

## 第3章

# 医学信息标准化

### 学 习 目 的

1. 了解标准、标准化、分类及编码的基本概念；
2. 掌握医学信息学常用标准。

### 引 言

随着国内外医学信息学研究的不断深入，特别是医疗卫生信息化进程的加快，生物医学信息的处理、存储和传输的标准及其标准化建设的重要性日益凸显。医学信息标准化的建设是医疗卫生保障体系建设的重要基础和技术支撑。

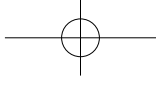
## 3.1 标准与标准化

在科学领域中，医学是专用名词极多、极深奥、极难以统一规范的领域之一，却又是与人类生命、健康关系极密切的领域之一。随着医疗卫生信息化的进展，传统手工操作时代“非标准化”的矛盾日益突出。远程医疗、医疗保险、社区卫生、区域卫生的发展，要求医学信息必须跨部门和地区进行交互，这更需要标准化。因此，标准化成了医疗卫生信息化的首要任务。

### 3.1.1 标准

国家标准 GB/T20000.1—2002 给“标准”下的定义是：“为了在一定的范围内获得最佳秩序，经协商一致制定并由公认机构批准，共同使用和重复使用的一种规范化文件。”在此定义后有一条附注：“标准宜以科学、技术和经验的综合成果为基础，以促进最佳共同效益为目的。”

因此，标准应具有如下一些特性：第一，它是一种规范化文件；第二，具有共同使



用和重复使用性质;第三,文件的制定必须有一定程序,经协商一致,并由公认机构批准,而这“公认机构”是负责为公共和常用事物的活动及结果制定和提供规则、指导原则的;第四,制定标准的目的是为了在“一定范围”内获得最佳秩序。

标准(standard)是指获得一致同意的,并由公认权威机构(如国际标准化组织, International Organization for Standardization, ISO)、国际电工委员会, International Electrotechnical Commission, IEC)认可的文件。这个权威机构负责为公共和常用事物的活动及结果制定规则及指导原则,其宗旨是使应用该标准的环境达到最佳的有序状态。

### 3.1.2 标准化

#### 1. 标准化的定义

国家标准 GB/T20000.1—2002,采用了国际标准化组织(ISO)的定义:“标准化是为了在一定范围内获得最佳秩序,对现实问题或潜在问题制定共同使用和重复使用的行为规范的活动。”

在“标准”这一名词后加上“化”,则转化为动词,表示转变为一种性质、状态或活动。因此,标准化首先应是制定和实施某一规范的活动;其次,该规范的内容涉及当前的现实问题或将显露的潜在问题;再次,制定该规范的目的是为了在一定范围内达到最佳秩序。

#### 2. 标准化的特征

根据标准化的定义,标准化具有如下特性。

(1) 明确的域(domain):即范围。某一标准一定是针对和适应某一域的需求。域的分界必须清晰,内容必须明确。就是说,域的边界必须清晰到能够明确判定什么是属于该域的,什么是不属于该域的。例如,研究药品的标准化问题,首先要确定是否包括中药、原料药、试剂、医院制剂等。

(2) 唯一性与完整性:所谓的“唯一性”,指在标准化的体系中,无论是一个对象,或是一组对象,应该有、而且只能有一个确定的代码与之对应。所谓完整性,就是在某一个“域”内的标准化体系应涵盖它所有的对象。由于对事物的理解总是在不断深化过程中,事物也总是在发展变化中,所以,标准化系统总是不断地修订和完善。

信息编码的完整性往往用设置“收容组”编码来保证。所谓“收容组”是指在相应位置设置一个其他类的特殊编码,当客观事物出现了没有对应编码情况时,可以将其归于相应类别的其他类编码。

(3) 权威性:标准必须在一个宽广的范围内被认可和执行才有意义,而一个标准的诸多内容不可能在这样宽广范围内被所有的用户完全认同和接受,因此必须由权威部门制定和颁布,并带有明确的约束性,甚至是强制性。因此权威性是标准化天生的特质。

