

死因数据中垃圾编码再分配方法研究进展

刘利群 万霞

中国医学科学院基础医学研究所/北京协和医学院基础学院,北京 100005

通信作者:万霞,Email:xiawan@ibms.pumc.edu.cn

【摘要】 垃圾编码存在于死因监测数据库中,会降低死因统计的准确性,进而影响卫生决策的准确性和有效性。国内外科研工作者们对于不同国家或地区死因数据库中垃圾编码的特点进行了很多研究,提出了专家咨询、按比例分配、利用死因链信息、建立统计方程等多种再分配垃圾编码的途径。本文对于目前常见方法的原理、应用以及应用场景的局限性进行了梳理和比较,从而为提高中国死因数据的准确性和有效性提供借鉴思路。

【关键词】 垃圾编码; 再分配; 死因数据; 综述

基金项目:中国医学科学院医学与健康科技创新工程(2016-I2M-3-001);中华医学基金会合作项目(15-208)

Progress in research on redistribution methods for garbage codes in causes of death data

Liu Liqun, Wan Xia

Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences/School of Basic Medicine, Peking Union Medical College, Beijing 100005, China

Corresponding author: Wan Xia, Email: xiawan@ibms.pumc.edu.cn

【Abstract】 The existence of garbage codes in death cause surveillance data sets could influence the accuracy of the death cause statistics, and subsequently affect the precision and effectiveness of public health policy making. International and domestic researchers have studied the characteristics of garbage codes in various death cause data sets from different countries or regions in the world. They proposed several approaches for redistributing garbage codes, such as expert consultancy, fixed proportional reassignment, using the information about death cause chain, building statistical models, and so on. This paper summarizes and compares the principles, applications and limitation of application scenarios of currently common methods for garbage code redistribution in order to provide some references for improving the accuracy and usefulness of the death cause data in China.

【Key words】 Garbage code; Redistribution; Death data; Review

Fund programs: Chinese Academy of Medical Sciences Innovation Fund for Medical Sciences (2016-I2M-3-001); China Medical Board Collaboration Project (15-208)

一、垃圾编码

死因别死亡专率是反映人群健康水平最基本、最重要的指标之一,也是卫生决策者做出决策的重要依据指标,因此在全世界多个国家和地区被广泛应用。其来源主要依靠死因监测等数据库中所记录的死亡报告卡所填写的根本死因。WHO对根本死因下了一个定义:“(a)引起直接导致死

亡的一系列病态事件的那些疾病或损伤,或者(b)造成致命损伤的事故或暴力的情况”,即根本死因不包括症状、体征和临床死亡方式,如“心力衰竭”“呼吸衰竭”等^[1]。在死亡报告卡根本死因填写的过程中,主要是根据死者生前死因链进行填写。但是,由于部分死因上报人员的水平有限,对根本死因的理解有误,往往会填写出一些不能作为根本死

DOI:10.3760/cma.j.cn112338-20211025-00818

收稿日期 2021-10-25 本文编辑 万玉立

引用格式:刘利群,万霞.死因数据中垃圾编码再分配方法研究进展[J].中华流行病学杂志,2022,43(5):784-788.

DOI:10.3760/cma.j.cn112338-20211025-00818.

Liu LQ, Wan X. Progress in research on redistribution methods for garbage codes in causes of death data[J]. Chin J Epidemiol, 2022, 43(5):784-788. DOI: 10.3760/cma.j.cn112338-20211025-00818.



