 数据元值域代码表 14

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。

本文件由北京市卫生健康委员会提出并归口。

本文件由北京市卫生健康委组织实施。

本文件起草单位：

本文件主要起草人：

队列研究基本信息数据元规范

1 范围

本文件规定的队列研究所需基本信息数据元的属性描述及代码制定。

本文件适用于医疗卫生机构开展队列研究的登记、处理、交换、共享和应用。高等学校、科研院所开展的队列研究的信息化管理可参照执行。

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。

GB/T 2659.1 世界各国和地区及其行政区划名称代码 第1部分：国家和地区代码

GB/T 7408 数据元和交换格式 信息交换 日期和时间表示法

GB/T 8561 专业技术职务代码

GB/T 13000 信息技术 通用多八位编码字符集（UCS）

GB/T 14396 疾病分类与代码

GB/T 23705 数字城市地理信息公共平台地名/地址编码规则

WS/T 303 卫生健康信息数据元标准化规则

WS 363.1 卫生健康信息数据元目录 第1部分：总则

WS 364.5 卫生健康信息数据元值域代码 第5部分：健康危险因素

WS 364.10 卫生健康信息数据元值域代码 第10部分：医学诊断

3 术语和定义

下列术语和定义适用于本文件。

3.1

队列研究 cohort study

一种观察性研究设计，用于研究特定人群中暴露因素与结局之间的关系。在队列研究中，研究者选取一组有共同特征的人群（称为队列），并在一段时间内对他们进行随访，记录他们的健康结局。

3.2

队列研究基本信息 fundamental information for cohort study

队列研究数据登记、处理、交换、共享和应用所需的基本信息，包括研究项目基本情况、伦理及知情同意、研究方案设计、随访情况、暴露因素、结局变量、执行现状、研究成果和数据共享的资料。

4 数据元属性

4.1 数据元公用属性

数据元公用属性应按表1描述。

表 1 数据元公用属性

属性种类	数据元属性名称	属性值
标识类	相关环境	V1.0
	相关环境	卫生信息
关系类	分类模式	分类法

4.2 数据元专用属性描述规则

- 4.2.1 数据元包括数据元标识符、中文名称、英文名称、定义、字段类型、数据格式和说明。
- 4.2.2 数据元名称、定义、字段类型、数据格式和说明应符合 WS 363.1 的规定。
- 4.2.3 数据元标识符用字母和数字表示，编码结构见图 1。编码长度为 13 位，其中：

——域标识用2位大写英文字母表示，统一为“CO”；

——子域代码用2位字母表示,应符合表2的要求。域标识和子域代码之间加“.”；

——亚类别码用2位数字表示，数字大小无含义，取值范围为00～99。亚类别码从00开始编码，子域代码与亚类别码之间加“.”；

——顺序码用4位数字表示，按先后顺序赋码，取值范围为0001～9999。顺序码从0001开始编码，亚类别码与顺序码之间加“.”。

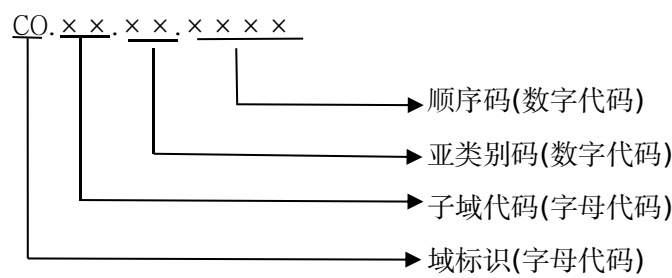


图1 队列研究基本信息数据元标识符结构图

表 2 队列研究基本信息数据元子域

序号	子域代码	名称	定义
1	XM	研究项目基本情况	研究项目名称、研究单位、负责人、临床平台注册等基本信息。
2	LL	伦理及知情同意	伦理机构、批件编码、知情同意书等相关信息。
3	FA	研究方案设计	研究项目样本量、研究对象、抽样方法等相关信息。
4	SF	随访信息	随访频率、随访内容等信息。
5	SJ	研究数据信息	研究收集变量的相关信息。

6	GG	研究方案更改	研究方案更改时间、原因、内容等相关变量。
7	ZX	研究执行现状	队列研究状态、执行异常原因、首例对象入组日期、已入组对象例数等相关信息。
8	CG	研究成果	项目产生的论文、专利、软件著作权产出等相关信息。
9	GX	数据共享	研究项目公开、共享方式等相关信息。
10	QT	队列其他信息	以上子域中未涵盖的队列研究项目相关信息。

