

根本死因自动编码工具的应用研究

姬一兵 王黎君 周脉耕

【导读】 为提高常规死因监测中根本死因编码的准确性,按照 WHO 的 ICD-10 根本死因编码规则及国际上同类软件开发设计原理,并参照美国编码策略表,通过计算机语言实现根本死因编码的自动化,研制基于全国死因登记报告信息系统的根本死因自动编码工具,其根本死因编码正确率达到 85%,实现了对死亡个案数据的自动编码,且编码正确率达到国际同类软件水平。

【关键词】 根本死因;自动编码;编码正确率

Exploratory development of automated coding software on the underlying causes of death Ji Yi-bing, WANG Li-jun, ZHOU Mai-geng. Center for Public Health Surveillance and Information Service, Chinese Center for Disease Control and Prevention, Beijing 102206, China
Corresponding author: WANG Li-jun, Email: wangli_19@yahoo.com.cn

【Introduction】 To develop an automated coding software related to the underlying causes of death, based on the National Registration Information System on deaths, which could improve the quality of coding on the underlying causes of death in the conventional death surveillance system. Following the coding rules of the underlying death cause of ICD-10 and the design on principles of software of underlying death cause automated coding from some other countries, as well as in accordance with the coding strategy table from the USA, we implemented the automatization of the underlying death cause coding. Based on national registration information system on cause of death, an automated coding software of underlying death cause was developed with the coding correction rate closed to 85%. The automated coding software of underlying death cause could code the death cases of underlying death cause with high rate of correction, similar to that of the same kind softwares developed in other countries.

【Key words】 Underlying death cause; Automated coding; Coding correct rate

居民死亡水平反映一个国家或者地区的卫生保健水平和居民健康状况,同时死亡资料对研究疾病负担、期望寿命和疾病控制优先权都是不可或缺的基础信息^[1,2]。获取准确的死因对分析其顺位、构成及得出相对准确的死亡谱至关重要,并为卫生行政部门制定相关防病政策提供死因方面的依据。在对居民死亡信息的分析中,最重要的信息是死因链填写的准确性,而根本死亡原因编码的准确与否直接影响到死因数据分析结果的正确性和居民真实的死亡水平,进而影响到卫生行政部门制定的卫生政策和卫生资源的合理配置。但目前我国死因登记工作中对根本死因确定和编码存在较大问题,很多报告的个案未能正确判定根本死因及编码^[3]。为此本研究根据疾病和有关健康问题的国际统计分类第十次修订版(ICD-10)编码规则,并借鉴美国编码策略

表,基于全国死因登记报告信息系统,研制了根本死因自动编码工具,依据死因链实现根本死因的自动编码功能,提高个案信息编码的准确性,同时减轻人工编码的繁重工作,克服了短时间内通过常规编码培训无法提高基层专业人员编码水平的问题。

1. 研究方法:

(1)设计思路:根本死因自动编码工具是基于 ICD-10 判断根本死因的规则,包括总原则、选择规则、修饰规则和注释^[4],参考使用美国的编码策略表,推断各种疾病之间的病因关系和修饰关系^[5],通过计算机语言实现根本死因判断的自动化。

(2)《需求规格说明书》的制定:按照系统工程需求分析的方法,基于全国死因登记报告信息系统各级用户,并结合我国目前死因报告管理工作中的实际情况,通过查阅文献及研究国外同类软件的设计原理和开发流程,根据 ICD-10 根本死因判定的编码规则,与编码专家、计算机专业人员和各级死因业务系统用户进行充分研讨,提出了根本死因自动编码工具《需求规格说明书》。其中参照美国最新的编码

策略表,对疾病之间的病因关系和修饰关系进行判断,并以该说明书为标准实施质量控制,同时采用单纯随机抽样方法抽取 2011 年死因登记报告信息系统收集的全国死亡个案卡进行编码准确性验证。

(3)实施过程:分为三个阶段。包括编码工具需求分析、编码工具开发(委托中科软件科技股份有限公司)以及正式上线推广使用(图 1)。2011 年 8 月完成编码工具 1.0 版。测试结果表明自动编码软件对死亡个案根本死因判定的准确性为 85%,达到国外同类软件的编码正确率水平。