因的疾病,或者是死因链中的中间疾病或者症状,例如:“心衰”“败血病”“肺栓塞”“肺源性心脏病”等。这类疾病所对应的编码,被称为垃圾编码。

垃圾编码最早是由 Murray 和 Lopez^[2]于 1996 年首次提出。具体来说,就是指死因数据库中,有些记录的根本死因被填写为对于公共卫生统计不具有意义的编码^[2]。垃圾编码又可细分为 4 类^[3]:①在医疗机构中有意义但不应作为根本死因的,包括症状、体征、危险因素、后遗症等;②实际应为死因链上的中间原因的;③实际应为直接死因的;④一个疾病大类中不明确的。

垃圾编码的存在影响死因统计的准确性,从而影响国家卫生部门对于人群疾病发展水平和趋势的判断,以及对于疾病与危险因素间相关性的认知。不准确的判断降低了疾病防控策略的正确和有效性,可能给维持和提升人群的健康水平带来不利影响。随着全球疾病负担(Global Burden of Disease, GBD)研究取得进展,垃圾编码的范围有所扩充,划分也越来越细致,目前达到在国际疾病分类第 10 次修订本(ICD-10)编码小数点后一位的水平进行划分^[3,4]。为了增加死因数据对于公共卫生决策的支持力度,科学家们对于不同国家或地区死因统计数据中垃圾编码的特点及其再分配方法的研究也从未停止。为了更好地服务于公共卫生策略制定,保障人民健康,应了解国际上针对垃圾编码处理的最新研究进展,结合我国死因监测数据实际情况,思考如何借鉴他人的研究成果,加强我国死因分析的准确性和有效性。

二、垃圾编码再分配方法

每种垃圾编码再分配的方法中都包含两方面内容:再分配目标编码的范围,和某垃圾编码分配到每个目标编码的比例。某些方法通过一个分析过程同步获得目标编码范围和分配比例。由于死因模式受到年代、地点(不同国家或不同地区)、性别和年龄等的影响,因此,再分配过程一般在根据以上变量划分的亚组内分别进行。现具体阐述目前国内垃圾编码再分配的方法。

1. 专家咨询法和专家经验算法:专家咨询法自 GBD 研究之初就开始应用^[5],指回顾发表的文献,并同 GBD 研究专家组进行磋商后,确定某垃圾编码再分配的范围和比例。直至 GBD2010 研究期间,这仍是很多垃圾编码再分配时使用的方法^[3]。与最初相比发展之处在于,研究者会建立一套基于专家经验的算法,并且考虑目标死因的基本性质,注意其长期变化趋势的合理性和连续性,特别是在 ICD 版本更替的前后^[3]。

相对来说,专家咨询法操作简单,但是受专家的主观影响较大,不同专家可能得出不同的结果,因此,可能存在结果不稳定的现象。专家经验算法相对来说,比简单专家咨询法更优,但是,由于建模的需要,要求有一定的数据量。

2. 给定分配比例:在数据库中除了根本死因,其他信息内容很少的情况下,直接按照所有目标编码的构成比来分配垃圾编码^[3]。这种方法操作也相对比较简单,其思路类

似于缺失数据单值插补的方法,也是目前国内垃圾编码再分配常用的方法。有时,在给定分配比例的时候,也会结合数据库本身之外的其他信息。例如:HIV 感染和其他一些相对“常规”的根本死因如“多类真菌感染”“其他类型的免疫缺陷疾病”等,在死因监测库中可能被误写为同样的一组垃圾编码,包括“卡波西肉瘤”“免疫缺损”“霉菌感染”等(参考 GBD2017 研究论文附件 1 中的表 6,可见这一组垃圾编码的详细列表)^[4]。换句话说即这些垃圾编码可以被再分配到 HIV 感染或其他所谓“常规”疾病。考虑到艾滋病的死亡率在不同的年代、地点及人群中可能迥然不同,GBD 研究人员先以地区为单位,按每 5 年为一个阶段,计算这组垃圾编码的分年龄别分性别死亡率,并与 1980–1984 年(HIV 感染被发现的初期)的分年龄别分性别死亡率相比较。根据经验,如果某一个 5 年阶段内,这组垃圾编码在一个地区-性别-年龄别亚组内的死亡率与该亚组 1980–1984 年死亡率相比较,上升超过 5%,则将 5% 之上的部分再分配为 HIV 感染,而将 5% 及以下的部分,再分配到“常规”疾病,分配比例就按照数据库中“常规”疾病本来的构成比来执行^[4]。这种方法的分配比例基于同某一个年代点值相比较的结果,因此,具有一定的局限性。