4.2.4 数据元标识符的亚类别代码应符合以下要求：

- 研究项目基本情况亚类别码为“00”；
- 伦理及知情同意亚类别码为“00”；
- 研究方案设计亚类别码符合表 3 的要求；
- 随访信息亚类别码为“00”；
- 研究数据信息亚类别码为“00”；
- 研究执行现状亚类别码为“00”；
- 研究方案更改亚类别为“00”；
- 研究成果亚类别码符合表 4 的要求；
- 数据共享亚类别为“00”。

表 3 研究方案设计亚类别码

亚类别码编号	名称	定义
01	研究目的	与队列研究目的的数据元亚类
02	研究设计	与队列研究设计相关的数据元亚类
03	研究特点	队列研究的优势、劣势相关的数据元亚类
04	样本量信息	与队列包含的研究对象数目相关的数据元亚类
05	人口学资料	与研究对象人口学资料相关的数据元亚类
06	抽样方法	与抽样方法相关的数据元亚类
07	纳入排除标准	与研究对象纳入标准或排除标准相关的数据元亚类

表 4 研究成果亚类别码

亚类别码编号	名称	定义
01	论文	产生的论文相关的数据元亚类
02	专利	获批专利相关的数据元亚类
03	软件著作权	获批软件著作权的数据元亚类
04	其他成果	其他成果相关的数据元亚类

5 数据元

5.1 研究项目基本情况数据元专用属性见表 5。

表 5 研究项目基本情况数据元专用属性

内部编码	中文名称	英文名称	定义	字段类型	数据格式	说明
CO.XM.00.0001	中文名称	Chinese name	队列研究方案的	S1	AN..200	

			中文名称			
CO.XM.00.0002	英文名称	English name	队列研究方案的英文名称	S1	AN..200	
CO.XM.00.0003	项目负责人姓名	Investigator' s name	队列研究项目的主要责任人在公安户籍管理部门正式登记注册的姓氏和名称	S1	AN..50	GB/T 13000
CO.XM.00.0004	项目负责人专业技术职务	Investigator' s title of a technical post	项目负责人所在单位的专业技术职务。	S3	N..3	GB/T 8561
CO.XM.00.0005	项目负责人行政职务	Investigator' s title of a position	项目负责人在所在单位的行政职务。	S1	AN..20	
CO.XM.00.0006	项目负责人所属单位	Workplace of investigator	负责人所在单位名称的全称	S1	AN..100	
CO.XM.00.0007	项目牵头单位	Leading institution	项目牵头单位名称全称	S1	AN..100	
CO.XM.00.0008	项目资助来源名称	Funding	主要资助项目名称	S1	AN..100	
CO.XM.00.0009	项目资助来源编号	Funding' s number	主要资助项目对应的编号	S1	AN..50	
CO.XM.00.0010	项目联系人姓名	Contact name	研究单位负责沟通和协调的人员在公安户籍管理部门正式登记注册的姓氏和名称	S1	AN..50	
CO.XM.00.0011	项目联系人专业技术职务	Contact title of a technical post	项目联系人所在单位的专业技术职务	S3	N..3	GB/T 8561
CO.XM.00.0012	项目联系人行政职务	Contact title of a position	项目联系人在所述单位的行政职务。	S1	AN..20	
CO.XM.00.0013	项目联系人所属单位	Workplace of contact	项目联系人所在单位名称的全称	S1	AN..100	
CO.XM.00.0014	项目联系人所属单位地址	Address of contact' s workplace	项目联系人所属单位地址	S1	AN..100	GB/T 23705
CO.XM.00.0015	项目联系人联系方式	Information of the contact person	项目联系人电话号码, 可以是手机	S1	AN..100	

			或座机号码			
CO.XM.00.0016	研究开始日期	Start time	研究启动会召开当日的公元纪年日期	D	D8	
CO.XM.00.0017	研究结束日期	End time	对最后一例研究参与者完成全部研究相关数据采集当日的公元纪年日期	D	D8	
CO.XM.00.0018	临床研究注册	Registration	是否已完成临床研究注册	L	T/F	
CO.XM.00.0018	临床研究注册平台	Registration platform	临床研究注册平台名称	S1	AN..100	
CO.XM.00.0019	临床研究注册平台注册号	Registration identify number	临床研究注册平台的注册号	S1	AN..100	
CO.XM.00.0020	卫生健康委员会备案	Municipal Health Commission record	是否在卫生健康委员会备案	L	T/F	
CO.XM.00.0021	卫生健康委员会备案号	Municipal Health Commission record number	卫生健康委员会备案号	S1	AN..100	
CO.XM.00.0022	人类遗传资源	Human genetic resources	是否进行人类遗传资源备案	S2	N1	1: 是 2: 否 3: 不涉及
CO.XM.00.0023	人类遗传资源备案号	Human genetic resources record number	人类遗传资源备案号	S1	AN..100	

