

· 技术与方法 ·

# 基于医院信息系统的伤害住院病例流行病学分析主题数据库的构建

李顺飞<sup>1</sup>, 刘丽华<sup>2</sup>, 刘建超<sup>2</sup>, 唐晓东<sup>1</sup>, 李申龙<sup>3△</sup>

(1. 中国人民解放军第一五〇医院, 河南洛阳 471031; 2. 中国人民解放军总医院医院管理研究所, 北京 100853; 3. 军事医学科学院疾病预防控制中心, 北京 100071)

**摘要:**目的 探讨基于医院信息系统(HIS)业务数据库的伤害住院病例主题化研究方法。方法 通过体系框架设计、表结构设计、多维数据模型设计和数据抽取集成, 构建了伤害住院病例流行病学分析主题数据库。结果 对 HIS 业务数据的挖掘利用进行了探索, 同时为伤害研究提供了新的方式。结论 将主题数据库引入到伤害医院病例监测工作中有积极意义。

**关键词:** 医院信息系统; 伤害住院病例; 数据库; 流行病学

doi:10.3969/j.issn.1671-8348.2012.04.017

文献标识码: A

文章编号: 1671-8348(2012)04-0357-03

## Theme database constructing of injured inpatients' epidemiological analysis on base of HIS

Li Shunfei<sup>1</sup>, Liu Lihua<sup>2</sup>, Liu Jianchao<sup>2</sup>, Tang Xiaodong<sup>1</sup>, Li Shenlong<sup>3△</sup>

(1. 150 Hospital of PLA, Luoyang, Henan 471031, China; 2. Institute of Hospital Management, General Hospital of PLA, Beijing 100853, China; 3. Institute of Disease Prevention and Control, Academy of Military Medical Sciences, Beijing 100071, China)

**Abstract:** **Objective** To study the method of theme research on injured inpatients based on operation database of HIS. **Methods** Through the design of system framework, table structure, multidimensional data model, and data extraction and integration, the theme database of injured inpatients' epidemiological analysis was constructed. **Results** The excavation and application of operation data were explored and the new way was provided to injury research. **Conclusion** Importing the theme database into the surveillance of injury inpatients has active significance.

**Key words:** hospital information systems; injury inpatients; database; epidemiology

随着医疗卫生水平的提高, 传染病发病率的下降, 伤害的危害及其防治问题日益显露出来。伤害住院病例造成的疾病负担和寿命损失在所有伤害病例中均占有较大比例, 对其进行流行病学研究可以掌握伤害流行特征, 为制订有效的预防干预措施提供重要依据。医院信息系统(HIS)的推广应用积累了海量的业务数据, 然而针对历史数据的应用大多停留在查询和报表的层次上, 缺乏对数据的集成和深层分析。本研究拟通过构建基于 HIS 的伤害住院病例流行病学分析主题数据库, 探讨基于 HIS 业务数据的主题化研究方法。

### 1 体系框架设计

以伤害住院病例流行病学分析为主题, 以数据库和数据挖掘技术为手段, 基于 HIS 业务数据库相关信息, 设计伤害住院病例流行病学分析体系框架。框架包括三部分, (1) 基础数据库: 即 HIS 后台业务数据库, 数据内容包括其中的伤害住院病例流行病学分析相关信息; (2) 主题数据库<sup>[1-3]</sup>: 即面向业务主题的数据组织存储, 其数据结构由规范的基本表构成, 其构建方法是基于基础数据库, 以多维数据模型结构对业务主题信息进行抽取、集成、重构; (3) 数据挖掘利用: 即基于主题数据库进行的决策分析。体系框架见图 1。

### 2 表结构设计及与 HIS 后台的对应

通过定义伤害住院病例流行病学分析主题数据库的表结构, 集成 HIS 业务数据库中相关信息, 在此数据环境中, 根据不同的研究需要, 定义其数据模型, 以便进行分析和数据挖掘。数据库的字段设置以满足流行病学分析为最基本要求<sup>[4-5]</sup>, 为辅助分析性研究增加了诊断、费用和伤害特征等信息; 字段名称取 HIS 中相应名称, 便于理解和应用。经过反复论证, 主题数据库共设置数据项 43 个, 包括患者基本信息、住院信息、诊断信息和费用信息等, 数据主要来源于患者主索引、患者住院主

