

肿瘤登记编码审核中的常见问题及处理

Common Problems on Cancer Registration Coding Checked by IARC-CHECK Program and Their Solutions

CHEN Jian-guo, ZHU Jian

陈建国^{1,2}, 朱 健¹

(1.启东肝癌防治研究所, 江苏 启东 226200;

2.南通大学肿瘤医院/肿瘤研究所, 江苏 南通 226361)

摘要: 全文介绍 IARCcrgTools 2.05 软件的主要设计思想和主要模块, 并着重描述 IARC-CHECK 程序的输出文件、出错文件、警示文件的形式和意义。结合软件的应用, 列举了肿瘤登记资料编码审核中常见的问题、原因及其处理解决的方法。

关键词: 肿瘤登记; CHECK 程序; 国际疾病分类; 肿瘤学编码

中图分类号: R730.1 **文献标识码:** A **文章编号:** 1004-0242(2012)07-0502-05

目前我国县(区)以上肿瘤登记处已接近 200 家, 恶性肿瘤的登记、报告、编码及上报已成为肿瘤登记处的常规工作; 但资料能够被应用采纳的还不足 50 家^[1], 可见肿瘤登记资料的质量亟待提高。肿瘤编码涉及综合知识的应用, 为帮助审核人员对登记报告资料进行编码审核, 国际癌症研究机构(IARC)和国际癌症登记协会(IACR)联合编制了专门的软件用于资料的有效性和逻辑性的审核。现分析讨论肿瘤编码审核发现的常见问题及处理。

1 编码审核软件 CHECK 及设计思想

1.1 软件及模块

目前所采用的软件为 IARCcrgTools 2.05 (2008) 版^[2], 在 IACR 的网站免费下载, 地址: http://www.iacr.com.fr/IARCcrgTools_205.ZIP。该软件需要的计算机环境为: Windows 2000/XP 以上, 5 MB 以上硬盘空间, 64 MB 以上内存。

程序所需要的批处理文件通常为文本文件(.txt)及固定长度记录(fixed length record)。软件可提供在线帮助(Online help)。

IARCcrgTools 2.05 主要有 3 个模块: (1)Con-

versions program(转换程序); (2)IARC-CHECK program(审核程序); (3)IARC/IACR multiple primary check program(多重原发肿瘤审核程序)。

1.2 设计思想

1.2.1 对不同版本的肿瘤编码系统进行转换

该软件的转换(Conversions)模块可以方便地进行 ICD 多种版本的转换, 可以根据实际情况选择。目前可进行相互转换的 ICD 版本及 ICD-O 版本见图 1。



图 1 IARCcrgTools 软件中的转换(Conversions)模块

收稿日期: 2012-04-05

E-mail: qdchenjg@gmail.com

1.2.2 对肿瘤编码中的逻辑问题提出警示

通过模块 IARC-CHECK 程序,审核登记资料中的变量的有效性(Validity)及一致性(consistency)。

系统在对肿瘤登记资料进行审核后,对明显的错误,例如“无效的年龄”(invalid age)、无效的 ICD-O-3(M)编码(invalid ICD-O-3(M) code);即形态学编码、无效的性别/部位组合(invalid sex/site combination)等信息,直接提出出错(error)警告,该文件名的后缀(扩展名)为.err。

对于可能的问题,也即在“一般情况下”不太可能(不常)出现的变量间的组合,审核程序会提出警示(warning)警告,该文件名的后缀为.chk。

2 输出文件的形式和意义

若假定工作文件夹在“J:\QD2008CanSample\”,文件为“2008CanIncidence.txt”,输出文件名为“2008Can.out”(用户自定义),则软件正常运行后将产生4个主要文件。这些文件本质上均为文本文件(注意:文本中应去掉中文字符),可参见 IARC Check Program 帮助(help)菜单中对输出文件的说明^[2]。

2.1 运行后的汇总文件

可以在工作文件夹“J:\QD2008CanSample\2008Can.txt”中找到,文件名为:“2008Can.txt”。可用记事本打开,其内容提示,原输入文本文件中,有多少条记录已经处理,有多少条记录有出错,有多少条记录提出警示。

2.2 主要结果输出文件

主要结果输出文件(output data file)可以在工作文件夹“J:\QD2008CanSample\2008Can.out”中找到,该文件就是在软件运行时用户自定义的“2008Can.out”文件。用记事本打开,该文件中包括输入的登记资料中全部通过的记录。假定原有3565条记录,若没有出错记录和警示记录,则该文件中还是3565条记录;若出错记录有27条、警示记录有13条,则“2008Can.out”文件中的记录为3538条,即27条出错记录已被排除在外,而13条警示记录仍然被包括在文件“2008Can.out”中。