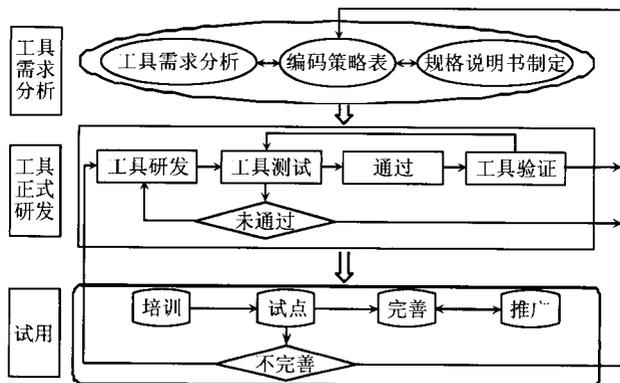


图 1 根本死因自动编码工具设计流程

(4)编码工具测试:从全国死因登记报告系统中选择北京市 2011 年 1—3 月的死因报告个案,共计 21 550 例。使用 SAS 软件从此期间病例中筛选出编码链为一码的个案为 8451 例,二码为 7715 例,三码为 4049 例,四码为 1128 例,五码为 106 例;从中随机抽取一至四码的个案各 600 例,五码个案 100 例,由 4 名编码专家对个案进行根本死因编码,并以此为金标准。每名专家负责 625 例个案的编码,其中一至四码个案各 150 例,五码个案 25 例。使用自动编码软件对专家编码后的个案重新编码,比对软件自动编码与金标准的一致性,以获得自动编码的正确率。

(5)根本死因推断流程:将居民死亡医学证明书的死因链上各死因按图 2 依次处理后,推断出根本死亡原因。首先是将死亡个案死因链上所列疾病编码数据保留整数前 3 位,即为疾病的类目,然后编码通过有效编码表(VCODES)进行过滤,如存在有效编码表中则保留该编码用于下一步判断,反之则删除(该疾病不能用做根本死因);再将可作为根本死因的编码置于病因关系表(TABA)中进行疾病间的病因关系判断,如死因链上各疾病在 TABA 中存在对应关系,则可判断该两行疾病间存在病因关系,将

下位疾病或损伤作为初步根本死因。而初步根本死因还需经过修饰关系表(TABB)检验判定其是否为最终根本死因。

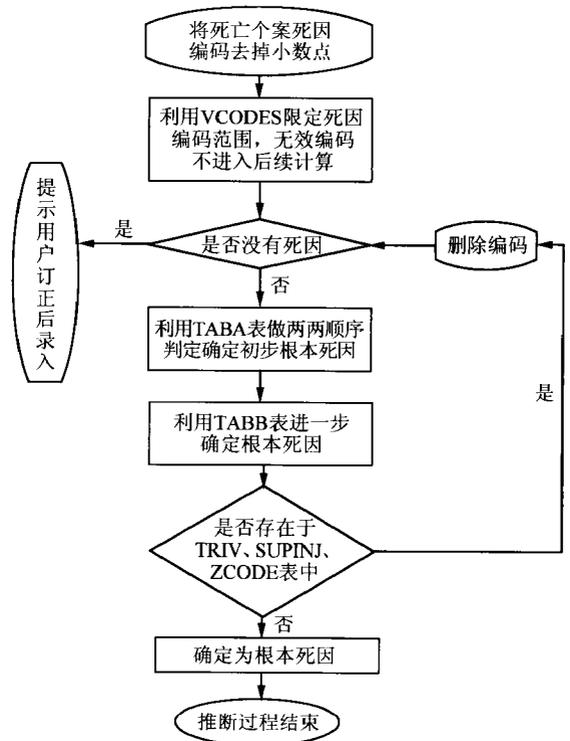


图 2 根本死因推断流程图

(6)应用试点的选择:以北京市石景山和昌平区作为试点,并选择 2011 年 11—12 月死因登记系统报告的死亡个案,应用根本死因自动编码工具此两区 11—12 月网报个案进行编码。通过基层专业人员的应用以发现该软件在实际使用中存在的问题,并比对 2 个区编码专业人员手工编码与根本死因自动编码工具编码的一致性,以评价软件实用性。

2. 结果:

(1)根本死因自动编码工具:依据 WHO 对根本死亡原因判断的基本原则和修饰规则,参考美国编码策略表,结合我国死因监测实际工作,形成根本死因自动编码的业务流程和开发框架,并与软件公司合作研制出根本死因自动编码工具。该软件能够对全国死因登记报告信息系统中上报死亡个案的死因链上各直接死因间的病因关系做出推断,并最终判断出造成患者死亡最特异、最严重的根本死因。其中包括单条个案处理、多条个案的批处理、字典表查看和维护及帮助 4 个功能选项(图 3)。