西方发达国家中许多标准往往是由一家或几家技术先进的公司率先发起制定和使

用的,再作为企业内标准,然后被其他公司所仿效与遵循,最终成为行业标准直至国家、国际标准,这在高新技术产业中相当普遍。

### 3. 元数据与数据元

讨论信息标准化,必须先了解两个基本概念:元数据和数据元,它们是信息表达和信息标准化的基础。

#### 1) 元数据

元数据是对信息资源的规范化描述,它是按照一定标准,从信息资源中抽取相应的特征组成的一个特征元素集合。这种规范化描述可以准确和完备地说明信息资源的各项特征。不同类型的数据资源可能会有不同的元数据标准。

元数据内容标准从数据结构、格式、语义、语法、功能各方面来制定标准,促使了数据的规范化、标准化,提高了数据库的建库量。

元数据为信息的管理、发现和获取提供一种实际而简便的方法。通过元数据,人们能够对信息资源进行详细、深入的了解,包括信息资源的格式、质量、处理方法和获取方法等各方面细节,并有数据维护、历史资料维护的功能(对数据生产者而言)。元数据标准可适用于资料共享、数据发布、数据集编目、数据交换、网络查询服务等,也是数据集元数据整理、建库、汇编、发布的标准格式。元数据标准的制定是为提高数据库建库质量,使数据加工达到规范化、标准化,促进科学数据的标准化,加强数据交流与共享。

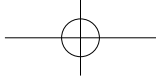
#### 2) 数据元

数据元又称数据元素,是用一组属性描述定义、标识、表示及允许值的数据单元。在一定的语境下被认为是不可再分的最小数据单元。通常用于构建一个语义正确、独立且无歧义的特定概念语义的信息单元。数据元一般由三部分组成。

(1) 对象类:思想、概念或真实世界中的事物的集合,它们具有清晰的边界和含义,其特征和行为遵循同样的规则。对象类是人们希望研究、搜集和存储它们的相关数据的事物,比如汽车、人、房屋、订单等。

(2) 特性:对象类中的所有成员共同具有的一个有别于其他数据元的显著特征。特性是人们用来区分和描述对象的一种手段。特性的例子包括颜色、性别、年龄、收入、地址等。

(3) 表示:它描述了数据被表达的方式。表示与数据元的值域关系密切。一个数据元的值域是数据元的所有允许值的集合。例如,对于“个人所得税金额”这个数据元,它的值域可以是一系列非负整数(带有货币单位),这是一种非枚举型的值域。而“个人所得税比率”对收入进行分段划分,并给每一段赋予一个比率,它的值域就是这些比率的集合。此时称它的值域为枚举型值域。如果脱离值域,“个人所得税”实际上是一个数据元概念,即一个对象类与特性的组合。因此,一个数据元概念是由一个对象类和一个特性组成,一个数据元是由一个数据元概念和一个表示组成。



### 3.1.3 标准化发展历程

#### 1. 古代的标准化

早在远古时代，人类就表现出无主观意识的标准化行为，例如世界各地出土的石器时代的石刀、石斧形态都惊人地相似，这是人们在长期生产实践中，共同探索、模仿所形成的一些约定俗成的概念和规范。

在漫长的进化过程中，当人类产生了语言，创造了文字，这标志着人类能主观去统一、规范事务和概念，体现了人类初期有意识的、朴素的标准化活动。

古代标准化典范有古罗马战车的建造者，他们设置车轴的长度基于两匹马的宽度，这轴长便成为开发道路的标准，也是制造战车乃至马车、客车轮子距离的规范，现在标准铁路轨道规范正是源自这种古老的启示。

在我国古代，秦始皇统一中国后，先后颁布政令，对度、量、衡、文字、货币、道路等进行全国范围的、空前的标准化，“车同轨，书同文”成为古代世界标准化的杰出典范。而北宋时期，毕昇于公元 1041—1048 年发明的活字印刷术则孕育了近代标准化的原理和方法，并成为人类文明的光辉范例。

#### 2. 近代工业标准化

自 18 世纪末英国发生工业革命以来，大机器生产方式促使了工业标准化飞速发展。典型范例有伊莱·惠特尼于 1789 年发明的工序生产方法，设计了可通用互换零件组装步枪，被誉为“标准化之父”。此后层出不穷的标准深入生产参数系列化、行业标准化、作业和管理标准化的各个领域，成为社会有序化的支柱之一。1959 年，ISO 决议将每年 10 月 14 日定为“国际标准日”。

