



中华人民共和国卫生行业标准

WS 364.1—2011

卫生信息数据元值域代码 第 1 部分：总则

Classification and coding for value domain of health data element—
Part 1: General specification

2011-08-02 发布

2012-02-01 实施

中华人民共和国卫生部 发布

前 言

WS 364《卫生信息数据元值域代码》分为以下十七个部分：

- 第 1 部分：总则；
- 第 2 部分：标识；
- 第 3 部分：人口学及社会经济学特征；
- 第 4 部分：健康史；
- 第 5 部分：健康危险因素；
- 第 6 部分：主诉与症状；
- 第 7 部分：体格检查；
- 第 8 部分：临床辅助检查；
- 第 9 部分：实验室检查；
- 第 10 部分：医学诊断；
- 第 11 部分：医学评估；
- 第 12 部分：计划与干预；
- 第 13 部分：卫生费用；
- 第 14 部分：卫生机构；
- 第 15 部分：卫生人员；
- 第 16 部分：药品、设备与材料；
- 第 17 部分：卫生管理。

本部分为 WS 364 的第 1 部分。

本部分由卫生部卫生信息标准专业委员会提出。

本部分主要起草单位：卫生部统计信息中心、中国人民解放军第四军医大学、中华医学会。

本部分主要起草人：孟群、徐勇勇、胡建平、刘丹红、汤学军、杨鹏、饶克勤。

卫生信息数据元值域代码

第 1 部分:总则

1 范围

WS 364 的本部分规定了卫生信息数据元值域代码标准的数据元值域的编码方法、代码表格式和表示要求、代码表的命名与标识。

本部分适用于卫生信息数据元值域代码标准的编制。

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,仅所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(包括所有的修改单)适用于本文件。

GB/T 10113 分类与编码通用术语

GB/T 18391.1 信息技术 元数据注册系统(MDR) 第 1 部分:框架

GB/T 18391.3 信息技术 元数据注册系统(MDR) 第 3 部分:注册系统元模型和基本属性

WS/T 303 卫生信息数据元标准化规则

3 术语和定义

下列术语和定义适用于本文件。

3.1

值域 value domain

允许值的集合。

[GB/T 18391.1]

3.2

类别 category

具有某种共同属性(或特征)的事物(或概念)的集合。

[GB/T 10113]

3.3

分类 classification

按照选定的属性(或特征)区分分类对象,并将具有某种共同属性(或特征)的分类对象集合在一起的过程。

3.4

线分类法 method of line classification

将分类对象按选定的若干属性(或特征)逐次地分为若干层级,每个层级又分为若干类目,不同层级类目之间构成隶属关系。这种分类方法称为线分类法。

3.5

面分类法 method of area classification

选定分类对象的若干属性(或特征),将分类对象按每属性(或特征)划分成一组独立的类目,每一组类目构成一个“面”。再按一定顺序将各个“面”平行排列。使用时根据需要有关“面”中的相应类目按

“面”的指定排列顺序组配在一起,形成一个新的复合类目。这种分类方法称为面分类法。

3.6

代码 code

表示特定事物(或概念)的一个或一组字符。这些字符可以是阿拉伯数字、拉丁字母或便于电子计算机和人识别与处理的其他符号。

[GB/T 10113]

3.7

编码 coding

给事物(或概念)赋予代码的过程。

3.8

代码结构 code structure

一个完整代码的组成方式和长度的综合表示。

3.9

代码类型 code type

从某一个方面(如含义、结构、长度、组成等)来表示代码的某种特性。

如:从含义上可分为有含义代码和无含义代码;

从结构上可分为层次码和顺序码等;

从长度上可分为等长代码和不等长代码;

从组成上可分为数字代码和字母代码等。

[GB/T 10113]