5.2 伦理及知情同意数据元专用属性描述见表 6。

表 6 伦理及知情同意数据元专用属性

内部编码	中文名称	英文名称	定义	字段类型	数据格式	说明
CO.LL.00.0001	伦理申请状态	Ethical application status	伦理申请所处阶段	S3	N1	详见表 15
CO.LL.00.0002	伦理委员会	Ethics Committee	伦理委员会名称	S1	AN..100	

CO.LL.00.0003	批件编码	Approval number	队列项目的伦理批件编码	S1	AN..100	
CO.LL.00.0004	伦理批准日期	Approval date	伦理获批的日期	D	D8	
CO.LL.00.0005	伦理批件	Ethical approval upload	伦理批件上传	B	B..2	
CO.LL.00.0006	知情同意书签署	Informed consent	队列研究受试者是否签署同意参加研究的知情同意书	S2	N1	1: 是 2: 否 9: 不适用
CO.LL.00.0007	知情同意书模板	Informed consent upload	如“知情同意书签署”选择“是”，在此上传知情同意书模板，文件格式为PDF或word版本	B	B..3	

5.3 研究方案设计数据元专用属性描述见表 7。

表 7 研究方案设计数据元专用属性

内部编码	中文名称	英文名称	定义	字段类型	数据格式	说明
CO.FA.01.0001	研究目的	Objectives	描述主要研究目的，例如流行病学调查/暴露与疾病关联/观察干预效果/其他等	S1	AN..200	
CO.FA.02.0001	研究对象来源	Study design	队列研究受试者来源	S2	N1	1: 社区 2: 专病 9: 其他
CO.FA.02.0002	其他研究对象来源	Other study design	如“队列研究受试者来源”选择“其他”，在此描述	S1	AN..100	
CO.FA.03.0001	优势	Strength	简要描述登记的队列研究的主要特点和优势	S1	AN..200	
CO.FA.03.0002	劣势	Limitation	简要描述登记的主要劣势	S1	AN..200	
CO.FA.04.0001	总样本量	Sample size	队列研究包含的研究对象总数量	N	N..10	

CO.FA.04.0002	计划首要暴露	Primary exposure of the plan	研究计划中首要暴露变量	S1	AN..200	
CO.FA.04.0003	计划首要暴露的样本量	Sample Size in primary exposure group	首要暴露组研究包含的研究对象数量	S1	AN..200	
CO.FA.04.0004	计划首要结局事件	primary event of plan	研究计划中首要结局事件	S1	AN..200	
CO.FA.04.0005	计划首要结局事件样本量	Sample Size of primary event	研究计划中首要结局事件的样本量	S1	AN..200	
CO.FA.05.0001	起始年龄	Minimum age value	研究对象年龄范围的最小值。研究参与者从出生当日公元纪年起到入组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	S1	N3	0 – 365
CO.FA.05.0002	起始年龄计量单位	Unit of minimum age	起始年龄的计量单位。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	1: 年 2: 天
CO.FA.05.0003	最大年龄	Maximum age value	研究对象年龄范围的最大值。研究参与者从出生当日公元纪年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	S1	N3	0 – 365
CO.FA.05.0004	最大年龄计量单位	Unit of age	最大年龄的计量单位。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	1: 年 2: 天
CO.FA.05.0005	国家	Country	队列研究开展的国家	S3	AN3	GB/T 2659.1

		Distribution				
CO.FA.05.0006	地区	Geographic Distribution	队列研究开展的行政区划	S1	AN..100	
CO.FA.06.0001	抽样方法	Sampling methods	队列研究的抽样方法	S3	N1	详见表 16
CO.FA.06.0002	其他的抽样方法	Other sampling methods	如“抽样方法”选择“其他”，在此说明	S1	AN..20	
CO.FA.07.0001	纳入标准	Inclusion criteria	队列研究的纳入标准的内容及参考	S1	AN..200	
CO.FA.07.0002	排除标准	Exclusion criteria	研究的排除标准的内容及参考	S1	AN..200	

5.4 随访信息数据元专用属性描述见表 8。

表 8 随访信息数据元专用属性

内部编码	中文名称	英文名称	定义	字段类型	数据格式	说明
CO.SF.00.00001	是否进行随访	Follow-up	队列研究是否对研究参与者进行访视	L	T/F	
CO.SF.00.00002	随访频率	Frequency	研究周期内参与者入组后随访的时间间隔或时间节点	S1	AN..20	
CO.SF.00.00003	随访内容	Follow-up Content	随访过程中收集的资料, 包括各随访时间点调查的变量名称等	S1	AN..2000	