2.3 出错文件

出错文件(error file)可以在工作文件夹在“J:

\QD2008CanSample\2008CanIncidence.err”中找到,注意:该文件名不叫“2008Can.err”!而是与原始输入文件名“2008CanIncidence”同名,只是把后缀“.txt”改为了“.err”。用记事本打开,可见第1行为软件操作的日期和时间;第2行提示输入文件的路径和文件名;第3行提示:The follwing records must be checked,意为:下列的记录必须作审核,然后显示有错误的27条记录的具体信息。注意最右栏中,提示了错误的类型。例如,无效的年龄(invalid age)、ICD-O-3形态编码错误(ICDO-3(M))、性别/部位的组合错误(sex/site combination)等等。注意:这27条“出错”的记录,未被写入输出文件“2008Can.out”。

2.4 警示文件

警示文件(warning file)可以在工作文件夹“J:\QD2008CanSample\2008CanIncidence.chk”中找到,注意该文件与原始输入文件名“2008CanIncidence”同名,只是把后缀“.txt”改为了“.chk”。用记事本打开,第1、第2行信息同样提示运行日期和原输入文件名,第3行提示:The follwing records should be checked,意为:下列的记录应当作审核,然后显示有13条记录(13 individuanl records)、有14处警示(14 warnings)之问题需要审核。注意:这些“警示”的记录,同时已被写入输出文件“2008Can.out”。

3 常见问题和处理

肿瘤的发生有显著的人群特点和差异,某些肿瘤为男性特有,某些反之;某些肿瘤有年龄特点和形态学的差别,这些特征间的组合可以通过逻辑性与专业知识来判断。在日常的登记工作中,有部分已经通过登记处专业人员的专业知识,及时发现并予以更正;而在CHECK程序中发现的问题,仍然需要人工进行再核对,作出最后的判断和核准。

CHECK把常见的不符合逻辑的变量(字段或项目)间的组合可归纳为9种情况,即:年龄/发病和出生日期、年龄/部位/组织学、部位/组织学、性别/部位、性别/组织学、行为/部位、行为/组织学、分化程度/组织学、以及诊断依据/组织学。这些不符合逻辑的变量主要涉及年龄、性别、部位、组织学(形态学)、行为、分化程度(分级)及诊断依据等。现简述常见的

原因及处理方法。

3.1 年龄

年龄的计算以发病年龄为例:发病年龄=(发病日期-出生日期)/365.25。在 Excel 表格中,可以采用公式:INT (“发病日期”-“出生日期”/365.25),INT 为“取整函数”。或者 DATEDIF (“出生日期”,“发病日期”,“Y”)。

3.2 性别

在男女性都有可能发生的肿瘤部位,例如肝癌、肺癌、直肠癌、白血病等,当发生性别编码错误,Check 软件是无法“发现”的。只有当男性和女性“特有”的肿瘤,例如男性肿瘤“阴茎癌”、“前列腺癌”,女性肿瘤“宫颈癌”、“卵巢癌”等,发生在“异性”身上,才有可能发现。

工作人员应增强工作责任心,加强编码核对,以避免性别错误。除非有特殊说明,男性通常编为“1”,女性通常编为“2”,性别不明为“9”(几乎没有)。当发现性别异常编码错误,首先应当确定是“性别”错误,还是部位错误。例如,当碰到性别编码为“1”(男)、而部位编码为“C51.9”(外阴恶性肿瘤)时,就要查清究竟是某女性患外阴恶性肿瘤(C51.9),还是某男性患前列腺恶性肿瘤(C61.9)。

3.3 部位

肿瘤部位是肿瘤登记中的重要项目(变量),也是最复杂的项目,因而是最容易出错的变量。由于肿瘤好发于任何器官的任何部位,而且又衍生出许多的“亚部位”,对此进行编码是非常“头痛”的事情。而且其错误可以出现在肿瘤从诊断、登记、报告及编码的各个环节,因此做好肿瘤登记工作,就要在肿瘤登记的学习、培训、编码、核对、上报的各个环节做好质量控制。

3.3.1 亚部位(“-”编码)

大部分肿瘤部位有亚部位(sub-site),例如根据 ICD-O-3,胃癌(C16)可分为贲门(C16.0)、胃底(C16.1)、胃体(C16.2)、胃窦(C16.3)、幽门(C16.4)、胃小弯(C16.5)、胃大弯(C16.6)、交搭跨越(C16.8)及胃的其它(C16.9)等亚部位癌。而“肠癌”应属部位不明,它必须编码归入小肠(C17)、结肠(C18)、直肠乙状结肠连接(C19)、直肠(C20)、肛门和肛管(C21)等部位的癌症。肠的这些部位再进一步细分为亚部

