

· 标准 · 方案 · 指南 ·

基于区域大数据平台的卒中高危人群队列研究 最小数据集标准专家共识

中国疾病预防控制中心慢性非传染性疾病预防控制中心, 国家神经系统疾病临床医学研究中心,
国家神经系统疾病医疗质量控制中心, 中国卒中学会脑血管病大数据与信息标准化分会

摘要: 开展基于真实世界的卒中高危人群大型队列研究对卒中防控具有重要意义, 然而各区域大数据平台数据元结构、变量定义和范围不一致, 难以实现不同区域的数据共享、汇总和分析。本研究在梳理卒中危险因素类别和定义的基础上, 结合区域大数据平台已有数据基础, 建立了一套统一的卒中高危人群队列研究最小数据集标准, 并形成专家共识。以期为实现区域内和区域间真实世界数据的比较、整合和共享提供参考, 在卒中危险因素队列研究、防控措施实施和评价中发挥作用。

关键词: 卒中; 危险因素; 大数据; 队列研究; 数据集

中图分类号: R743.3

文献标识码: A

文章编号: 2096-5087 (2021) 12-1189-10

Minimum dataset standard for cohort study of high-risk population of stroke based on regional big data platform: a consensus

National Center for Chronic and Non-communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention; China National Clinical Research Center for Neurological Diseases; National Center for Healthcare Quality Management in Neurological Diseases; Cerebrovascular Disease Big Data and Information Standardization Branch of Chinese Stroke Association

Abstract: A large cohort study of high-risk population of stroke based on the real world is of great significance for stroke prevention and control. However, the data element structures, variable definitions and scopes of regional big data platforms are inconsistent, which will be an obstacle for data sharing, summary, and analysis among different regions. In this study, we formed an expert consensus on a unified minimum dataset standard for the cohort study of high-risk population of stroke, considering the categories and definitions of risk factors of stroke, and the existing database of the regional big data platforms. The consensus shall provide a reference for the comparison, integration, and sharing of real world data within and between regions, and play an important role in the cohort study on risk factors of stroke, as well as the implementation and evaluation of prevention and control measures.

Keywords: stroke; risk factor; big data; cohort study; dataset

卒中是中国成年人致死、致残的首位病因^[1]。预防卒中发生,减轻卒中疾病负担的根本是针对卒中中的危险因素进行积极的、早期的和有效的干预,卒中高危人群是防控工作的关键群体^[1]。当前,各地积极推

进区域大数据平台建设,并基于平台开发了慢性病智能直报管理系统^[2-4]。但是各个区域大数据平台的建设构架和数据元结构不一致,且变量的定义和范围也不统一,难以实现不同区域的数据共享、汇总和分析,大数据应用受到一定限制。由于数据多元性不足,与真实世界研究存在差距,传统队列研究结果容易产生偏倚。因此,为了使研究人群更接近真实世界,充分整合存在区域差异的多来源数据,有必要建立一套科学、统一的卒中高危人群队列研究的最小数据集标准,实现区域内和区域间数据的比较、整合和共享,在卒中危险因素研究、防控措施实施和评价中发挥重要作用。

DOI: 10.19485/j.cnki.issn2096-5087.2021.12.001

基金项目: “十三五”国家重点研发计划“重大慢性非传染性疾病防控研究”(2017YFC1310902, 2017YFC1310901, 2017YFC1310903);
国家自然科学基金面上项目(81872721);宁波市医疗卫生品牌
学科资助(PPXK2018-10)

通信作者: 刘世炜, E-mail: liusw@chinacdc.cn;

李子孝, E-mail: lizixiao2008@hotmail.com

1 概要

卒中的危险因素种类繁多、来源广泛。根据中华医学会神经病学分会脑血管病学组完成并发布的《中国脑血管病一级预防指南2019》^[1], 卒中的危险因素包括: (1)不可干预的危险因素, 如年龄、性别、种族和遗传因素; (2)可干预的危险因素, 如行为生活方式(吸烟、饮食与营养、缺乏身体活动、超重与肥胖、饮酒等)和疾病危险因素(高血压、糖尿病、心房颤动、其他心脏病、血脂异常、无症状颈动脉狭窄、代谢综合征、高同型半胱氨酸等)。基于当前国家卫生政策、医疗卫生信息化和区域大数据平台建设现状, 卒中危险因素研究的数据来源不尽相同, 如年龄、性别、种族等人口学信息和行为生活方式、健康状况等信息主要来自居民健康档案; 高血压、糖尿病、卒中等慢性病随访和死亡结局数据大多来自疾病预防控制系统; 慢性病的发病、诊断、治疗等信息大多来自医院信息系统和实验室检测信息系统; 居民体检数据大多来自体检信息系统。

中国疾病预防控制中心慢性非传染性疾病预防控制中心(以下简称慢病中心)作为全国慢性病防控指导中心, 对慢性病防控信息化和大数据应用具有丰富的实践经验, 并在全国建立了多个应用中心。2004年以来, 开展了多次中国慢性病及危险因素监测和脑血管病流行病学调查^[5], 积累了丰富的数据分析和科研经验。为了更好地整合区域内和区域间不同来源数据,

充分发挥真实世界数据的优势, 科学、高效地开展基于区域大数据平台的卒中高危人群队列研究, 慢病中心联合国家神经系统疾病临床医学研究中心、国家神经系统疾病医疗质量控制中心和中国卒中学会脑血管病大数据与信息标准化分会共同制定了一套基于区域大数据平台的卒中高危人群队列研究最小数据集标准, 以期为区域内和区域间实现数据共享, 建立扩展性良好的卒中高危人群队列提供数据基础。

2 卒中危险因素的分类和定义

卒中危险因素的确定主要通过文献回顾、现场调研和专家咨询等方法。检索并梳理卒中及其危险因素的相关文献, 如《2016版中国脑血管病诊治指南与共识》^[6]、《中国心血管疾病预防指南(2017)》^[7]、《中国2型糖尿病防治指南(2017年版)》^[8]、《中国成人血脂异常防治指南(2016年修订版)》^[9]、《中国慢性病及其危险因素监测报告2013》^[5]以及疾病负担研究^[10-11]和卒中危险因素研究论文^[10-37], 初步制定卒中危险因素的分类和定义。现场调研湖北省宜昌市和浙江省宁波市的大数据平台, 分析平台上的卒中及其危险因素的定義、分类和数据来源, 了解平台构建卒中危险因素的参考依据, 进一步完善卒中危险因素的分类和定义。最后, 通过专家咨询确定卒中危险因素的分类和定义。见表1。