5.5 研究数据信息数据元专用属性描述见表 9。

表 9 研究数据信息数据元专用属性

内部编码	中文名称	英文名称	定义	字段类型	数据格式	说明
CO.SJ.00.00001	主要暴露因素	Exposure	队列研究中收集的主要暴露因素的变量名称	S1	AN..2000	
CO.SJ.00.00002	主要结局变量	Primary outcome	队列研究中收集的主要结局的变量名称	S1	AN..2000	
CO.SJ.00.00003	次要结局变量	Secondary outcome	队列研究中收集的研究对象的次要结局的变量名称	S1	AN..2000	
CO.SJ.00.00004	编码规则	Code	队列研究中收集的变量的标识符	S3	N1	1: 研究元数据 (SDTM)

						2: 临床研究数据 (CDASH) 3: 其他 4: 不适用
CO.SJ.00.0005	其他编码规则	Other code	如“编码规则”选“其他”，在此说明	S1	AN..200	
CO.SJ.00.0006	上传核心变量清单	Core variable list	上传队列研究中核心变量清单，清单文件的格式为 PDF 或 word 版。变量清单内容包括队列研究中收集的变量详细信息，如变量中文名称、变量代码、变量类型和测量方法等	B	B..10	

5.6 研究方案更改数据元专用属性描述见表 10。

表 10 研究方案更改数据元专用属性

内部编码	中文名称	英文名称	定义	字段类型	数据格式	说明
CO.GG.00.0001	研究方案调整	Adjustment	研究方案是否发生调整	L	T/F	
CO.GG.00.0002	研究方案调整日期	Adjustment date	研究方案调整当日的公元纪年日期	D	D8	
CO.GG.00.0003	研究方案调整原因	Adjustment Reason	研究方案调整原因描述	S1	AN..200	
CO.GG.00.0004	研究方案调整内容	Adjustment content	描述研究方案更改内容，包括更改前内容和更改后内容	S1	AN..2000	
CO.GG.00.0005	伦理审批	Ethical approval	更改后研究方案是否上报伦理委员会	S2	N1	1: 是 2: 否 9: 不详
CO.GG.00.0006	更改后备案	Record	更改后的研究方案是否向所属医疗机构备案	S2	N1	1: 是 2: 否 9: 不详

5.7 研究执行现状数据元专用属性见表 11。

表 11 研究执行现状数据元专用属性

内部编码	中文名称	英文名称	定义	字段类型	数据格式	说明
CO.ZX.00.0001	队列研究状态	Study status	目前队列研究所处的阶段	S3	N1	详见表 17
CO.ZX.00.0002	队列研究执行异常原因	Reason	如“队列研究状态”选择“暂停”、“提前终止”，在此说明	S1	AN..200	
CO.ZX.00.0003	首例研究对象入组日期	Enrollment time of first	首例研究对象接收研究相关数据采集的公元纪年日期	D	D8	
CO.ZX.00.0004	已入组研究对象例数	Number of enrolled subjects	截止至目前，的已入组研究研究对象例数	S1	N..200	
CO.ZX.00.0005	随访研究对象例数	Cases of subjects with at least one follow-up visit	已入组研究对象中，已完成一次或多次随访、规范收集信息的例数。以人为单位，单个人的多次随访计 1 例	S1	N..200	
CO.ZX.00.0006	已发生终点事件研究对象例数	Cases of endpoint events	以任务书规定的首要终点事件为准，已入组研究对象中，已发生一次或多次终点事件的例数。以人为单位，单个人多次发生终点事件计 1 例	S1	N..200	
CO.ZX.00.0007	已失访研究对象例数	Lost cases	未发生任务书规定的任何终点事件，同时目前已无法联系上该对象，或该对象明确拒绝继续参与研究、提供信息	S1	N..200	

5.8 研究成果数据元数据元专用属性见表 12。

表 12 研究成果数据元表述方式

内部编码	中文名称	英文名称	定义	数据类型	表示格式	说明
CO.CG.01.0001	有无论文发表	Article	是否已发表该队列的相关论文	L	T/F	
CO.CG.01.0002	论文发表数目	Number of articles	如“有论文发表”，填写已公开发表的论文数量	S1	N..3	