#### 3. 信息时代标准化

当人类进入信息化时代，标准化变得尤为重要和紧迫。ISO、IEC 等国际标准化组织逐步促成了全面的信息技术标准化体系，涉及信息处理、软件与软件工程、数据交换与通信网络技术、信息安全等各个方面，极大地顺应和促进了信息化日新月异的飞速发展。

早期医院信息管理局限在单个部门，以单机版为主，例如检验结果数据处理、收费处的划价收费等，对标准化的需求不显著。即便是整个医院的信息管理系统，为了进行数据的交换，其标准化也是粗糙和不规范的。

随着医疗卫生信息化的深入发展，医院的信息管理必须集成来自不同厂家的不同应用软件，以实现全院的信息交互和共享。医疗保险、区域卫生信息管理、公共卫生平台则要求信息在不同医院之间，进而在一个城市、一个省，甚至在全国实现共享。这时，数据、通信、接口等一系列信息标准化的问题就成为当务之急。

### 3.1.4 信息标准化

狭义的信息标准化是指信息表达上的标准化，是在一定范围内人们能共同使用的

对某类、某些、某个客体抽象的描述与表达。医学信息的标准化是特指信息标准化在医学领域的具体应用。语言文字可能是人类最早实现标准化，并且连续几千年持续不断努力维护其高水准标准化程度的实例。

计算机广泛引入信息处理技术以来，信息标准化的表达方式常常用数字、字符等抽象符号表达，这是因为计算机处理起这些抽象符号较之信息的其他表达方式（如语言、文字、图形、图像）更节省、更快捷、更方便。

广义的信息标准化指对整个信息的处理，包括信息传递与通信、数据流程、信息处理技术与方法、信息处理设备等。

### 1. 信息的表达

信息表的类标准是信息标准化的基础，也是最常见的。像分类编码类，即名称和内涵的标准化、代码化。下面提到的 ICD-10（International Classification of Diseases，国际疾病分类）、SNOMED（Systematized Nomenclature of Human and Veterinary Medicine）等均属于此类。

### 2. 信息的交换

信息的交换是要解决不同的系统之间或不同的部门、企业之间对信息共享的问题。信息交换的标准往往比信息的表达要复杂。信息交换标准更注意信息的格式，而忽略信息的内容。从信息交换的角度来看，应同时包括信息表达的标准化和信息内容的可相互理解性。这多少有一点像语言学习中的句法与词法的关系。

因为远程医疗和区域卫生信息系统日益流行，所以信息交换标准也变得越来越重要。HL7（health level 7）、XML（extensible markup language）和 DICOM3（digital imaging and communication in medicine）是卫生信息交换类标准中常见的典型例子。

HL7 是一整套网络信息交换标准中专门用于医疗卫生信息交换的标准。因为用户层是第七层通信协议，因此叫 health level 7。HL7 主要是规定了当一个事件发生时，相应的信息应该用什么样的格式通知需要信息的一方。

XML 是更为广泛的一个信息交换标准，是 extensible markup language（可扩展置标语言）的缩写，是 W3C（World Wide Web Consortium）组织于 1998 年 2 月发布的标准。W3C 组织制定 XML 标准的初衷是，定义一种互联网上交换数据的标准。是国际互联网上数据交换标准语言，有取代 HTML 的趋势。

DICOM 标准是由 ACR（American College of Radiology）及 NEMA（National Electrical Manufacturers Association）所形成的联合委员会于 1983 年以后陆续发展而成的医学数字成像和通信标准。简而言之，DICOM 是有关医学影像的表达方式（存储）和信息交换的标准。

### 3. 信息的处理与流程

还有许多标准是规范信息处理流程的，这一类标准的制定常常对信息系统的开发与推广有着十分重要的意义。众所周知，我国财务电算化是信息化产业较为成功的领域之一，其根本的原因，就是财务管理信息处理流程的规范化和标准化。另一个正面



的例子是中国人民解放军医院能在短短的两三年内成功地统一使用同一个医院信息系统产品（在美国是 VA 系统），最重要的就是军队有足够的权威统一制定和推行信息处理与流程的标准。