位,例如结肠再分为盲肠(C18.0)、阑尾(C18.1)、升结肠(C18.2)、结肠肝曲(C18.3)、横结肠(C18.4)、结肠脾曲(C18.5)、降结肠(C18.6)、乙状结肠(C18.7)、结肠交搭跨越(C18.8)以及结肠的其它(C18.9)等亚部位的癌症。肝和肝内胆管癌(C22),可分肝(C22.0)和肝内胆管(C22.1)癌。肺癌(C34)分主气管(C34.0)、肺上叶(C34.1)、肺中叶(C34.2)、肺下叶(C34.3)、肺交搭跨越(C34.8)及肺的其它(C34.9)等亚部位癌。

对于“白血病”,在 ICD-10 及以前的版本中,不同形态学的白血病类型是通过部位分类来区别的。例如在 ICD-10 中,分为淋巴样白血病(C91)、髓样白血病(C92)、单核细胞白血病(C93)、特指细胞类型的其它白血病(C94)、未特指细胞类型的白血病(C95)等,再根据细胞的形态学,归入相应的“亚部位”(“-”)中。但在 ICD-O-3 版本中,所有的白血病都归入 C42.1,而名目繁多的白血病的编码,是通过形态学编码(M)来实现的。例如,淋巴样白血病编码为 C42.1M9820,慢性髓样白血病编码为 C42.1M9863,单核细胞白血病编码为 C42.1M9860 等。其它还有描述白血病的术语,例如急性、慢性、亚急性、中性、B 细胞、T 细胞等,都可查到相应的 M 编码,例如,前体 T 细胞急性淋巴细胞白血病,编码为 C42.1M9837。

3.3.2 常见的部位(亚部位)编码错误的原因

(1) 资料收集时信息的缺少,例如只报告“肠癌”、“白血病”等。

(2) 资料摘录时漏摘重要信息,例如“肺中叶小细胞癌”,而在病史摘录时、或者报告时,只报告“肺癌”。

(3) 资料编码时业务不熟悉,例如,把胃恶性淋巴瘤,根据 ICD-10 而编码为 C85.9M9590。而在 ICD-O 中应当归入部位“胃”(C16.9),因此胃恶性淋巴瘤的 ICD-O-3 编码应为:C16.9M9590。

对于发生在源头上的错误,应当通过登记报告系统的培训、督查、核对和随访来解决。通过提高摘录能力和加强责任心,可以杜绝或减少对于资料摘录时易出现的错误。对于编码中的特殊点和难点,可以通过反复培训、练习和复核来减少错误出现的机会。

3.4 形态学(组织学)

形态学编码是 ICD-O-3 编码区别 ICD-10 编码的重要部分。因为在 ICD-O-3 中,无论是良性肿瘤、还是恶性肿瘤,其肿瘤所在部位的编码,即 C 编码都相同,主要依据形态学编码及行为学编码来区别。形态学编码是一组 4 位数的数字编码表,范围在 8000~9989。但这组编码还常与行为学编码连用,作为形态学编码的进一步阐明。

例如,8042/3 燕麦细胞癌 (C34._),斜杠后的“3”,表明其行为学为“恶性”,“C34.”表示其只出现在肺部(部位推荐码)。再如,9950/3 真性红细胞增多症(C42.1),斜杠后的“3”,表明该病的行为学为“恶性”,“C42.1”表示其部位归在“骨髓”。

3.5 行为(行为学)

也有称为“动态”。“行为”是描述肿瘤性质的,是指“/”后面的编码,/0 为良性,/1 为交界恶性,/2 为原位癌,/3 为恶性,原发部位,/6 为恶性,继发部位,/9 为恶性,原发或继发未肯定。癌症登记处通常记录“恶性”(/3)的肿瘤,因此,行为编码一般为“/3”。其容易出现的问题来自不太可能出现的部分和行为学,如表 1。此外,在 ICD-O-3 文献^[3]第 78~148 页中未列出的形态学编码,程序都视作不太可能出现的编码。例如,8000/2。在 Check 软件中,这些不太可能的,组合将被写入 .chk 警示文件。

表 1 不太可能出现的部位和行为学

部位编码	行为学编码
C40._	2
C41._	2
C42._	2
C47._	2
C49._	2
C70._	2
C71._	2
C72._	2

3.6 分化程度(分级)