对于单条个案根本死因的推断,可将死亡医学证明书上死因链列出的各死因编码依次录入 A、B、C、D 后的空白处,死因链上第二部分疾病或损伤录

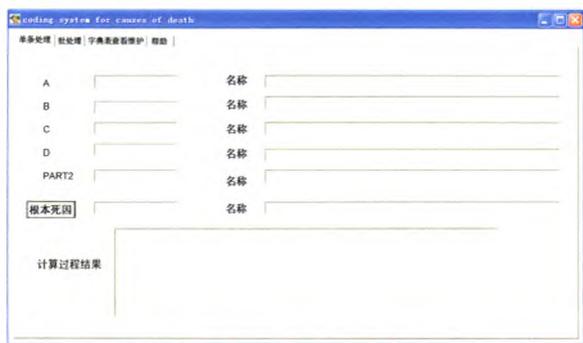


图3 根本死因自动编码工具相关界面

入软件界面的PART2后空白处,然后点击根本死因按钮,则系统会自动对各疾病之间的病因关系进行判断,并推断出患者的根本死亡原因,相应编码则生成在根本死因按钮后的空白处。同时该系统还支持批处理过程,可将全国死因登记报告信息系统内的死亡数据导出成CSV或者XLS格式的文件,然后导入到系统中,同时进行多条死亡个案的根本死因的推断,在相应的CSV或者XLS格式的数据库中生成相应的根本死亡原因列。

(2)自动编码工具正确性测试:选取北京市2011年1—3月通过全国死因登记报告信息系统报告的死亡个案21 550条,随机抽取死因链上一至四码的个案各600例、五码的个案100例,由编码专家对抽取的2500例个案进行根本死因编码,并以专家编码为金标准,再使用自动编码工具对2500条记录进行根本死因编码,比对编码工具所编的根本死因与专家编码结果,表明自动编码工具所编根本死因的总正确率为84.76%(表1)。

表1 根本死因自动编码工具正确性测试结果

编码位数	例数	比例 (%)	累计百分比 (%)	测试例数	编码正确率 (%)	累积编码正确率 (%)
一码	8 451	39.22	39.22	600	95.83	95.83
二码	7 715	35.80	75.02	600	85.17	90.50
三码	4 049	18.79	93.81	600	83.67	88.22
四码	1 128	5.23	99.04	600	76.67	85.33
五码	106	0.49	99.95	100	71.00	84.76
无效编码	11	0.052	100.00	-	-	-
合计	21 550	-	-	2500	84.76	-

(3)自动编码工具试点应用:根本死因自动编码工具在北京市昌平和石景山区试点应用结果表明,死因链上填写的病因少则编码正确率高,反之亦然(表2)。说明死因链的复杂性是影响根本死因自动编码工具正确性的重要因素。因此尽可能全面地对疾病间的病因关系和修饰关系建立关联,进而完善

编码策略表,是解决提高自动编码工具编码正确性的关键。

表2 自动编码工具试点应用结果

地区	编码位数	例数	一致性(%)	累积一致性(%)
昌平	1	301	89.70	89.70
昌平	2	201	80.60	86.06
昌平	3	62	80.65	85.46
昌平	4	49	79.59	84.99
昌平	5	2	50.00	84.88
石景山	1	99	98.99	98.99
石景山	2	124	82.26	89.69
石景山	3	107	75.70	85.15
石景山	4	51	76.47	83.99
石景山	5	8	75.00	83.80
石景山	-	2	-	-
合计	-	1006	84.46	-

3. 讨论:目前我国死亡数据的上报工作主要依托于中国疾病预防控制中心的全国死因登记报告信息系统,形成国家-省-市-县四级网络直报平台,由各级医疗机构将死亡个案通过死因系统直接报至国家数据库,再由中国疾病预防控制中心对全国死亡数据进行统计分析,从死因构成、死因顺位、死亡个案的时间和空间分布等方面对我国居民的死亡情况做出评价。因此所有分析均基于死亡个案记录中死因链的填写,其准确性直接影响到分析结果的正确性,从而影响到卫生决策部门对我国居民主要健康危害的判断及卫生资源的合理和优化配置。

根本死因自动编码工具实现了对单条记录的编码和对多条记录的批处理编码功能,编码正确率达到85%,已达到国外同类软件的编码正确率水平^[6,7]。1967年美国生命统计中心已开发死因数据的自动编码系统(Mortality Medical Data System),后经自动化死因统计国际合作联盟将死因自动编码系统推广至其他欧美国家,法国、瑞典等国相继在美国的死因自动编码系统基础上结合本国实际情况开发出了各自国家的死因自动编码系统^[8]。本次研究的根本死因自动编码工具也是借鉴美国和其他国家的自动编码软件,参照美国的编码策略表及软件相关功能,对我国死因监测数据进行自动编码。相对国外同类编码系统,本软件基于我国死因登记报告信息系统所收集的死亡数据,应用ICD-10的根本死因编码规则,侧重于对死因链列出的疾病或者损伤中毒编码之间逻辑关系的判断分析,提供单条记录与多条记录的批处理功能。同时在开发过程中兼顾与全国死因登记信息系统的兼容性,以后还将自动编码工具嵌入到死因登记信息系统中,提供专业人