疾病诊断相关组（diagnosis-related groups, DRG）描述由某医院治疗的病人的类型，虽然表现出来的是用代码来区别相关的疾病组，实际上，其本质是规定了一套信息处理的流程、逻辑与标准。

## 3.2 分类与编码

分类和编码是信息标准化的主要方法之一。分类是某一领域内概念化的序化和原理的序化。编码是指定一个对象或事物的类别或者类别集合的过程。

### 3.2.1 基本概念

#### 1. 分类

分类（classification）是为了某一目的，依据某一原理，采取一种分类准则，将依从这一准则的、具有共同属性和特征的信息归并在一起，并依从这一准则有序地排列。

分类的准则首先取决于某一领域的应用目的，然后依从于这一目的，根据某一概念分类，再将这些类别依照属性关系有序排列。所谓属性关系表现为甲包含了乙和丙，即乙或丙是甲的一种。同时，这种概念的序化系统，或明确或潜在地反映了其中包含某种原理的有序化。

#### 2. 编码

编码（coding）是指定一个对象或事物的类别（如多轴分类）或者类别集合的过程。类别通常是用代码来表示的，即将一个表示对象或事物信息的某种符号体系（文字）转换成便于人或计算机识别和处理的另一种符号体系（代码）的过程。

### 3.2.2 分类与编码的基本原则

在对信息进行分类和编码时，应遵循以下原则：

（1）科学性。要以当代先进的医学科学水平为基准，分类目的有科学依据，分类轴心要体现对象的本质特性，编码有科学意义。

（2）系统性。分类的对象必须按照其内在的特性和规律进行排序，并形成科学严谨、结构合理、层次分明的分类体系。

（3）准确性。分类的类目应独立明确、相互排斥、互不包括。类目下的亚目，从属关系清楚、层次分明。代码确切有序，不要随意空码、跳码。

(4) 唯一性。应确定统一的代码元素集, 严格做到一码一义, 避免一码多义或一义多码, 使整个分类编码系统井然有序、精确无误。

(5) 冗余性。一个分类编码系统除了应包括现有的所有对象外, 还应预留一定的空项, 以适应发展中不断涌现出来的新对象。这些预留的空项又必须依据分类编码原理和内在属性关系而定, 新的对象将参照与原有对象的属性关系填充到相应的预留空项中, 而不是简单堆放在原系统之后。

(6) 结构化。代码与对象的特性以及内涵应有结构化的对应关系, 代码的不同位置标识了对象的特性及其与周围的层次关系。

(7) 实用性。分类和代码都要有实用价值, 符合实际需要。它不能过于简单而失去准确性, 又不能过于烦琐而应用困难。

(8) 可操作性。分类编码应力求简单明了, 易于学习掌握, 同时要便于计算机输入。

### 3.2.3 分类与编码的方法

#### 1. 分类的方法

数据分类的基本方法有 3 种: 线分类法、面分类法、混合分类法。

(1) 线分类法: 所谓线分类法(表 3-1), 是依据某一属性或特征, 逐层分解展开, 形成分类体系, 例如 ICD 分类法。线分类法又称树型分类编码, 是最常使用的一种编码方法。树型编码是将客体逐层细化的编码方法。每个子层的细化只与其上一层有关, 而其他主层无关, 就像一棵倒长的树, 由树根、树干、树杈直至树叶组成。

线分类法的优点: 层次性好, 能较好地反映类目之间的逻辑关系; 实用方便, 既符合手工处理信息的传统习惯, 又便于电子计算机处理信息。

线分类法的缺点: 结构弹性较差, 分类结构一经确定, 不易改动; 效率较低, 当分类层次较多时, 代码位数较长。

表 3-1 线分类法实例

代 码	类 型 名 称
68	医疗器械
6830	医用 X 射线设备
683010	通用 X 射线诊断设备
68301001	便携式 X 射线诊断机
68301002	便携式交直流两用 X 射线诊断机
.....	.....
683020	X 射线断层诊断设备
68302001	直线轨迹断层 X 射线机
68302002	多轨迹断层 X 射线机
.....	.....