表1 卒中危险因素的定義

危险因素	定义
年龄	年龄越大, 风险越高
性别	男性
民族	少数民族
文化程度	文化程度越低, 风险越高
高脂血症	血清总胆固醇 ≥ 6.2 mmol/L (240 mg/dL) 或三酰甘油 ≥ 2.3 mmol/L (200 mg/dL) 或高密度脂蛋白胆固醇 ≤ 1.0 mmol/L (40 mg/dL) 或低密度脂蛋白胆固醇 ≥ 4.1 mmol/L (160 mg/dL)
高血压	在未用降压药物的情况下, 非同日3次测量, 收缩压 ≥ 140 mm Hg和(或)舒张压 ≥ 90 mm Hg; 或已被乡镇(社区)级或以上医院确诊为高血压且近2周服降压药者
糖尿病	典型糖尿病症状(烦渴多饮、多尿、多食、不明原因的体重下降)且随机静脉血浆葡萄糖 ≥ 11.1 mmol/L (200 mg/dL) 或空腹静脉血浆葡萄糖 ≥ 7.0 mmol/L (126 mg/dL) 或葡萄糖负荷后2小时血糖无典型糖尿病症状者, 复查确认葡萄糖负荷后2小时血糖 ≥ 11.1 mmol/L (200 mg/dL)
血糖异常	空腹静脉血浆葡萄糖受损: 空腹静脉血浆葡萄糖6.1~7.0 mmol/L且口服葡萄糖耐量试验2小时血浆葡萄糖 < 7.8 mmol/L; 糖耐量减低: 空腹血糖 < 7.0 mmol/L且服糖后2小时血糖为7.8~11.1 mmol/L
糖化血红蛋白异常	非贫血和血红蛋白异常疾病患者的糖化血红蛋白值 $\geq 7\%$
超重/肥胖	体质指数 ≥ 24.0 kg/m ²
中心性肥胖	成年男性腰围 ≥ 90 cm, 成年女性腰围 ≥ 85 cm
吸烟	连续或累计吸烟6个月及以上, 包括现在每天吸烟、现在偶尔吸烟、现在不吸烟以前吸烟
二手烟暴露	不吸烟者在1周内有一天以上的时间吸入吸烟者呼出的烟雾每天至少15 min

表 1 (续)

危险因素	定义
饮酒	偶尔饮酒: 饮酒, 但未达到过每周至少饮酒1次, 连续饮酒1年以上; 既往饮酒: 曾经达到过每周至少饮酒1次, 连续饮酒1年以上, 但1年前戒酒; 目前饮酒: 过去1年内, 每周至少饮酒1次, 连续饮酒1年以上
不合理饮食	摄入盐 ≥ 6 g/d; 油 ≥ 30 g/d; 畜禽肉 > 75 g/d; 水产品 < 40 g/d; 蛋类 > 50 g/d; 蔬菜类 < 300 g/d; 水果类 < 200 g/d; 水 $< 1\ 500$ mL/d; 谷类 < 125 g/d; 饮用浓茶; 饮用浓咖啡
身体活动不足	每周 < 150 min中等强度体力活动, 或 < 75 min高等强度体力活动, 或 < 600 代谢当量(MET)的中-高等强度混合型体力活动(中等强度体力活动MET值=4.0, 高等强度体力活动MET值=8.0, 骑单车体力活动和步行体力活动MET值=4.0) (中等强度体力活动MET $\times$ 分钟/周=4.0 $\times$ 中等强度体力活动分钟数 $\times$ 中等强度体力活动天数/周; 高等强度体力活动MET $\times$ 分钟/周=8.0 $\times$ 高等强度体力活动分钟数 $\times$ 高等强度体力活动天数/周; 步行体力活动MET $\times$ 分钟/周=4.0 $\times$ 步行体力活动分钟数 $\times$ 步行体力活动天数/周; 骑单车体力活动MET $\times$ 分钟/周=4.0 $\times$ 骑单车体力活动分钟数 $\times$ 骑单车体力活动天数/周)
病史或症状	无症状性颈动脉狭窄; 心血管病(缺血性心血管病、先天性高血压、冠心病、心肌炎、心肌病和心律失常等); 房颤; 左心室肥厚; 既往充血性心力衰竭史; 其他心脏病; 短暂脑缺血发作或卒中或血栓栓塞; 高同型半胱氨酸血症; 高凝状态; 脂蛋白升高; 脂代谢异常史; 高脂血症史; 睡眠呼吸障碍; 偏头痛; 脑外伤史; 肝、肾功能异常; 脑血管畸形; 脑肿瘤; 炎症(类风湿关节炎、系统性红斑狼疮、牙周炎、慢性支气管炎、幽门螺杆菌感染、肺炎支原体感染、巨细胞病毒感染)
用药史	绝经后激素治疗、降血脂药物、抗血小板药物、抗凝药物、降压药物、激素替代治疗、使用口服避孕药
家族史	一至三级亲属中至少1人患心脑血管疾病(心肌梗死, 卒中等)
低出生体重	出生体重 $< 2\ 500$ g
吸毒	相对非吸毒者, 吸毒者风险较高
精神压力	精神压力越大, 风险越高
肾脏功能受损	尿蛋白/肌酐、肾小球滤过率: 等级越高, 肾脏功能受损越严重: 等级1: 尿蛋白/肌酐 < 30 mg/g, 肾小球滤过率 ≥ 60 mL $\cdot$ min $^{-1}\cdot(1.73\text{ m}^2)^{-1}$; 等级2: 尿蛋白/肌酐 $30\sim 299$ mg/g, 肾小球滤过率 $30\sim 59$ mL $\cdot$ min $^{-1}\cdot(1.73\text{ m}^2)^{-1}$; 等级3: 尿蛋白/肌酐 ≥ 300 mg/g, 肾小球滤过率 $15\sim 29$ mL $\cdot$ min $^{-1}\cdot(1.73\text{ m}^2)^{-1}$; 等级4: 尿蛋白/肌酐 ≥ 300 mg/g, 肾小球滤过率 < 15 mL $\cdot$ min $^{-1}\cdot(1.73\text{ m}^2)^{-1}$ 。 血肌酐值越高, 肾脏功能受损越严重
血常规	血浆纤维蛋白原增高; 超敏C反应蛋白浓度增高; 血常规白细胞计数增高; 脂蛋白相关磷脂酶A2增高
空气污染	室外空气污染: PM _{2.5} 质量浓度 $> 4.15\ \mu\text{g}/\text{m}^3$; NO ₂ 值越高; SO ₂ 值越高 室内空气污染: 家庭使用固体燃料做饭(煤、木材、木炭、粪便和农业残留物)