3. 建立线性方程:基于前述两种方法的局限性,有学者提出了模型法,即在垃圾编码和目标编码间建立简单的线性方程。下面以“未特指的心血管疾病”及“心衰”这两个垃圾编码为例,介绍两种建立线性方程的原理。

(1) Lozano 等^[6]再分配“未特指的心血管疾病”:2001 年,Lozano 等^[6]在分析多个国家的死因数据库后发现,死因数据中“未特指的心血管疾病”编码的占比与“缺血性心脏病”(ischemic heart disease, IHD)编码的占比间存在负相关。即“未特指的心血管疾病”(是垃圾编码)中有一部分应该编码到 IHD。因此,Lozano 等^[6]采用一组与 IHD 死亡密切相关的变量(X_i)及可获取到的相应数据,通过最小二乘法拟合方程,从而估算真实的 IHD 的死亡率(I_T)。死因库中观察到的 IHD 死亡率(I_0)与 I_T 的差值,则为被错误标记成“未特指的心血管疾病”而实际上应为 IHD 的这部分死亡率(I_M)。而后,可以进一步计算应从死因库中观察到的“未特指的心血管疾病”死亡率(G_0)中校正过来的比例,即计算 I_M 在 G_0 中的占比(见公式 1, I_M/G_0 为需要估算的参数)。

$$I_0 = f(X_i \dots) - (I_M/G_0) \cdot G_0 \quad (1)$$

在应用时,Lozano 等^[6]根据死因数据库的实际情况,将一部分国家划分为“未特指的心血管疾病”占比低的国家,其余为“未特指的心血管疾病”占比高的国家,并用哑变量 $D_{high} = 0$ 或 1 来标记。他们将前者的死因编码工作定义为“标准流程”,对应其“未特指的心血管疾病”占比的平均水平为“标准百分比”。在这样的划分后,公式 1 进一步变换为:

$$I_0 = \beta_{li}(X_i \dots) + \beta_2 \cdot G_0 + \beta_3 \cdot G_0 \cdot D_{high} + \varepsilon \quad (2)$$

其中 β_{li} 指代 $f(X_i \dots)$ 中每个 X_i 所采用的回归系数, β_2 和 β_3 替

代了公式 1 中的“ $-(I_M/G_0)$ ”,其估计值都应在 $-1\sim 0$ 。 β_2 和 β_3 分别表示“标准百分比”之内和之上的“未特指的心血管疾病”应重编为IHD的比例。

(2)Ahern等^[7]再分配“心衰”:Ahern等^[7]也介绍了一种利用负相关关系和线性回归模型寻找目标编码范围和分配比例的方法,使用“心衰”作为示例垃圾编码。基于“心衰”的病理生理学,他们列出了它可能被再分配到的12个目标死因组(12 target groups,共包含49种疾病)。从死因库中提取出根本死因为“心衰”或目标疾病的记录,并按照性别和年龄别分成亚组,在每个亚组内分别建立回归模型:以每个地区每一年为最小单位,计算各“地区-年份”内“心衰”及其目标死因组的百分比(构成比),13个百分比之和为100%;所有“地区-年份”累积在一起,便可得到13列数据,用于下列的回归分析:

$$\%TG = \alpha + \beta[\%HF] + \varepsilon \quad (3)$$

式中, $\%HF$ 为“心衰”的构成比,即某地区-年份内根本死因为“心衰”的记录数/该地区-年份内所有记录数; $\%TG$ 为12个目标死因组的构成比,即“某地区-年份内根本死因为某一个目标死因组中疾病的记录数/该地区-年份内所有记录数”。可见,公式3实际上对应12个具体的方程。