(2) 面分类法: 所谓面分类法, 是将所选定的分类对象的若干属性或特征视为若干个“面”, 每个“面”中又可分成彼此独立的若干类目。使用时, 可根据需要将这些“面”



中的类目组合在一起，形成一个复合类目。例如在医院财务统计中，收费项目的“面”包括了住院费、药品费、手术费、检查费等类目；付费方式的“面”中包含了医疗保险、公费、自费等类目。这两个方面类目组合，就形成复合类目，如全年“医疗保险病人的药品总费用”“自费病人的检查总费用”……从而为医院的“合理收费”提供统计依据。

面分类法的优点：具有较大的弹性，一个“面”内类目的改变，不会影响其他的“面”；适应性强，可根据需要组成任何类目，同时也便于机器处理信息；易于添加和修改类目。

面分类法的缺点：不能充分利用容量，可组配的类目很多，但有时实际应用的类目不多；难于手工处理信息。

(3) 混合分类法：混合分类法是将线分类法和面分类法组合使用，以其中一种分类法为主，另一种作为补充的信息分类方法。例如药品的分类法就是混合分类法。

### 2. 编码的方法

(1) 数字编码 (number codes)：将一个未用过的数字给予一个新类别，编定此种类别的数字只能用在特定的类别。常用于病历号的编码。

(2) 助记编码 (mnemonic codes)：由一个或多个和类别有关的字符组成，这种代码编码容易，用户易于记忆，使用方便。例如使用英文词汇的首字母组合作为代码：CHD (congenital heart disease) 表示先天性心脏病；EEG (electroencephalogram) 表示脑电图。

(3) 阶层编码 (hierarchical codes)：对每个附加层次的细节进行延伸。阶层编码在相关细节的层次和相关的母阶层产生信息。例如，ICD-10。

(4) 并排码 (juxtaposition codes)：是由区段 (segment) 所组成的合成码，每个区段提供相关类别的特征。例如，ICPC (international classification of primary care) 分类系统中，“N”表示神经系统疾病。

(5) 组合码 (combination codes)：根据排序原则将不同的类别进行编码，并组合成一个编码的分类系统。例如，供应室的器械包分类。

(6) 加值码 (value addition codes)：利用二进制法来加总代表不同的分类码。例如，20= 抽烟，21= 过胖，22= 高胆固醇，3= 抽烟且过胖。

## 3.3 医学信息标准

医学信息标准主要包括医学信息的标准、医学信息交换的标准、医学信息处理与流程的标准等。

### 3.3.1 制定医学信息标准的艰巨性

医学信息因其具有种类繁多、数量庞大、量化困难等特点，所以要做到真正的医学信息标准化是一项十分艰巨的任务。

(1) 医学信息面广量大、种类繁多,包括数值、文字、图像、声音、气味等,各种类别的信息表示内容不一、表达形式不一,难以标准化。

(2) 患者信息数量庞大,而且十分复杂细致,个性突出,共性和可重复性差,加之病人流动频繁、病情多变,形成极为复杂的海量信息。

(3) 医学信息量化困难。它不同于工程信息,各变量的相互关系及变化规律难以用数学语言表达。例如头痛的性质和程度会因患者的个性特质、痛域高低不同而表达不一。

(4) 自然语言标准化困难。病历中的病史、病程记录、病情讨论分析多采用自然语言,常因医师的学术水平、文化素养、书写习惯不同而迥然不一,自然语言标准化是全球共同的难题。

(5) 共享性突出。复诊、转诊、会诊需要共享同一患者信息;电子病历需要共享不同专业和医生的信息;社区医疗、区域医疗则要共享不同地域的信息。范围越大,标准化难度越大。

因此,医学信息标准化是一项十分艰巨的任务。

### 3.3.2 医学信息标准的类型

医学信息标准是一个宽泛的范畴,类型也有多种,主要种类如下:医学信息的标准、医学信息交换的标准、医学信息处理与流程的标准和医学信息应用软件和硬件的标准。

(1) 医学信息的标准。医学信息的标准主要是指信息表达类标准,是标准化的基础,它更注重信息本身的内容,它分门别类地定义各个医学专有名词的代码,形成医学分类系统或医学词汇表,例如在 3.4 节介绍的 ICD、SNOMED 等。