3 最小数据集标准

基于卒中危险因素分类和定义,初步建立卒中高危人群队列研究最小数据集标准,结合区域大数据平台现场调研和专家咨询,经过多次修订,制定卒中高危人群队列研究最小数据集标准。该数据集标准以国家卫生行业标准和国家分类标准为依据,按照不同的数据来源和类别分组,共计5类9个数据集,包括不可改变类危险因素、行为方式类危险因素、疾病类危险因素、结局类信息和其他危险因素。其中,行为方式类危险因素和疾病类危险因素根据数据来源分为健康档案、体检信息和随访信息;结局类信息分为卒中发病和死亡信息;其他危险因素主要来自医院,按照具体数据来源分类。所有数据集通过主键ID和身份证号串联。

3.1 不可改变类危险因素 不可改变类危险因素主

要来源于居民健康档案,包括索引、性别、出生日期、民族、文化程度、职业和婚姻状况等人口学基本信息。可提供年龄、性别等危险因素信息,而民族、文化程度、职业和婚姻状况等可用于社会经济水平类危险因素探索。主键ID、身份证号和居民健康档案号可以确保每个个体的唯一性。民族、文化程度、职业和婚姻状况参考国家标准定义及分类。见表2(2-1-1)。

3.2 行为方式类危险因素 行为方式类危险因素主要来源于健康档案、体检信息和随访信息,包括索引、吸烟、饮酒、运动、饮食和职业相关数据。其中厨房排风设施和厨房燃料类型可间接反映室内环境污染程度。见表2(2-2-1)。

3.3 疾病类危险因素 疾病类危险因素根据数据来源分为3类:(1)居民健康档案中的医疗数据相关信息,主要为历史健康状态,包括索引、医疗费用支付方

式、药物过敏史、暴露史、疾病史、家族史、遗传史和残疾史,见表2(2-3-1)。(2)健康体检信息,反映居民体检时的健康状态,包括索引、体检日期、体检相关信息、体检机构名称及代码。其中本次体检ID用于区分同一患者的多次体检,体检机构名称和代码用以区别不同体检机构。见表2(2-3-2)。(3)随访信息,根据卒中发病和复发的高危人群分为高血压患者、糖尿病患者和卒中患者,通过对卒中重点人群的随访,收集过程信息并开展队列研究。高血压患者随访数据包括索引、患者临床信息以及相关行为习惯,其中高血压确诊机构等级和随访方式参考统一标准,见表2(2-3-3)。糖尿病随访数据包括索引、患者临床信息、药物使用信息及行为方式等,见表2(2-3-4)。卒中患者随访数据包括索引及其临床相关信息,可补充心血管病卡片信息,见表2(2-3-5)。

3.4 结局信息 结局信息分为:(1)报卡信息,主要为卒中的发病信息,反映高危人群的发病进展结局。心脑血管上报卡信息包括索引、上报日期、卒中诊断相关信息、病史、是否首次发病、发病日期等,可提供卒中发病信息进行卒中发病的队列研究,见表2(2-4-1)。(2)死亡信息,主要为居民死亡记录,反映卒中患者的死亡进展结局。居民死亡记录包括索引及死亡相关信息,最高死亡诊断依据代码和死亡诊断的最高单位代码参考相应代码表,见表2(2-4-2)。

3.5 其他危险因素 其他危险因素来源于医院信息系统。根据数据的具体来源分为门诊诊断和住院诊断信息、病案首页信息、心电图检查信息、实验室检测信息和影像检查信息。该类信息既可为疾病类危险因素提供更详细的临床数据,也可进一步探索卒中发病或死亡的其他危险因素。

表2 卒中高危人群队列研究最小数据集标准

字段名称	类型及长度	说明
ID	VARCHAR2 (32)	主键ID: SYS_GUID () (不可空)
IDENTITY_NO	VARCHAR2 (18)	身份证号: 用于关联唯一标识
2-1 不可改变类危险因素		
2-1-1 人口学基本信息		
HEALTH_CODE	VARCHAR2 (17)	居民健康档案号
HEALTH_DATE	VARCHAR2 (12)	建档日期: 格式YYYYMMDDHHmm
SEX	VARCHAR2 (1)	性别: 1.未知 2.男 3.女 4.未说明
BIRTHDATE	VARCHAR2 (8)	出生日期: 格式为YYYYMMDD (例如19910101)
BLOOD	VARCHAR2 (1)	ABO血型: 1.A型 2.B型 3.O型 4.AB型 5.不详
RH_TYPE	VARCHAR2 (1)	RH阴性: 1.否 2.是 3.不详
NATION	VARCHAR2 (3)	民族: 参考GB/T 3304—1991
EDUCATION	VARCHAR2 (1)	文化程度: 参考GB/T 4658—2006
OCCUPATION	VARCHAR2 (3)	职业: 参考GB/T 6565—1999
MARRIED	VARCHAR2 (2)	婚姻状况: 参考GB/T 4766—1984婚姻状况代码
REGISTER_ZONECODE	VARCHAR2 (6)	户籍地址编码
REGISTER_ADDR	VAHRCAR2 (100)	户籍地址
ADDZONE_CODE	VARCHAR2 (6)	常住地址编码
ADDR	VARCHAR2 (100)	常住地址
2-2 行为方式类危险因素		
2-2-1 居民行为方式类危险因素		
SMOKE_STATUS	VARCHAR2 (1)	吸烟状态: 1.从不吸烟 2.过去吸烟,已经戒烟 3.现在吸烟
SMOKE_AGE	NUMBER (2)	开始吸烟年龄,单位:岁
SMOKE_CESSATION_AGE	NUMBER (2)	戒烟年龄,单位:岁
SMOKE_CESSATION_TIME	NUMBER (2)	戒烟时长,单位:a
SECONDHAND_SMOKE_COUNT	VARCHAR2 (2)	每周二手烟暴露次数: 1.每天 2.平均每周4~6 d 3.平均每周1~3 d 4.没有 5.不知道/记不清
DAILY_SMOKERS	NUMBER (3)	平均日吸烟量,单位:支
DRINK_STATUS	VARCHAR2 (1)	饮酒状态: 1.喝过,在过去30天前 2.喝过,在过去30天内 3.没喝过