CO.CG.01.0003	论文清单	Article List	如“有论文发表”，列举具有代表性的论文信息，包括作者、论文题目、期刊名称、卷、册、doi、PMID	S1	AN..2000	
CO.CG.02.0001	有无或专利获批	Patent	是否获批该队列相关的专利	L	T/F	
CO.CG.02.0002	专利数量	Number of patents	如“有获批专利”，填写获批专利数量	S1	N..3	
CO.CG.02.0003	专利清单	Patent list	如“有获批专利”，列举具有代表性的获批专利信息，包括专利类型、发明人、专利号、获批时间等	S1	AN..2000	
CO.CG.03.0001	是否有软件著作权获批	Software copyright	是否获批队列相关软件著作权	L	T/F	
CO.CG.03.0002	软件著作权数量	Number of copyrights	如“有软件著作权获批”，填写获批软件著作权数量	S1	N..3	
CO.CG.03.0003	软件著作权清单	Software copyright list	如“有软件著作权获批”，列举具有代表性的软件著作权信息，包括软件名称、著作权人，开发完成时间、首次发表时间、登记号等	S1	AN..2000	
CO.CG.04.0001	是否有其他公开成果	Other products	是否有论文、专利和软件著作权之外的该队列研究相关的公开发表成果	L	T/F	
CO.CG.04.0002	其他成果列表	Other products list	如“有其他公开成果”，列举成果，包括类型、数量、名称、成果获得时间等	S1	AN..2000	

5.9 数据共享数据元专用属性见表 13。

表 13 数据共享数据元专用属性

内部编码	中文名称	英文名称	定义	数据类型	表示格式	说明
CO.GX.00.0001	是否数据公开	Data availability	队列研究收集的数据是否开放共享	L	T/F	
CO.GX.00.0002	数据共享形式	Data sharing	如队列研究项目数据公开，	S3	N1	详见表 18

			选择数据共享的形式			
CO.GX.00.0003	其他数据共享形式	Other data sharing	如“数据共享形式”，选“其他”请说明	S1	AN..200	
CO.GX.00.0004	数据共享联系人姓名	Contact for data	如队列研究项目数据公开，填写数据共享联系人在公安户籍管理部门正式登记注册的姓氏和名称	S1	AN..50	GB/T 13000
CO.GX.00.0005	数据共享联系方式	Contact information	获取公开共享数据的联系方式	S1	AN..200	
CO.GX.00.0006	数据管理单位	Data management affiliation	数据管理单位	S1	AN..200	

6 数据元值域代码表

6.1 数据元值域代码表编码规则

6.1.1 数据元值域代码表包括代码栏(代码指编码值,简称“值”)、编码对象名称栏(简称“值含义”)、说明栏,并可根据实际需要适当增减栏目。

6.1.2 数据元值域代码、编码对象名称的描述规则应符合 WS/T 363.1 的规定。

6.1.3 数据元值域代码表的命名为该标准的内部命名，用字母和数字混合表示，编码结构图见图2，编码长度为8位，分两层：

- 编码第一和第二位为第一层，为值域代码表标识，用 C0 表示；
- 编码后六位为第二层，为顺序码，数字大小无含义，取值范围为 000000–999999。

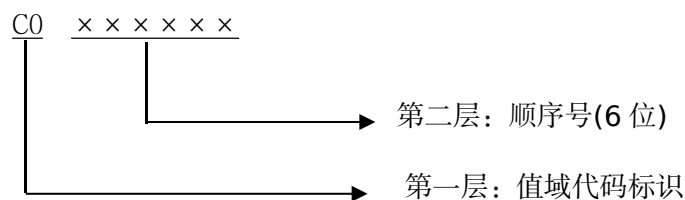


图 2 数据元值域代码编码结构图

6.2 数据元阈值代码表按表 14 描述。

表 14 数据元值域代码表

序号	表号	值域代码表编码	值域代码表名称
1	表 15	CO010001	伦理申请状态代码表
2	表 16	CO020001	抽样方法代码表
3	表 17	CO030001	研究状态代码表
4	表 18	CO040001	数据共享形式代码表

6.3 值域表

数据元按照表15-18取值。

表 15 CO010001 伦理申请状态代码表

值	值含义	说明
0	未申请	
1	已申请，尚未批复	
2	已批复	
9	不详	

表 16 CO020001 抽样方法代码表

值	值含义	说明
1	分层抽样	
2	整群抽样	
3	典型抽样	
4	随机抽样	
5	多阶段抽样	
9	其他	

表 17 CO030001 研究状态代码表

值	值含义	说明
0	未启动	
1	已启动，尚未开始招募	

2	招募中	
3	招募满随访中	
4	暂停	
5	提前终止	
6	完成	
9	不详	

表 18 C0040001 数据共享形式代码表

值	值含义	说明
1	提交研究计划书，审批后共享所需数据	
2	联邦学习	
3	提供全部原始数据	
4	其他	