由于垃圾编码与目标死因之间存在负相关,因此,在第一轮估算后,Ahern等^[7]找到 β 的估计值为正数且具有统计学意义($P < 0.05$)的回归方程,剔除了其对应的目标死因组,重新计算各“地区-年份”内“心衰”及剩余目标死因组的百分比(构成比之和仍为100%),而后进入第二轮估算。以此类推,直到没有方程的 β 估计值为正数且具有统计学意义为止。找到最终 β 的估计值为负数且具有统计学意义的回归方程,其对应的目标死因组(其中的疾病)即是“心衰”再分配的目标编码。先根据这些方程的 α 值,将“心衰”分配到每个目标死因组,再根据组内所有目标编码的构成比,将“心衰”最终分配到目标编码。

自GBD2013研究开始,“心衰”“高血压”“动脉粥样硬化”等一系列垃圾编码的再分配都利用了这一方法^[8]。方程的形式也在渐渐改进^[4,8-10]。

4. 利用死因链信息:在推断根本死因时,主要是依据死者生前的死因链进行推断。因此,死因链在垃圾编码再分配过程中也具有重要的利用价值。与前文中介绍的方法相比,利用死因链信息的垃圾编码再分配方法不预先设定目标编码范围,而是基于死因库的真实情况来“发现”所有目标编码,并同时获得其分配比例。

(1)GBD研究中再分配“意外中毒”:GBD研究人员发现,根本死因被记录为“意外中毒”的,其中有一部分人的真实死因可能是成瘾物质或药物的过量使用。他们在死因链中寻找药品/毒品名称;当发现某死者同时使用多种药品/毒品时,按照它们的致死能力(详见GBD2017研究论文附件1,小节2.7,表A)选择最优势的一个^[4]。

(2)粗化精确匹配法再分配“心衰”:将粗化精确匹配法

应用于垃圾编码再分配的思路是来源于前述的“给定分配比例”,但结合了死因链的信息。其中“精确”体现在对于死因链和其他疾病诊断的利用方面,“粗化”体现在划分亚组方面。下面以Stevens等^[11]再分配“心衰”的过程作为示例进行介绍。

Stevens等^[11]从死因库中找出根本死因为“心衰”的记录,先观察每条记录的死因链和其他疾病诊断。如果从这两处能找到可取的根本死因的,直接修改“心衰”为该死因;如此处理好一部分,剩余的则是等待再分配的记录。再从死因库中找出死因链里包含“心衰”,但根本死因不是“心衰”,既不是其他垃圾编码,也不属于伤害的记录,这些是分配的对照记录。利用死因记录中的其他一些变量作为分亚组变量,例如性别、年龄、文化水平、死亡地点、生活地区等,将待分配记录和对照记录按亚组对应起来。为了使尽可能多的待分配记录得到再分配,每个亚组内应尽量包含至少一条对照记录。为做到这点,在合理范围内,亚组变量的分级可以粗略一些。以每个亚组内所有对照记录涉及的根本死因为目标编码,将其构成比作为分配比例,在亚组内再分配“心衰”;合并所有亚组的再分配结果,以及前述的直接修改“心衰”的结果,即是最终的“心衰”再分配结果。

自2019年开始,GBD研究人员在“心衰”“败血症”“栓塞”“肝衰竭”“呼吸衰竭”等多个垃圾编码的再分配上,应用了粗化精确匹配法的原理(操作过程和分亚组方式有所区别)^[10]。对于死因数据库中不具有死因链信息的国家和地区,GBD2019研究在上述垃圾编码的再分配时,综合应用了其他国家和地区的分配模式结果^[10]。

5. 建立较为复杂的方程:近几年来,有研究者开始尝试使用更为复杂的建模过程来再分配垃圾编码。下面主要介绍Foreman等^[12]及Ellingsen等^[13]的研究。