表 2 (续)

字段名称	类型及长度	说明
DRINK_FREQUENCY	VARCHAR2 (1)	饮酒频率: 1.从不 2.偶尔 3.经常 4.每天
DRINK_EACH	NUMBER (3)	日均饮酒量, 单位: g
DRINK_TYPE	VARCHAR2 (1)	饮酒种类: 1.白酒 2.啤酒 3.红酒 4.黄酒 5.其他
DRINK_AGE	NUMBER (3)	开始饮酒年龄, 单位: 岁
DRINKEN	VARCHAR2 (1)	近1年内是否曾醉酒: 1.是 2.否
DRINK_QUIT	VARCHAR2 (1)	是否戒酒: 1.未戒酒 2.已戒酒
QUIT_DRINK_AGE	NUMBER (3)	戒酒年龄, 单位: 岁
FLAG_EXERCISE	VARCHAR2 (1)	是否运动锻炼: 1.是 2.否
EXERCISE_FREQUENCY	VARCHAR2 (2)	运动频率: 1.从不 2.偶尔 2.1 1~3 d/月 2.2 少于1 d/月 3.经常 3.1 5~6 d/周 3.2 3~4 d/周 3.3 1~2 d/周 4.每天
EXERCISE_TIME	NUMBER (3)	每次锻炼的时间, 单位: min
EXERCISE_LONG	NUMBER (2)	坚持锻炼的时间, 单位: a
EXERCISE_TYPE	VARCHAR2 (100)	锻炼方式
DIET	VARCHAR2 (1)	饮食习惯: 1.荤素均衡 2.荤食为主 3.素食为主 4.嗜盐 5.嗜油 6.嗜糖
SALT_RANK	VARCHAR2 (2)	食盐摄入量分级: 1.轻 2.中 3.重
EXPOSE_HISTORY	VARCHAR2 (1)	接触史: 1.无 2.有
WORK_TYPE	VARCHAR2 (50)	工种: 参考GB 6565—1996
WORK_TIME	NUMBER (3)	工作时间, 单位: a
DUST	VARCHAR2 (100)	职业病接触粉尘: 1.无 2.有
DUST_PROTECT	VARCHAR2 (1)	粉尘防护措施: 1.无 2.有
RADIO	VARCHAR2 (100)	职业病接触放射性物质: 1.无 2.有
RADIO_PROTECT	VARCHAR2 (1)	放射性物质防护措施: 1.无 2.有
PHYSICAL_FACTOR	VARCHAR2 (100)	职业病接触物理因素: 1.无 2.有
PHYSICAL_PROTECT	VARCHAR2 (1)	物理因素防护措施: 1.无 2.有
CHEMICAL_FACTOR	VARCHAR2 (100)	职业病接触化学物质: 1.无 2.有
CHEMICAL_PROTECT	VARCHAR2 (1)	化学物质防护措施: 1.无 2.有
OTHER_FACTOR	VARCHAR2 (100)	职业病接触其他危险因素: 1.无 2.有
OTHER_PROTECT	VARCHAR2 (1)	职业病其他危险因素防护措施: 1.无 2.有
KITCHEN_EXHAUST	VARCHAR2 (1)	厨房排风设施: 1.无 2.油烟机 3.换气扇 4.烟囱
FUEL_TYPE	VARCHAR2 (1)	厨房燃料类型: 1.液化气 2.煤 3.天然气 4.沼气 5.柴火 6.其他
DATA_SOURCE	VARCHAR2 (10)	数据来源: 1.健康档案 2.体检 3.其他
HEALTH_CODE	VARCHAR2 (17)	健康档案号: 如果本条数据来自于健康档案
CHECK_ID	VARCHAR2 (18)	体检ID: 如果本条数据来自于体检
TRANSACTION_DATE	VARCHAR2 (20)	本条数据的记录时间: YYYYMMDDHHmmSS
2-3 疾病类危险因素		
2-3-1 基本医疗信息 (健康档案)		
INSURANCE_PAY	VARCHAR2 (1)	医疗费用支付方式: 1.城镇职工基本医疗保险 2.城镇居民基本医疗保险 3.新农合 4.贫困救助 5.商业医疗保险 6.全公费 7.全自费 8.其他
ALLERGY	VARCHAR2 (1)	药物过敏史: 1.无 2.青霉素 3.磺胺 4.链霉素 5.其他
EXPOSURE_HISTORY	VARCHAR2 (1)	暴露史: 1.化学品 2.毒物 3.射线 4.其他
HYPERTEN_HISTORY	VARCHAR2 (1)	高血压既往史: 1.无 2.有
HYPERTEN_TIME	VARCHAR2 (12)	高血压确诊时间: YYYYMMDDHHmm
DIABETES_HISTORY	VARCHAR2 (1)	糖尿病既往史: 1.无 2.有
DIABETES_TIME	VARCHAR2 (12)	糖尿病确诊时间: YYYYMMDDHHmm
CHD_HISTORY	VARCHAR2 (1)	冠心病既往史: 1.无 2.有
CHD_TIME	VARCHAR2 (12)	冠心病确诊时间: YYYYMMDDHHmm

表 2 (续)