记录、诊断记录、诊断分类记录、住院患者费用记录等, 其与 HIS 业务数据的对应关系见表 1。

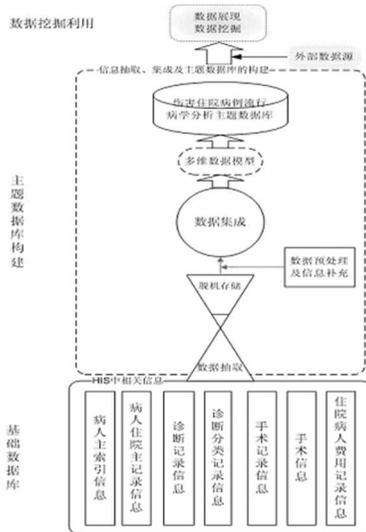


图 1 伤害住院病例流行病学分析体系框架

### 3 多维数据模型

本研究的主题是伤害住院病例流行病学分析, 可将此数据库模型设计为星形模型<sup>[6-8]</sup>。星形架构是管理关系数据以进行多维使用的最常用方式, 比较适合对单个主题建模。星形模型的数据库通常包括一个大的包含大批数据并且不含冗余的事实表(中心表)和一组小的维表(附属表), 星形结构中的各个维表同事实表以维标识符互联, 形成一个多维数据表, 然后在此基础上进行各角度的预计算, 将计算结构存储形成多维数据库。

△ 通讯作者, Tel: (010) 66948391; E-mail: lishenlong@sohu.com.

表 1 伤害住院病例主题数据库的数据项与 HIS 业务数据的对应关系

序号	字段中文名称	字段名	类型	长度	HIS 对应(表名. 字段名)	采集系统
1	患者标识	PATIENT_ID	C	10	PAT_VISIT. PATIENT_ID	身份登记
2	姓名	NAME	C	8	PAT_MASTER_INDEX. NAME	身份登记
3	性别	SEX	C	4	PAT_MASTER_INDEX. SEX	身份登记
4	出生日期	DATE_OF_BIRTH	D	-	PAT_MASTER_INDEX. DATE_OF_BIRTH	身份登记
5	入院时间	ADMISSION_DATE_TIME	D	-	PAT_VISIT. ADMISSION_DATE_TIME	住院登记
6	出院时间	DISCHARGE_DATE_TIME	D	-	PAT_VISIT. DISCHARGE_DATE_TIME	病房分系统
7	职业	OCCUPATION	C	1	PAT_VISIT. OCCUPATION	住院登记
8	入院方式	PATIENT_CLASS	C	1	PAT_VISIT. PATIENT_CLASS	住院登记
9	入院病情	PAT_ADM_CONDITION	C	1	PAT_VISIT. PAT_ADM_CONDITION	住院登记
⋮	⋮	⋮	⋮	⋮	⋮	⋮
38	出院诊断名称	DIAGNOSE_DIS	C	80	DIAGNOSIS. DIAGNOSIS_DESC	病房分系统
39	治疗结果	DIAGNOSE_RESULT	C	4	DIAGNOSIS. TREAT_RESULT	病房分系统
40	损伤和中毒原因	MAR_TOXI	C	80	DIAGNOSIS. DIAGNOSIS_DESC	病房分系统
41	总费用	COSTS_SUM	N	10,2	MEDICAL_COSTS. COSTS	收费分系统
42	伤害性质	SHXZ	C	16	无	手工补填
43	伤害部位	SHBW	C	16	无	手工补填

- :表示无此项。

表 2 维度和粒度设计

维度	粒度
入院时间	年、月
年龄	<15、15~<25、25~<35、35~<45、45~<55、55~<65、65~
性别	男、女
身份	地方、离退休干部、在职干部、士官、士兵、学员、免减费家属
入院方式	门诊、急诊、转入
入院病情	危、急、一般
治疗结果	治愈、好转、无效、死亡
伤害亚类	上肢骨折、下肢骨折、脱臼等 23 种
伤害病种	脑震荡、中暑、掌骨骨折等 277 种
伤害原因诊断	军训中意外损伤、滑到或摔倒、在体育运动中损伤等 71 种
伤害性质	骨折、扭伤/拉伤、挫伤/擦伤、锐器伤/咬伤/开放伤、器官系统损伤、脑震荡/脑挫裂伤、烧烫伤、其他
伤害部位	上肢、头部、下肢、躯干、中毒窒息触电冻伤等、消化系统、神经系统、呼吸系统