3.10

无含义代码 unmeaning code

对编码对象只起标识作用,而无任何其他附加含义的代码。

3.11

有含义代码 meaning code

除对编码对象起标识作用外,还具有其他特定含义的代码。

3.12

数字型代码 numeric code

由阿拉伯数字(0~9)构成的代码。

注:此种类型的代码仅仅是以阿拉伯数字的形式表示,但不是数值型,不可直接用于计算。

3.13

字母型代码 alphabetic code

由字母构成的代码。

注:其中所称字母通常为英文字母(I、O因与1、0相似,通常不使用)。

3.14

字母数字型代码 alphanumeric code

由字母和阿拉伯数字混合构成的代码。

3.15

层次码 layer code

以编码对象的隶属关系为排列顺序而组成的有层级关系的代码。

3.16

顺序码 sequential code

按照阿拉伯数字或字母的自然顺序来表示编码对象的代码。亦称“流水码”。

注:通常情况下,顺序码是连续的,代码之间不出现断点。但在特殊情况下,可采用等距离(间隔)跳跃式编码。

3.17

系列顺序码 alignment-sequence code

根据编码对象属性(或特征)的相同或相似,将编码对象分为若干组。再将顺序码分为相应的若干系列(也称为“段”),并分别赋给各编码对象组。在同一系列内对编码对象连续编码,并预留扩展空间。这样编制的代码称为系列顺序码。

3.18

等长代码 code of equal length

在同一个代码体系中,所有编码对象的代码长度都相等。

3.19

不等长代码 code of different length

在一个完整的代码体系中,代码总长度不完全相同。

3.20

标识符 identifier

在特定语境中,可唯一性地标识与之相关联的事物的一系列字符,可看做用来识别特定对象的数据元的编码值。

[GB/T 1839.3]

4 数据元值域的编码方法

4.1 代码结构

数据元值域代码结构设计遵守以下要求:

- 代码结构设计注重代码的标识作用,避免承载过多的信息,以保证结构的简练。
- 代码结构符合信息处理的基本方法,尽量与系统内、外的相关标准结构协调一致。
- 代码结构确保代码的添加、删除和修改不破坏代码结构。
- 代码应采用便于使用的符号。

数据元值域代码结构描述遵守以下要求:

- 应明确描述所采用的代码种类、代码结构以及编码方法。
- 层次码的代码结构还可用示意图表示,如图 1 所示。
- 当代码结构复杂时,可用示例说明。

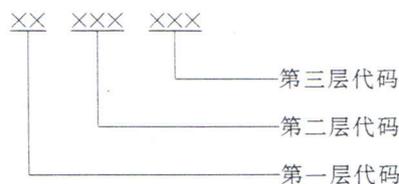


图 1 数据元值域代码结构示意

4.2 代码长度

4.2.1 设计要求

数据元值域代码长度设计遵守以下要求:

- 在保证需求的前提下,代码长度应尽量简短。
- 尽可能使用等长代码,不宜使用不等长代码。

c) 代码预留空间应满足编码对象的发展要求。

4.2.2 计算公式

代码长度的计算公式：

$$N = \sum_{i=0}^n \log a_i Q_i \dots\dots\dots (1)$$

式中：

- N —— 代码总长度；
- n —— 层次码的层数；
- i —— 第 i 层(或第 i 面)；
- a_i —— 第 i 层(或第 i 面)的代码字符集的字符个数；
- Q_i —— 第 i 层(或第 i 面)的编码对象的总数量。

当为顺序码或系列顺序码时, i=1 则以上公式变为：

$$N = \log a Q \dots\dots\dots (2)$$

4.3 代码类型及形式

代码类型及形式应符合下列要求：

- a) 代码字符可选择使用数字型代码、字母型代码、字母数字型代码；
- b) 代码字符应正确无误、易认易读。应避免使用容易被混淆和误解的字符。在一个标准中, 音相近、形相似的字符应避免同时出现, 如字母“l”与数字“1”；
- c) 代码最好全部用数字或全部用字母表示。字母数字混合的形式一般在特殊位置使用, 不宜在随机的位置使用；
- d) 采用数字型代码时, 如果有收容类目时其代码采用末位数字为“9”的代码；
- e) 选用顺序码时, 代码一般要等长。例如: 用 001~999, 而不用 1~999; 采用层次码时, 同层次的代码要等长；
- f) 在同一个标准中, 代码书写形式要一致, 包括字母的大、小写, 代码的字体字号。