字段名称	类型及长度	说明
COPD_HISTORY	VARCHAR2 (1)	慢性阻塞性肺疾病既往史: 1.无 2.有
COPD_TIME	VARCHAR2 (12)	慢性阻塞性肺疾病确诊时间: YYYYMMDDHHmm
CANCER_HISTORY	VARCHAR2 (1)	恶性肿瘤既往史: 1.无 2.有
CANCER_TIME	VARCHAR2 (12)	恶性肿瘤确诊时间: YYYYMMDDHHmm
STROKE_HISTORY	VARCHAR2 (1)	卒中既往史: 1.无 2.有
STROKE_TIME	VARCHAR2 (12)	卒中确诊时间: YYYYMMDDHHmm
MENTAL_HISTORY	VARCHAR2 (1)	严重精神障碍既往史: 1.无 2.有
MENTAL_TIME	VARCHAR2 (12)	严重精神障碍确诊时间: YYYYMMDDHHmm
TB_HISTORY	VARCHAR2 (1)	结核病既往史: 1.无 2.有
TB_TIME	VARCHAR2 (12)	结核病确诊时间: YYYYMMDDHHmm
HEPATITIS_HISTORY	VARCHAR2 (1)	肝炎既往史: 1.无 2.有
HEPATITIS	VARCHAR2 (12)	肝炎确诊时间: YYYYMMDDHHmm
FAMILY_HISTORY_FATHER	VARCHAR2 (2)	家族史(父亲): 1.无 2.高血压 3.糖尿病 4.冠心病 5.慢性阻塞性肺疾病 6.恶性肿瘤 7.卒中 8.严重精神障碍 9.结核病 10.肝炎 11.先天畸形 12.其他
FAMILY_HISTORY_MOTHER	VARCHAR2 (2)	家族史(母亲): 1.无 2.高血压 3.糖尿病 4.冠心病 5.慢性阻塞性肺疾病 6.恶性肿瘤 7.卒中 8.严重精神障碍 9.结核病 10.肝炎 11.先天畸形 12.其他
FAMILY_HISTORY_SIBLING	VARCHAR2 (2)	家族史(兄弟姐妹): 1.无 2.高血压 3.糖尿病 4.冠心病 5.慢性阻塞性肺疾病 6.恶性肿瘤 7.卒中 8.严重精神障碍 9.结核病 10.肝炎 11.先天畸形 12.其他
FAMILY_HISTORY_CHILD	VARCHAR2 (2)	家族史(子女): 1.无 2.高血压 3.糖尿病 4.冠心病 5.慢性阻塞性肺疾病 6.恶性肿瘤 7.卒中 8.严重精神障碍 9.结核病 10.肝炎 11.先天畸形 12.其他
INHERIT_HISTORY	VARCHAR2 (1)	遗传史: 1.无 2.有
INHERIT_DISEASE	VARCHAR2 (100)	遗传疾病名称
DISABILITY	VARCHAR2 (2)	残疾情况: 1.无残疾 2.视力残疾 3.听力残疾 4.言语残疾 5.肢体残疾 6.智力残疾 7.精神残疾 8.其他残疾
2-3-2 基本医疗信息 (居民体检信息)		
CHECK_ID	VARCHAR2 (18)	本次体检ID: 用于区分同一患者多次体检
CHECK_DATE	VARCHAR2 (20)	本次体检日期: 格式为YYYYMMDDHHmmSS
SYMPTOMS	VARCHAR2 (100)	本次体检症状: 1.无症状 2.头痛 3.头晕 4.心悸 5.胸闷 6.胸痛 7.慢性咳嗽 8.咳痰 9.呼吸困难 10.多饮 11.多尿 12.体重下降 13.乏力 14.关节肿痛 15.视力模糊 16.手脚麻木 17.尿急 18.尿痛 19.便秘 20.腹泻 21.恶心呕吐 22.眼花 23.耳鸣 24.乳房胀痛 25.其他
TEMPERATURE	NUMBER (3,1)	体温, 单位: ℃
PULSE_RATE	NUMBER (3)	脉率, 单位: 次/min
BREATHE_RATE	NUMBER (3)	呼吸频率, 单位: 次/min
SBP	NUMBER (3)	收缩压, 单位: mm Hg
DBP	NUMBER (3)	舒张压, 单位: mm Hg
HEIGHT	NUMBER (3,1)	身高, 单位: cm, 保留1位小数
WEIGHT	NUMBER (3,1)	体重, 单位: kg, 保留1位小数
BMI	NUMBER (3,1)	体质指数, 单位: kg/m ²
WAISTLINE	NUMBER (3,1)	腰围, 单位: cm, 保留1位小数
HIPLINE	NUMBER (3,1)	臀围, 单位: cm, 保留1位小数
MENTALITY	VARCHAR2 (2)	心理状态: 1.紧张 2.焦虑 3.抑郁 4.其他
HARZARD_STATUS	VARCHAR2 (2)	危险因素控制代码: 1.戒烟 2.健康饮酒 3.饮食 4.锻炼 5.减体重 6.建议接种疫苗 7.其他
HEART_RATE	NUMBER (3)	心率, 单位: 次/min
HEART_RATE_TYPE	VARCHAR2 (1)	心律: 1.齐 2.不齐 3.绝对不齐

表 2 (续)