(1)Foreman等^[12]再分配美国死因数据库中的垃圾编码:Foreman等^[12]分析了美国1979-2011年的死因数据库,将根本死因中出现的垃圾编码分为9类,并为每一类垃圾编码列出了对应的目标编码范围,例如“部位不明确的肿瘤”只可能被再分配到肿瘤,“心衰”只可能被再分配到慢性非传染性疾病,且需要去除肿瘤、精神疾病及神经系统疾病,“未明原因的死亡”可以被再分配到任何疾病。他们选出死因链中包含某个垃圾编码,且根本死因是该垃圾编码对应的目标编码之一的记录,作为这个垃圾编码再分配建模的分析子集(training data set)。以分析子集中所有目标编码(即根本死因)为因变量,以死亡记录中的人口学信息、死亡地点、死亡年份、死因链中所有不是垃圾编码的疾病为自变量,建立贝叶斯混合效应多项式logistic回归模型,估算在一定性别、年龄、死亡地点和年份等的情境下,死因链中包含某些疾病时,根本死因为各个目标编码的概率。这里作为自变量的疾病需要注意两点,其一,不是仅排除预分配的垃圾编码,是排除所有的垃圾编码;其二,仅使用死因链中的疾病,不考虑其他疾病诊断部分。使用模型结果的方法为:当一条记录的根本死因为该垃圾编码,死因链中出现

了某些疾病,则找到相应性别、年龄、死亡地点和年份下,该疾病对应的出现概率最大的目标编码,将该垃圾编码再分配到这个目标编码。

这一方法背后的假设是将“某垃圾编码和某疾病同时出现于死因链中时,根本死因最有可能是某目标编码”的模式扩大到“该疾病出现于死因链中,该垃圾编码在根本死因位置”的情况下(人口学、死亡地点、死亡年份等同时)。这种假设其实并没有考虑到从死因链发展到根本死因的病理生理过程,所以虽然统计建模方法先进,且考虑了众多协变量,但在实际应用中应该谨慎核查再分配的结果。

(2) Ellingsen 等^[13]再分配“未明原因的伤害”:利用挪威的死因登记数据研究了如何再分配“未明原因的伤害”。首先,从死因库中提取出根本死因属于伤害的记录;将伤害分为 7 类,除了“未明原因的伤害”外,还有道路交通伤害、意外跌倒、意外中毒、其他意外事件、故意自害及被人加害。第二步,以其他 6 类伤害为因变量,以人口学信息、死亡年份、死亡地点、是否描述受伤场景、是否进行解剖、身体损伤的性质(ICD-10 中 S 和 T 开头的编码)等为自变量,建立多项式 logistic 回归模型。建模过程重复 1 000 次,每次都是将分析数据库随机分为两部分,67% 的记录进入“训练集”,33% 的记录进入“验证集”,利用训练集建模后,将模型形式代入验证集,观察其预测结果与实际情况间的符合程度。最终,将预测效果最好的模型形式应用于根本死因为“未明原因的伤害”的记录,利用记录中的上述信息,估计死者根本死因。

这一方法在预测时,全面考虑了死于伤害的死者多方面信息;因此,它对死因记录的质量要求相对较高。Ellingsen 等^[13]报告他们的研究结果为,97% 的“未明原因的伤害”被再分配为“意外跌倒”。出现这种高度集中现象的原因还值得进一步探讨。

6. 其他来源数据:两项在巴西开展的研究^[14-15]利用住院信息系统中患者的主要诊断来重新填写同一人死亡后证明书上的根本死因(如果本来为垃圾编码)。这种方法不具有普遍性,原因有两个方面:第一,其应用条件严苛,需要住院信息数据库质量高,且每个人都能找到唯一编码来完成链接;第二,需要耗费较多的时间和人力。

三、讨论

由此可见,死因数据中垃圾编码的问题已经在全世界很多国家和地区引起了重视。虽然已有研究者提出了几种建立复杂统计模型再分配垃圾编码的方法,但 GBD 研究目前为止主要应用的仍是专家咨询法、给定比例分配法、简单线性回归以及粗化精确匹配^[10]。其原因可能包括:第一,新的方法还处于提出并讨论阶段,它们再分配垃圾编码的结果是否符合疾病的病理生理发展过程、当地人群的疾病变化趋势等,尚未经过强有力的验证;第二,越是复杂的、涉及变量多的模型,在不同区域、年代等之间需要进行的微调越多,GBD 研究的对象是全球范围内的数据,过多的模型调整可能影响分析效率,甚至造成硬件设备不能支持分析过程