3.1 定义事实 伤害住院病例流行病学分析多维数据模型的事实是对伤害住院病例进行流行病学分析的基本指标。定义的指标包括人次、医疗费用及住院日。

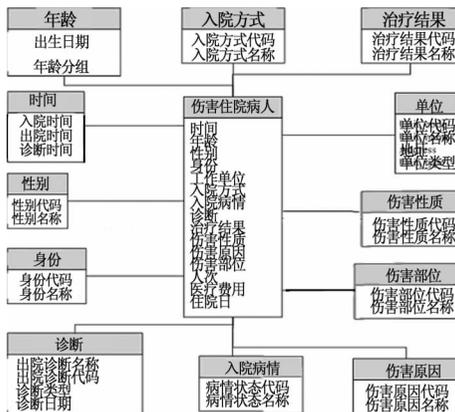


图 2 伤害住院病例流行病学分析多维数据模型

3.2 维度和粒度设计 在设计维度表模型的过程中考虑到维度模型是用户分析数据的窗口,维度的设计提供了维度属性的定义,一个对象的维度包含大量的属性。这些属性应具有这样一些特征,即可用文字描述、离散值、有规定限制以及在分析过程中可以提供标题。本研究的维度和粒度具体设计见表 2。

3.3 多维数据模型设计 以伤害住院病例流行病学分析为主题的星形模式多维模型见图 2,包含了一个中心事实表(伤害

住院患者)、12 个维表(时间、年龄、性别等)和 3 个度量(人次、医疗费用、住院日)。

4 数据抽取集成

本研究利用 Kettle 进行数据抽取集成,Kettle 是 Extract-Transform Load 的缩写。ETL 工具集,被设计用来实现 ETL 需要,即抽取、转换、装入和加载数据;具有易于生成、易于发布、易于维护,从模型直接执行任务的特点。数据抽取集成的具体步骤包括:构建数据流,配置数据库参数,获取数据,选择字段,将抽取的数据插入目标数据库。

5 数据预处理

HIS 中的医疗数据存在多样性、不完整性、不一致性、冗余性,甚至可能包含错误或存在偏离期望的离群数据。因此,数据预处理是必不可少的环节。本研究主要采用了数据清理、数据变换和调查补充等方法。

5.1 数据清理 数据清理的处理方法包括缺失数据调查填补、异常值的筛检及处理。通过对数据进行逻辑校验,对其中的逻辑错误进行标注,后核实修正;对于有标准字典的数据项直接用标准字典的相应内容更新;对影响研究病例流行特征的异常病例进行剔除,剔除了以伤害诊断进行住院治疗的非真实伤害病例。

5.2 数据变换 数据变换是将数据转换或统一成适合于分析的形式,主要是找到数据的特征表示,用转换方法挖掘变量的有效信息。如将入院时间转换为年、月,以便进行趋势分析;将

实际年龄转换为设置的年龄组以进行年龄特征描述。

**5.3 数据调查补充** 数据调查补充主要针对伤害特征属性数据,当前 HIS 中相关信息只在出院诊断、损伤与中毒原因诊断中有所体现,但满足不了伤害流行病学分析的需要。因此,本研究尝试通过对照 HIS 中的诊断信息、电子病历记录的方法,参考《伤害监测指南》和我国“医院伤害监测报告卡”的相关分类内容,逐个提炼并标准化病例的伤害性质、伤害部位信息。如出院诊断是“股骨干骨折”,提炼标准化后的伤害性质是“骨折”,伤害部位为“下肢”。

## 6 讨 论

该研究旨在提供一种挖掘现有业务历史资料对伤害类病例进行深入分析的方法。之前,国家和军队也开展了许多相关的研究<sup>[9-10]</sup>,其方法主要是通过进行现场问卷调查、医院急诊监测和住院病历信息整理再处理等方法,信息采集耗费人力、物力。HIS 业务数据库中信息的电子化存储方式便于加工整理、易于获得,且涵盖了医疗费用、住院时间等医疗资源消耗方面的信息,为开展伤害研究提供了一个新的思路。然而,由于当前的 HIS 设计初始没有考虑伤害监测功能,加之各单位 HIS 的应用情况不尽相同,致使信息有选择性偏移,这需要通过系统的改造与重建来完善。