字段名称	类型及长度	说明
HEART_NOISET	VARCHAR2 (1)	心脏杂音: 1.无 2.有
TJJG_NAME	VARCHAR2 (100)	体检机构名称
TJJG_CODE	VARCHAR2 (20)	体检机构代码
2-3-3 高血压患者随访数据		
FOLLOW_ID	VARCHAR2 (18)	本次随访ID: 用于区分同一患者多次随访记录
SBP	NUMBER (3)	收缩压, 单位: mm Hg
DBP	NUMBER (3)	舒张压, 单位: mm Hg
FLAG_HBP	VARCHAR2 (1)	是否高血压: 1.是 2.否
CONFIRM_ORG_RANK	VARCHAR2 (2)	高血压确诊机构等级: 参考CM-005-确诊医院等级
SYMPTOM	VARCHAR2 (300)	出现的症状
CHECK_WAY	VARCHAR2 (2)	高血压检出途径: 1.健康档案 2.首诊测压 3.普查 4.门诊就诊 5.免费测压点测压 6.其他
HBP_TYPE	VARCHAR2 (1)	高血压类型: 1.原发性高血压 2.继发性高血压 3.不详
HBP_GRADE	VARCHAR2 (1)	高血压分级: 1.一级 2.二级 3.三级
COMPLICATION	VARCHAR2 (2)	高血压并发症: 1.缺血性卒中 2.脑出血 3.短暂性脑出血发作 4.心肌梗死 5.心绞痛 6.充血性心力衰竭 7.糖尿病肾病 8.肾功能衰竭 9.无
FAMILY_DISEASE	VARCHAR2 (2)	直系亲属疾病: 1.高血压 2.冠心病 3.卒中 4.糖尿病 5.无
SELFCARE	VARCHAR2 (2)	生活自理行为: 1.可自理 2.轻度依赖 3.中度依赖 4.不能自理
MEDICINE_COMPLIANCE	VARCHAR2 (1)	服药依从性: 1.规律 2.间断 3.不服药
DRUG_NAME	VARCHAR2 (50)	服用药物名称
FOLLOW_UP_TYPE	VARCHAR2 (2)	随访方式: 参考 CV06.00.207-随访方式代码表
FOLLOW_UP_TIME	VARCHAR2 (20)	本次随访日期: 格式为YYYYMMDDHHmmSS
2-3-4 糖尿病随访数据		
FOLLOW_ID	VARCHAR2 (18)	本次随访ID: 用于区分同一患者的多次随访
FOLLOW_TIME	VARCHAR2 (20)	随访日期: 格式为YYYYMMDDHHmmSS
FLAG_DIABETES	VARCHAR2 (1)	是否诊断为糖尿病: 1.是 2.否
CONFIRM_ORG_RANK	VARCHAR2 (2)	糖尿病确诊机构等级: 参考CM-005-确诊医院等级
SYMPTOM	VARCHAR2 (50)	本次随访出现的症状
DIABETES_TYPE	VARCHAR2 (1)	糖尿病种类: 参考CM-101
FBG	NUMBER (4,1)	空腹血糖值, 单位: mmol/L
BSTH	NUMBER (4,1)	餐后2小时血糖, 单位: mmol/L
FLAG_FAMILY_HISTORY	VARCHAR2 (1)	是否糖尿病家族史: 1.是 2.否 3.未知 4.拒绝回答
COMPLICATION	VARCHAR2 (1)	糖尿病并发症代码: 参考CM-102
DRUG_NAME	VARCHAR2 (50)	服用药物名称
DRUG_FREQUENCY	VARCHAR2 (20)	药物使用频率
DRUG_DOSE_UNIT	VARCHAR2 (6)	药物剂量单位
DRUG_DOSE_PER	NUMBER (5,2)	药物单次剂量
DRUG_DOSE_TOTAL	NUMBER (12,2)	药物使用总剂量
DRUG_WAY	VARCHAR2 (3)	药物使用途径: 参考 CV06.00.102用药途径代码
MEDICINE_COMPLIANCE	VARCHAR2 (1)	服药依从性: 1.规律 2.间断 3.不服药
FOLLOW_DOCTOR	VARCHAR2 (1)	遵医行为: 1.良好 2.一般 3.差
FOLLOW_EVALUATION	VARCHAR2 (1)	随访结果评价: 参考 CV05.10.012随访评价结果代码
DIET_REASONABLE	VARCHAR2 (1)	饮食合理性: 1.合理 2.基本合理 3.不合理
FLAG_LBS	VARCHAR2 (1)	低血糖反应: 1.无 2.偶尔 3.频繁
MENTALITY	VARCHAR2 (2)	心理状态: 1.紧张 2.焦虑 3.抑郁 4.其他

表 2 (续)

字段名称	类型及长度	说明
2-3-5 卒中患者随访数据		
FOLLOW_UP_TIME	VARCHAR2 (20)	本次随访日期: 格式为 YYYYMMDDHHmmSS
FOLLOW_UP_TYPE	VARCHAR2 (2)	随访方式: 参考 CV06.00.207-随访方式代码
FOLLOW_ID	VARCHAR2 (18)	本次随访ID: 区分同一患者的多次随访记录
STROKE_TYPE	VARCHAR2 (1)	卒中类型: 1.蛛网膜下出血 2.脑出血 3.脑梗死 4.未特指
LAST_ATTACK_TIME	VARCHAR2 (20)	最近一次发病日期: 格式为 YYYYMMDDHHmmSS
FIRST_ATTACK_TIME	VARCHAR2 (20)	第一次发病日期: 格式为 YYYYMMDDHHmmSS
FLAG_IN_HOSPITAL	VARCHAR2 (1)	是否到医院就诊: 1.是 2.否
CONFIRM_TIME	VARCHAR2 (20)	卒中确诊日期: 格式为 YYYYMMDDHHmmSS
TREATMENT	VARCHAR2 (500)	卒中治疗情况
RANKIN_SCORE	NUMBER (3)	RANKIN 评分
MEDICINE_CODE	VARCHAR2 (2)	服药情况代码: 参考 CV06.00.230-用药情况代码
RECURRENCE_TIME	VARCHAR2 (20)	卒中复发时间: 格式为 YYYYMMDDHHmmSS
FLAG_DEAD	VARCHAR2 (1)	是否发病 28 d 内死亡: 1.是 2.否
BODY_STATE	VARCHAR2 (1)	目前患者状态: 1.稳定 2.好转 3.恶化
STROKE_FAMILY_HISTORY	VARCHAR2 (1)	卒中家族史: 1.有 2.无 3.不详
FLAG_SEQUELAE	VARCHAR2 (1)	后遗症: 1.是 2.否
COMPLICATION	VARCHAR2 (2)	卒中并发症: 参考 CV05.01.045-卒中并发症情况代码
2-4 结局信息		
2-4-1 心脑血管上报卡信息数据		
REPORT_ID	VARCHAR2 (18)	上报ID: 用于区分同一患者的多次上报
REPORT_TIME	VARCHAR2 (20)	上报日期
ICD_CODE	VARCHAR2 (20)	卒中诊断 ICD 编码
STROKE_TYPE	VARCHAR2 (2)	卒中类型: 1.蛛网膜下出血 2.脑出血 3.脑梗死 4.未特指
STROKE_DIAG_EVIDENCE	VARCHAR2 (2)	卒中诊断依据: 1.急性起病 2.症状体征 3.CT 4.MRI 5.血管造影 6.腰穿 7.手术 8.尸检或病理 9.死亡补发
AMI_DIAGNOSE	VARCHAR2 (2)	急性心肌梗死诊断依据: 1.心肌坏死生化标志物升高 2.心脏缺血症状 3.心电图提示新发缺血性改变 4.心电图提示病理性 Q 波形成 5.影像学证据提示新发局部室壁运动异常或存活心肌丢失 6.手术 7.尸检或病理 8.死亡补发
SCD_DIAGNOSE	VARCHAR2 (2)	心脏性猝死诊断依据: 1.临床 2.推断 3.死亡补发
SCD_DIAGNOSE_DEDUCE	VARCHAR2 (2)	心脏性猝死推断: 1.有目击者情况下发病 1 h 内死亡且排除可能非心脏原因死亡 2.无目击者情况下 24 h 内突发死亡且排除可能非心脏原因死亡
DISEASE_HISTORY	VARCHAR2 (2)	病史: 参考 CM-004-病史代码
FLAG_FIRST_ATTACK	VARCHAR2 (1)	是否首次发病: 1.是 2.否
ATTACK_TIME	VARCHAR2 (20)	发病日期
CONFIRM_ORG_RANK	VARCHAR2 (2)	确诊医疗机构等级: 参考 CM-005-确诊医院等级
FLAG_RESIDENT	VARCHAR2 (1)	是否常住人口: 1.是 2.否
2-4-2 居民死亡记录数据		
DEAD_TIME	VARCHAR2 (20)	死亡时间: 格式为 YYYYMMDDHHmmSS
FLAG_DEAD_HOSPITAL	VARCHAR2 (1)	是否院内死亡: 1.是 2.否
DEAD_DIAGNOSIS_NAME	VARCHAR2 (50)	根本死亡诊断名称
DEAD_ICD_CODE	VARCHAR2 (20)	死亡诊断 ICD 编码: ICD-10
DEAD_EVIDENCE	VARCHAR2 (200)	死亡依据
DEAD_DESCRIPTION	VARCHAR2 (200)	死亡情况描述
DEAD_INDUCION	VARCHAR2 (200)	导致死亡的诱因描述
DEAD_EVIDENCE_CODE	VARCHAR2 (200)	最高死亡诊断依据代码: 参考 CV05.01.037-死亡最高诊断依据类别代码