描述肿瘤的分化程度,即病理学上的“分级”(grade)。病理分级实际上根据是否为“实体”肿瘤和“造血和内皮系统”肿瘤而有区别。

对实体肿瘤分化程度的分级见表 2,对造血系和网状内皮系统肿瘤的分化程度的分级见表 3。

分级(分化程度)中的常见问题是造血和内皮系

表 2 实体肿瘤的分级分化编码

分级分化	编码	英文
Grade I 高(好)	1	Well differentiated Differentiated, NOS
Grade II 中	2	Moderately differentiated Moderately well differentiated Intermediate differentiation
Grade III 低(差)	3	Poorly differentiated
Grade IV 未间变	4	Undifferentiated Anaplastic
不详	9	Grade or differentiation not determined, not stated or not applicable

表 3 造血和内皮系统肿瘤的分级分化编码

细胞类型	分级编码	英文
T 细胞	5	T-cell
B 细胞	6	B-cell
前 B		Pre-B
B 前体		B-precursor
无标记细胞	7	Null-cell
非 T-非 B		Non T-non B
NK 细胞	8	NK cell
自然杀伤细胞		Natural killer cell
细胞类型不详	9	Cell type not determined, not stated or not applicable

统肿瘤的分级编码填写成 1、2、3 或 4;或者,实体肿瘤的分级编码填写成 5、6、7 或 8。在 Check 软件中,有关分化程度的不太可能的组合将被写入 .chk 警示文件。

3.7 诊断依据

这里所指的“诊断依据”,是指可能与“形态学”描述不相匹配的肿瘤诊断描述的情况,而不是具体指实验室检查、B 超、CT、手术等“诊断”依据。事实上,如果把诊断依据中例如“B 超”诊断的编码写成“生化”或“手术”等诊断的编码,Check 软件是不会作出审核的。这是因为,ICD-O-3 形态学编码并不用于区别诊断依据的目的。不过,CHECK 软件“知道”,某些特定的形态学诊断,若没有组织学检查就不大可能得到诊断。例如,前面提到的 8042/3 燕麦细胞癌(C34._),若没有提及组织学诊断,即没有提到“镜检确认”的诊断依据(例如编码“7”),而给出的形态学编码为“8042”,CHECK 软件能识别这样的错误。

但是,对于某些疾病的描述,即使没有提到“镜检确认”,但 CHECK 软件也不认为“有错”。例如,眼部恶性黑色素瘤(malignant melanoma of the eye),则

表 4 可被接受的非镜检确认的形态学编码

肿瘤名称	英文	形态学编码
肿瘤,未特指	Neoplasm, NOS	8000
胰岛细胞瘤	Islet cell tumours,	8150-8154
胃泌素瘤	Gastrinomas	
肝细胞癌	Hepatocarcinoma	8170
垂体肿瘤	Pituitary tumours	8270-8281
眼部黑色素瘤	Melanoma of the eye	8720 且部位为 C69._
皮肤黑色素瘤	Melanoma of skin	8720 且部位为 C44._
肉瘤,未特指	Sarcoma, NOS	8800
肾母细胞瘤,未特指	Nephroblastoma, NOS	8960
绒毛膜上皮癌,未特指	Choriocarcinoma, NOS	9100
卡波西肉瘤	Kaposi sarcoma	9140
颅咽管瘤	Craniopharyngioma	9350
神经胶质瘤	Glioma	9380
室管膜下巨细胞星形细胞瘤	Subependymal giant cell astrocytoma	9384/1
成神经细胞瘤	Neuroblastoma, NOS	9500
成视网膜细胞瘤	Retinoblastoma, NOS	9510
脑膜瘤	Meningioma	9530-9539
淋巴瘤,未特指	Lymphoma, NOS	9590
多发性骨髓瘤	Multiple myeloma	9732
瓦尔登斯特伦巨球蛋白血症	Waldenstrom macroglobulinemia	9761
白血病,未特指	Leukaemia, NOS	9800

即使在记录中没有提及诊断依据编码(7),CHECK软件也可通过审核,而不认为有错。因此,诊断依据/组织学质控用来检查特定的形态学诊断和非形态学诊断,某些组合被排除在这些通用规则之外^[2,3]。

非镜检确认的诊断(即一个并未列在“镜检确认诊断依据编码”字段中的诊断依据编码的记录)只接受下列的组织学编码(表4)。

除表4以外的肿瘤,若无镜检确认的诊断依据,则通常形态学诊断编码视为不太可能。在Check软

件中,这些不太可能的组合将被写入.chk警示文件。

参考文献:

- [1] 赵平,陈万青,孔灵芝. 中国癌症发病与死亡 2003-2007 [M]. 北京: 军事医学科学出版社, 2012.1-413.
- [2] IACR. IARCrgTools, Version 2.05. <http://www.iacr.com.fr> [CP/OL]. 2012-3-15.
- [3] 董景五编译. 国际疾病分类肿瘤学专辑第3版(ICD-O-3) [M]. 北京: 人民卫生出版社, 2003.1-467.