5 代码表格式和表示要求

5.1 代码表格式

数据元值域分类与代码表(或代码表)应以表格的形式列出。依据 WS/T 303 要求, 表格由代码栏(代码指编码值, 可简称为“值”)、编码对象名称栏(在代码表中可简称“值含义”)、说明栏组成, 并可根据实际需要适当增减栏目。

当表格比较简单时, 为了减少篇幅, 可以在一页中排两列或两列以上的表格。

5.2 代码表书写要求

数据元值域代码表书写要求如下：

- a) 代码(值)栏: 代码一般在代码栏内左起顶格书写; 当代码层次较多时, 代码栏可按层次再进行划分；
- b) 名称(值含义)栏: 编码对象名称在名称栏内左起顶格书写, 每个编码对象名称占一行。当编码对象名称较长时, 可延续至下一行, 延续部分要与上一行对齐。采用线分类法时, 第一层次的编码对象名称左起顶格书写, 第二层空一个字, 依此类推；
- c) 说明栏: 说明的内容在说明栏左起顶格排。

代码表的具体示例参见附录 A 中表 A.1 和表 A.2。

6 代码表的命名与标识

6.1 命名

代码表应具备在特定领域背景上获得权威认可的名称。代码表的名称应准确反映代码表作为数据元表示类属性之一的特征,不应放大或缩小其使用范围。

代码表的名称应简洁,传达明确的语义,体现代码表的本质内容。

6.2 标识

代码表应该在特定使用领域内具有唯一的标识符,用来识别表示数据元值域的编码体系。卫生信息数据元值域代码表的标识符根据卫生信息的归类确定。结构为:

CV+7 位数字,组成总长度为 11 位的字母数字混合码,包括 2 个分隔符号“.”。

按类别代码、顺序号从左向右顺序排列。结构如图 2 所示。

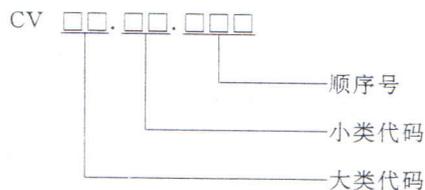


图 2 代码表标识符结构

CV——表示数据元值域的编码值,Coded Value;

大类代码——2 位数字,表示卫生信息分类中第一层(大类)的代码,见表 A. 1。

小类代码——2 位数字,表示卫生信息分类中第二层(小类)的代码,见表 A. 2。小类代码和大类代码之间加“.”隔开。

顺序号——用 3 位数字表示,代表每一类别下值域代码表的序号,数字大小无含义,从 001 开始编码。顺序号与类别代码之间加“.”隔开。

附 录 A
(资料性附录)
分类与代码表的编排

分类与代码表编排的示例见表 A.1 和表 A.2。

A.1 代码结构为一层时代码表的编排方式示例见表 A.1。

表 A.1 老年人生活自理能力自我评估代码表

值	值含义	说 明
1	可自理	0~3 分者为可自理
2	轻度依赖	4~8 分者为轻度依赖
3	中度依赖	9~18 分者为中度依赖
4	不能自理	≥19 分者为不能自理

A.2 代码结构为多层时代码表的编排方式见表 A.2。

表 A.2 药物类型代码表

代 码	名 称
01	抗生素类抗感染药物
0101	青霉素类抗生素
0102	头孢菌素类抗生素
0103	氨基糖苷类抗生素
.....	
02	非抗生素类抗感染药物
0201	磺胺类药及增效剂
0202	喹诺酮类抗感染药
0203	抗结核麻风分枝杆菌类药
.....	
.....