表 2 (续)

字段名称	类型及长度	说明
DEAD_INTERVAL	NUMBER (3)	发病到死亡的时间间隔
INTERVAL_UNIT	VARCHAR2 (1)	发病到死亡的时间间隔单位: 1.h 2.d 3.月 4.a
DEAD_ORG_CODE	VARCHAR2 (2)	出具死亡诊断的最高单位代码: 1.三级医院 2.二级医院 3.乡镇卫生院或社区卫生服务机构 4.村卫生室 5.其他医疗卫生机构 6.未就诊
DEAD_ORG_NAME	VARCHAR2 (50)	出具死亡诊断的单位名称

撰写委员会专家名单 刘世伟(中国疾病预防控制中心慢性非传染性疾病预防控制中心、中国疾病预防控制中心控烟办公室)、张培(宜昌市疾病预防控制中心)、李辉(宁波市疾病预防控制中心)、张良(宁波市疾病预防控制中心)、吴周志(宜昌市疾病预防控制中心)、徐承中(宜昌市疾病预防控制中心)、李子孝(首都医科大学附属北京天坛医院)、邢立颖(辽宁省疾病预防控制中心)、周立新(北京协和医院)、李思萱(宁波市疾病预防控制中心)、段东辉(宁波市疾病预防控制中心)、王永(宁波市疾病预防控制中心)、纪威(宁波市疾病预防控制中心)、姜勇(首都医科大学附属北京天坛医院)、王安心(首都医科大学附属北京天坛医院)、潘岳松(首都医科大学附属北京天坛医院)、曾新穎(中国疾病预防控制中心控烟办公室)、刘杨(上海市浦东新区疾病预防控制中心)、董文兰(中国疾病预防控制中心慢性非传染性疾病预防控制中心)、潘旭东(青岛大学附属医院)、张吉玉(山东省疾病预防控制中心)、殷黎(湖南省疾病预防控制中心)、韩丽媛(中国科学院大学宁波华美医院)

主要执笔人 刘世伟、张培、李辉、张良、吴周志、徐承中、曾新穎、刘杨、董文兰、潘旭东、李子孝

志谢 感谢山东众阳健康科技集团有限公司李涛、王雷、石晓阳、田志勇和张文杰等工程师在数据标准、最小数据集标准和数据字典等方面提供的大力支持

参考文献

- [1] 中华医学会神经病学分会, 中华医学会神经病学分会脑血管病学组. 中国脑血管病一级预防指南 2019 [J]. 中华神经科杂志, 2019, 52 (9): 684-700.
- [2] 吴周志, 徐承中, 徐勇, 等. 健康医疗大数据下疾病监测报告模式的变革——宜昌市健康管理大数据分析中心建设的启示 [J]. 疾病监测, 2019, 34 (1): 10-14.
- [3] 李辉, 张良, 李思萱, 等. 基于全民健康信息平台的宁波市慢性非传染性疾病智能直报模式建设 [J]. 首都公共卫生, 2019, 13 (3): 114-117.
- [4] 叶荔娜, 赵飞, 陈坚, 等. 基于智能电子健康档案平台的大数据应用研究与实践 [J]. 中国卫生信息管理杂志, 2019, 16 (6): 672-676.
- [5] 中国疾病预防控制中心. 中国慢性病及其危险因素监测报告 2013 [M]. 北京: 军事医学科学出版社, 2015.
- [6] 中华医学会神经病学分会. 2016 版中国脑血管病诊治指南与共识 [M]. 北京: 人民卫生出版社, 2016.
- [7] 中国心血管病预防指南 (2017) 写作组, 中华心血管病杂志编辑委员会. 中国心血管病预防指南 (2017) [J]. 中华心血管病杂志, 2018, 46 (1): 10-25.
- [8] 中华医学会糖尿病学分会. 中国 2 型糖尿病防治指南 (2017 年版) [J]. 中国实用内科杂志, 2018, 38 (4): 292-344.
- [9] 中国成人血脂异常防治指南修订联合委员会. 中国成人血脂异常防治指南 (2016 年修订版) [J]. 中国循环杂志, 2016, 31 (10): 937-953.
- [10] BAI Y, WANG Y L, SHANTSILA A, et al. The global burden of atrial fibrillation and stroke: a systematic review of the clinical epidemiology of atrial fibrillation in Asia [J]. Chest, 2017, 152 (4): 810-820.
- [11] O'DONNELL M J, CHIN S L, RANGARAJAN S, et al. Global and regional effects of potentially modifiable risk factors associated with acute stroke in 32 countries (INTERSTROKE): a case-control study [J]. Lancet, 2016, 388 (10046): 761-775.
- [12] ANDERSEN K K, OLSEN T S. Married, unmarried, divorced, and widowed and the risk of stroke [J]. Acta Neurol Scand, 2018, 138 (1): 41-46.
- [13] BRAGG F, LI L M, YANG L, et al. Risks and population burden of cardiovascular diseases associated with diabetes in China: a prospective study of 0.5 million adults [J]. PLoS Med, 2016, 13 (7): 1-17.
- [14] COOMBES E, JONES A P, HILLSDON M. The relationship of physical activity and overweight to objectively measured green space accessibility and use [J]. Soc Sci Med, 2010, 70 (6): 816-822.
- [15] DU H, LI L, BENNETT D, et al. Fresh fruit consumption and major cardiovascular disease in China [J]. New Engl J Med, 2016, 374 (14): 1332-1343.
- [16] GAN Y, WU J, LI L Q, et al. Association of smoking with risk of stroke in middle-aged and older Chinese: evidence from the China National Stroke Prevention Project [J/OL]. Medicine (Baltimore), 2018, 97 (47) [2021-11-13]. <https://pubmed.ncbi.nlm.nih.gov/30461631>. DOI: 10.1097/MD.00000000000013260.
- [17] IRIBARREN C, DARBINIAN J, KLATSKY A L, et al. Cohort study of exposure to environmental tobacco smoke and risk of first ischemic stroke and transient ischemic attack [J]. Neuroepidemiology, 2004, 23 (1/2): 38-44.
- [18] JOSHIPURA K J, ASCHERIO A, MANSON J E, et al. Fruit and vegetable intake in relation to risk of ischemic stroke [J]. JAMA, 1999, 282 (13): 1233-1239.
- [19] LEE K K, MILLER M R, SHAH A S V. Air pollution and stroke [J]. J Stroke, 2018, 20 (1): 2-11.
- [20] LIU S, LI Y, ZENG X, et al. Burden of Cardiovascular diseases in China, 1990-2016: findings from the 2016 Global Burden of