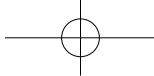
(2) 医学信息交换的标准。制定信息交换标准的目的就是解决不同系统之间数据难以准确、精细、完整地交互和通信的问题,为此,双方所传输的信息的语法和语义必须一致,才能“读懂”和“交流”。信息交换标准比信息表达类标准要复杂,因为需更注意信息交换时的格式和规则。例如在 3.5 节介绍的 HL7、DICOM 等。

(3) 医学信息处理与流程的标准。医学信息处理与流程的标准对于医学信息系统的开发与推广应用有着十分重要的意义,它规范了一个系统或不同系统之间信息的处理流程。

(4) 医学信息硬件与软件的标准。医学软件的标准大致包括以下三大方面:一是软件产品的标准;二是生产和管理软件工程的标准;三是软件开发环境的标准。这中间又以医学信息软件产品的标准最为困难。

医学信息软件种类繁多,这里以我国应用最广泛的医学信息软件——医院信息系统(HIS)为例予以说明。卫生部曾于 1997 年颁布了《医院信息系统软件基本功能规范》,对 HIS 的标准化、规范化起了重要指导作用;于 2002 年又重新修订颁发了《医院信息系统基本功能规范》,该规范强调了标准化是信息化的基础,并将 HIS 中数据、数据库、数据字典编码标准化作为一个独立章节予以阐述,突出了标准化在医院信息化建设中的重要地位。

硬件的标准化范围很广,例如计算机的标准化、网络布线的标准化、网络设备的



标准化，存储设备的标准化及原材料的标准化等。医学信息硬件与一般信息硬件相同，是医疗卫生信息系统建设的基础保障。

## 3.4 医学信息学常用标准

医学信息的标准是医学信息标准化的基础，本节介绍国际主要的几个标准：ICD、SNOMED、DRG、UMLS、ICPC、RCC、LOINC、MeSH。

### 3.4.1 国际疾病分类

#### 1. ICD概述

国际疾病分类是根据疾病的某些特征，按照规则将疾病分门别类，并用编码的方法来表示的系统。ICD是由世界卫生组织主持编写、发布，并要求各成员国在卫生统计中共同采用的国际权威的疾病分类方法。目前全世界通用的是第十次修订本《疾病和有关健康问题的国际统计分类》，通称为ICD-10。

#### 2. ICD发展简史

ICD已有120多年的发展历史，最初为对死亡率进行统一登记，1893年，国际统计研究所专门学会提出了一个分类方案《国际死亡原因编目》，并于1900年出版ICD第1版，列出192种疾病，以后基本上每10年修订一次。1940年第6次修订版由世界卫生组织（WHO）承担，首次引入了疾病分类，并强调继续保持按病因分类的哲学思想。1975年，出了第9次修改版，即全世界广泛使用的ICD-9。1992年，出了第10次修改版本ICD-10，共列出了14400种疾病，更名为《疾病和有关健康问题的国际统计分类》。

我国卫生部早在1987年就发布文件，要求医院采用ICD-9作为疾病分类统计报告标准，并于1993年由国家技术监督局发布《疾病分类与代码》的国家标准，将ICD-9完全等同于国家标准。北京协和医院世界卫生组织疾病分类合作中心（WHO Collaborating Center for the Family of International Classifications of Diseases）负责有关疾病分类的中文事宜，并协助卫生行政部门收集控制疾病分类资料的质量。

#### 3. ICD分类原理与方法

ICD分类依据疾病的四个主要特性，即病因、部位、病理和临床表现（包括症状、体征、分期、分型、性别、年龄、急慢性、发病时间等），每个特性构成了一个分类标准，形成一个分类轴心，因此ICD是一个多轴心的分类系统。

ICD的主要分类编码方法如下：分类有三个层次，首先是类目，类目下分亚目，亚目下分细目。通常在同一个层次的分类都是围绕疾病的一个特性，即围绕一个轴心展开的（个别情况有两个轴心）。

（1）类目：三位数编码，包括一个字母和两位数字。例如，S80表示小腿浅表损伤，