的情况;第三,不同国家和地区死因登记数据的质量差别很大,有些数据库可能不具备应用复杂垃圾编码再分配方法的条件。

我国目前进行全国和分省的疾病负担估算时,使用的垃圾编码再分配方法基本与 GBD 研究保持一致,但在分配比例方面根据中国实际情况进行了部分调整^[16-17]。针对垃圾编码再分配方法及其应用场景的验证和改进,本团队建议,我国可以从两个方面来深入开展研究,做出贡献。一是继续验证 GBD 研究中关于中国的垃圾编码的再分配结果是否符合中国实际情况。例如,本团队曾经发现,GBD 研究在处理中国人群死因监测数据中的“肺源性心脏病”时,采用的再分配比例不符合中国实际情况,导致了对中国人群 IHD 死亡流行趋势的误判,这一研究已经发表^[16],得到了 GBD 研究团队的认可,因此,在后续研究中分析中国数据时,根据本团队的研究结果,修改了“肺源性心脏病”再分配的比例。二是在比省更小一级的市(县)层面上开展研究,探索垃圾编码再分配各种方法能够适用的地域范围。

本团队认为,考虑死因垃圾编码的再分配时,最重要的基础理念是要尊重疾病在个体和人群中的发展规律,其次是从死者本身特性、死因登记工作的流程、地区风俗等多个方面,考虑影响死因登记质量的变量,作为划分亚组变量或是纳入统计方程的协变量。在死因链完整时,利用其中的疾病信息是有必要的。从这几个方面来判断,上面提到的方法中,Stevens 等^[11]和 Ellingsen 等^[13]的方法相对合理些,而 Foreman 等^[12]的方法虽然利用了死因链和其他信息,但有可能忽略了疾病在个体上的发展过程,反而不宜推广。

从根本上来说,WHO 关于疾病编码的规则和指导是死因编码工作应该参照的金标准^[18]。我国从 2005 年起开始使用“死因登记管理信息系统”,通过网络直报的方式收集死亡个案信息^[19]。同时,这些年来,中国 CDC 每年对地方的死因工作人员进行定期培训,提高了死因上报人员和编码人员的专业能力。这些工作从源头入手提升了死因监测数据的质量,使得我国的死因监测数据中垃圾编码的占比从 1980-2009 年的平均 49.2%,减少到 2010-2019 年的平均 28.3%,在 GBD 研究的数据质量评价中获得 3 星评价(0~5 星,5 星为质量最好,即垃圾编码占比最低)^[10]。可见,这些工作具有重要意义,应持续开展及继续发展。

利益冲突 所有作者声明无利益冲突

参 考 文 献

- [1] 北京协和医院世界卫生组织,国际分类家族合作中心. 疾病和有关健康问题的国际统计分类第十次修订本第二卷指导手册[M]. 2 版. 董景五,译. 北京:人民卫生出版社,2008.
- [2] Murray CJ, Lopez AD. The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020[M]. Cambridge, MA: Harvard University Press, 1996.