- Disease Study [J]. *JAMA Cardiol*, 2019, 4 (4): 342-352.
- [21] LV J, YU C, GUO Y, et al. Adherence to healthy lifestyle and cardiovascular diseases in the Chinese population [J]. *J Am Coll Cardiol*, 2017, 69 (9): 1116-1125.
- [22] MCHUTCHISON C A, BACKHOUSE E V, CVORO V, et al. Education, socioeconomic status, and intelligence in childhood and stroke risk in later life: a meta-analysis [J]. *Epidemiology*, 2017, 28 (4): 608-618.
- [23] MENTE A, O ' DONNELL M, RANGARAJAN S, et al. Urinary sodium excretion, blood pressure, cardiovascular disease, and mortality: a community-level prospective epidemiological cohort study [J]. *Lancet*, 2018, 392 (10146): 496-506.
- [24] PETERS S A E, HUXLEY R R, WOODWARD M. Diabetes as a risk factor for stroke in women compared with men: a systematic review and meta-analysis of 64 cohorts, including 775 385 individuals and 12 539 strokes [J]. *Lancet*, 2014, 383 (9933): 1973-1980.
- [25] PORTER A K, RODRIGUEZ D A, FRIZZELLE B G, et al. The association between neighborhood environments and physical activity from pregnancy to postpartum: a prospective cohort study [J]. *J Urban Health*, 2019, 96 (5): 703-719.
- [26] RICCI C, WOOD A, MULLER D, et al. Alcohol intake in relation to non-fatal and fatal coronary heart disease and stroke: EPIC-CVD case-cohort study [J/OL]. *BMJ*, 2018, 361 [2021-11-13]. <https://pubmed.ncbi.nlm.nih.gov/29844013>. DOI: 10.1136/bmj.k934.
- [27] RODRIGUEZ C J, SOLIMAN E Z, ALONSO A, et al. Atrial fibrillation incidence and risk factors in relation to race-ethnicity and the population attributable fraction of atrial fibrillation risk factors: the Multi-Ethnic Study of Atherosclerosis [J]. *Ann Epidemiol*, 2015, 25 (2): 71-76, 76.e1.
- [28] STRATH S J, KAMINSKY L A, AINSWORTH B E, et al. Guide to the assessment of physical activity: clinical and research applications: a scientific statement from the American Heart Association [J]. *Circulation*, 2013, 128 (20): 2259-2279.
- [29] STRAZZULLO P, D ' ELIA L, CAIRELLA G, et al. Excess body weight and incidence of stroke: meta-analysis of prospective studies with 2 million participants [J]. *Stroke*, 2010, 41 (5): e418-e426.
- [30] SUN L, CLARKE R, BENNETT D, et al. Causal associations of blood lipids with risk of ischemic stroke and intracerebral hemorrhage in Chinese adults [J]. *Nat Med*, 2019, 25 (4): 569-574.
- [31] TIAN T, JIN G, YU C, et al. Family history and stroke risk in China: evidence from a large cohort study [J]. *J Stroke*, 2017, 19 (2): 188-195.
- [32] VIRANI S S, ALONSO A, APARICIO H J, et al. Heart disease and stroke statistics-2021 update: a report from the American Heart Association [J]. *Circulation*, 2021, 143 (8): e254-e743.
- [33] WANG S, SHEN B, WU M, et al. Effects of socioeconomic status on risk of ischemic stroke: a case-control study in the Guangzhou population [J]. *BMC Public Health*, 2019, 19 (1) [2021-11-13]. <https://doi.org/10.1186/s12889-019-6998-4>.
- [34] WHELTON P K, CAREY R M, ARONOW W S, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines [J]. *Hypertension*, 2018, 71 (6): 1269-1324.
- [35] WOOD A M, KAPTOGE S, BUTTERWORTH A S, et al. Risk thresholds for alcohol consumption: combined analysis of individual-participant data for 599 912 current drinkers in 83 prospective studies [J]. *Lancet*, 2018, 391 (10129): 1513-1523.
- [36] WU S, WU B, LIU M, et al. Stroke in China: advances and challenges in epidemiology, prevention, and management [J]. *Lancet Neurol*, 2019, 18 (4): 394-405.
- [37] 李子孝, 王春娟, 王伊龙, 等. 卒中临床诊疗和疾病管理核心数据元及定义专家共识 [J]. *中国卒中杂志*, 2020, 15 (4): 416-434.

收稿日期: 2021-11-04 修回日期: 2021-11-13 本文编辑: 徐文璐