员在线完成根本死因自动编码的功能。

为检测编码工具的实用性,在 2011 年 11—12 月在北京昌平区石景山区进行试点应用,由疾病控制专业编码人员使用根本死因自动编码工具对辖区内报告的死亡个案编码,结果应用编码工具编码与专业人员编码的一致性达到 85%。其结果也表明一致性水平与死亡医学证明书上死因链填写的完善程度有关,说明死因链的复杂性是影响根本死因自动编码工具正确性的重要因素,即居民死亡医学证明书的死因链上填写的导致死亡疾病或者损伤的多少决定了对根本死因判断的复杂程度,造成患者死亡的直接原因越多,就越容易掩盖造成死亡的根本原因,从而增加了推断患者根本死亡原因的难度^[9]。

工具研发中还存在两方面不足。一是编码策略表需要进一步完善,制定适合我国人群的编码策略表;二是软件运算速度,随着个案记录的增加,运行时间较长,还有待优化。

参 考 文 献

- [1] Chen Z. The report of the third retrospective sampling survey of death cause. Beijing: Peking Union Medical College Press, 2008. (in Chinese)
陈竺. 全国第三次死因回顾抽样调查报告. 北京: 中国协和医科大学出版社, 2008.
- [2] Ezzati M, Lopez AD, Rodgers A, et al. Comparative quantification of health risks: global and regional burden of disease attributable to selected major risk factors. Geneva: World Health Organization, 2004.
- [3] Zhou MG, Wang YY, Ge H, et al. Study on the Quality of death-case-reporting-system in county and above levels' medical institutions in 2004. Chin J Epidemiol, 2006, 27(4): 328-332. (in Chinese)
周脉耕, 王玉英, 葛辉, 等. 中国 2004 年县及县以上医疗机构死亡病例报告质量评价. 中华流行病学杂志, 2006, 27(4): 328-332.
- [4] International Disease Classification Family Cooperation Center of World Health Organization at Peking Union Hospital. The Tenth Revision of International Statistical Classification of Diseases and Related Health Problems(疾病和有关健康问题的国际统计分类第十次修订本). 董景五, 主译. 2 版. 北京: 人民卫生出版社, 2008.
- [5] National Center for Health Statistics. Mortality Medical Data System (MMDS), Instruction Manual Part 2b -Instructions for Classifying Multiple Causes of Death. 2007.
- [6] McKenzie K, Walker S, Tong S. Assessment of the impact of the change from manual to automated coding on mortality statistics in Australia. HIM J, 2002, 30(3): 1-11.
- [7] Wall MM, Huang J, Oswald J, et al. Factors associated with reporting multiple cause of death. BMC Med Res Methodol, 2005, 5(1): 4.
- [8] http://www.cdc.gov/nchs/nvss/mmds/about_mmds.htm.
- [9] Redelings MD, Sorvillo F, Simon P. A comparison of underlying cause and multiple causes of death: US vital statistics, 2000-2001. Epidemiology, 2006, 17(1): 100-103.

(收稿日期: 2012-09-20)

(本文编辑: 张林东)

读者·作者·编者

关于中华医学会系列杂志投稿网址的声明

为维护广大读者和作者的权益以及中华医学会系列杂志的声誉,防止非法网站假冒我方网站诱导作者投稿、并通过骗取相关费用非法获利,现将中华医学系列杂志稿件管理系统网址公布如下,请广大作者加以甄别。

1. “稿件远程管理系统”网址: 中华医学会网站(<http://www.cma.org.cn>)首页的“业务中心”栏目、中华医学会杂志社网站(<http://www.medline.org.cn>)首页的“稿件远程管理系统”以及各中华医学会系列杂志官方网站接受投稿。作者可随时查阅到稿件处理情况。

2. 编辑部信息获取: 登录中华医学会杂志社网站(<http://www.medline.org.cn>)首页,在《中华医学会系列杂志一览表》中可查阅系列杂志名称、编辑部地址、联系电话等信息。

3. 费用支付: 中华医学会系列杂志视杂志具体情况,按照有关规定,酌情收取稿件处理费和版面费。稿件处理费作者在投稿时支付;版面费为该稿件通过专家审稿并决定刊用后才收取。

欢迎投稿,并与编辑部联系。特此声明。