- [3] Naghavi M, Makela S, Foreman K, et al. Algorithms for enhancing public health utility of national causes-of-death data[J]. *Popul Health Metr*, 2010, 8(1):9. DOI:10.1186/1478-7954-8-9.
- [4] GBD 2017 Causes of Death Collaborators. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980 - 2017: a systematic analysis for the Global Burden of Disease Study 2017[J]. *Lancet*, 2018, 392(10159):1736-1788. DOI: 10.1016/S0140-6736(18)32203-7.
- [5] 章亚男, 宇传华. 死因统计中的垃圾编码及其处理方法[J]. *公共卫生与预防医学*, 2016, 27(2):8-11. DOI:CNKI:SUN:FBYF.0.2016-02-003.
Zhang YN, Yu CH. Garbage codes and its redistribution method in the cause of death statistics[J]. *J Public Health Prev Med*, 2016, 27(2): 8-11. DOI: CNKI: SUN: FBYF. 0. 2016-02-003.
- [6] Lozano R, Murray CJL, Lopez AD, et al. Miscoding and misclassification of ischaemic heart disease mortality[R]. Geneva:World Health Organization, 2001.
- [7] Ahern RM, Lozano R, Naghavi M, et al. Improving the public health utility of global cardiovascular mortality data: the rise of ischemic heart disease[J]. *Popul Health Metr*, 2011, 9(1):8. DOI:10.1186/1478-7954-9-8.
- [8] GBD 2013 Mortality and Causes of Death Collaborators. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013:a systematic analysis for the Global Burden of Disease Study 2013[J]. *Lancet*, 2015, 385(9963):117-171. DOI:10.1016/S0140-6736(14)61682-2.
- [9] GBD 2015 Mortality and Causes of Death Collaborators. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015[J]. *Lancet*, 2016, 388(10053): 1459-1544. DOI: 10.1016/S0140-6736(16) 31012-1.
- [10] GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019[J]. *Lancet*, 2020, 396(10258): 1204-1222. DOI: 10.1016/S0140-6736(20) 30925-9.
- [11] Stevens GA, King G, Shibuya K. Deaths from heart failure: using coarsened exact matching to correct cause-of-death statistics[J]. *Popul Health Metr*, 2010, 8(1): 6. DOI: 10.1186/1478-7954-8-6.
- [12] Foreman KJ, Naghavi M, Ezzati M. Improving the usefulness of US mortality data: new methods for reclassification of underlying cause of death[J]. *Popul Health Metr*, 2016, 14(1): 14. DOI: 10.1186/s12963-016-0082-4.
- [13] Ellingsen CL, Ebbing M, Alfsen GC, et al. Injury death certificates without specification of the circumstances leading to the fatal injury-the Norwegian Cause of Death Registry 2005-2014[J]. *Popul Health Metr*, 2018, 16(1): 20. DOI:10.1186/s12963-018-0176-2.
- [14] Cascão AM, de Mello Jorge MHP, Costa AJL, et al. Use of primary diagnosis during hospitalization in the Unified Health System (Sistema Único de Saúde) to qualify information regarding the underlying cause of natural deaths among the elderly[J]. *Rev Bras Epidemiol*, 2016, 19(4):713-726. DOI:10.1590/1980-5497201600040003.
- [15] França E, Teixeira R, Ishitani L, et al. Ill-defined causes of death in Brazil: a redistribution method based on the investigation of such causes[J]. *Rev Saúde Pública*, 2014, 48(4):671-681. DOI:10.1590/s0034-8910.2014048005146.
- [16] Wan X, Yang GH. Is the mortality trend of ischemic heart disease by the GBD2013 study in China real? [J]. *Biomed Environ Sci*, 2017, 30(3): 204-209. DOI: 10.3967/bes2017.027.
- [17] Zhou MG, Wang HD, Zeng XY, et al. Mortality, morbidity, and risk factors in China and its provinces, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017[J]. *Lancet*, 2019, 394(10204):1145-1158. DOI: 10.1016/S0140-6736(19)30427-1.
- [18] WHO. ICD-10 Training Tool[EB/OL]. [2021-10-15]. <https://apps.who.int/classifications/apps/icd/icd10training/ICD-10%20training/Start/index.html>.
- [19] 万霞, 刘利群, 杨功焕. 2014-2016年云南省宣威市死因监测数据库清洗探讨[J]. *疾病监测*, 2018, 33(6):520-524. DOI:10.3784/j.issn.1003-9961.2018.06.017.
Wan X, Liu LQ, Yang GH. Discussion on data cleaning of vital registration system in Xuanwei, Yunnan, 2014-2016[J]. *Dis Surveill*, 2018, 33(6): 520-524. DOI: 10.3784/j.issn.1003-9961.2018.06.017.