本研究运用数据挖掘的有关思想和技术,立足于“军卫一号”医院信息系统,通过体系框架设计、表结构设计、多维数据模型设计和数据抽取集成等过程构建了伤害住院病例流行病学分析主题数据库。将主题数据库的应用引入到伤害住院病例监测工作中有其积极意义,既扩展了 HIS 业务数据的应用模式,也提高了伤害监测工作的效率。目前,国内已有解放军总医院、重庆医科大学等多家单位对数据挖掘技术在医院信息利用方面做了积极探索与实践<sup>[11-13]</sup>,但未发现利用数据挖掘技术对“伤害住院病例流行病学分析”进行研究的相关文献。本研究通过相关业务诊疗信息提取伤害特征信息的做法,为伤害特征研究提供了又一个思路,但其科学性需要进一步论证。通过利用本研究的方法对某综合性医院 HIS 数据库构建伤害住院病例流行病学分析主题数据库的实例分析,发现构造基于业务主题的星形模式多维模型,可以对业务数据进行多层次多

角度的分析,但数据挖掘方法和展现方式需要科学化、多样化。

## 参考文献:

- [1] Inmon WH. 数据仓库[M]. 王志海 译. 北京:机械工业出版社,2001.
- [2] 周炎涛,谢东,吴正国. 主题数据库技术的亲合性分析方法研究[J]. 计算机工程与应用,2005(12):168-170.
- [3] 张国庆,曹顺良,方焯,等. 基于构件的生命科学主题数据库构建方法[J]. 计算机应用研究,2007,24(6):12-14.
- [4] 刘丽华,曹秀堂,郝璐. 住院患者主题数据集统计集成与应用设计[J]. 中国医院,2007,11(12):8-12.
- [5] 刘玉,陈金雄. 病种分析主题数据仓库的建立[J]. 医学信息,2008,21(12):2141-2145.
- [6] Wong ST, Hoo KS Jr, Knowlton RC, et al. Design and applications of a multimodality image data warehouse framework[J]. J Am Med Inform Assoc,2002,9(3):239-254.
- [7] Han JW, Kamber M. 数据挖掘概念与技术[M]. 范明,孟小峰 译. 北京:机械工业出版社,2008:70-81.
- [8] 蒋伟进,唐代喜,刘青宝. 面向主题的关系——多维数据模型建立与数据集市[J]. 微机发展,2003,13(6):74-76.
- [9] 李恣,李丽萍. 3 622 例道路交通伤害住院患者的疗效和转归分析[J]. 疾病控制杂志,2007,11(4):383-385.
- [10] 沈敏,刘筱娴. 1 444 例女性伤害患者住院时间的影响因素分析[J]. 中国医院统计,2005,12(2):115-117.
- [11] 曹秀堂,郝璐,刘丽华. ODS 统计信息资源库的功能与应用[J]. 中国医院,2008,12(12):12-14.
- [12] 刘丽华,李昕,胡凯. 统计信息集成与管理决策支持系统概述[J]. 中国医院,2008,12(12):8-11.
- [13] 易静,苏新良,王润华. 决策树在乳腺癌高位淋巴结转移判别诊断中的应用[J]. 重庆医科大学学报,2009,34(8):606-609.

(收稿日期:2011-01-18 修回日期:2011-07-20)

(上接第 356 页)

清中却几乎没有 TNF- $\alpha$ 、IL-1 $\beta$ 、M-CSF 的表达,故得出结论,单核细胞和内皮细胞之间的相互作用并非单纯的物理性接触,但其具体的刺激机制还需进一步验证。TNF- $\alpha$  是维持和促进 DC 发育的最基本因子,并且可以阻止粒系细胞的分化,在诱导 DC 分化的过程也发挥举足轻重的作用。DC 的迁移能力是发挥其免疫功能的基础,在 DC 的趋化过程中发挥主要作用的是 CC 型趋化因子。DC 分化的不同阶段表达不同的趋化因子受体。成熟 DC 通过 CCR7 与 MIP-3 $\beta$ (CCL19),淋巴组织趋化因子(SLC<sup>[CCL21]</sup>)等的结合调节自身运动<sup>[10]</sup>。人脐静脉内皮表面有单核细胞趋化因子(MCP-1)表达,它与 DC 上的趋化因子受体 CCR2 结合,参与细胞的伸展与移出,介导了单核细胞穿过人脐静脉内皮细胞的运动。

本实验结果显示,用人原代脐静脉内皮细胞诱导脐血单核细胞,可以分化为具有典型 DC 形态、表型和功能的 DC,诱导后细胞表型标志 CD83、CD40、CD80、HLA-DR、CD1a 表达均显著增高,负载肿瘤细胞冻融抗原负载后,DC 各项表面标志表达进一步增加,且对肿瘤细胞的特异性杀伤活性比负载抗原前增加更为明显,说明负载肿瘤抗原后的成熟 DC 活化 T 细胞能力更强,是治疗肿瘤的主要 DC 来源。

脐静脉内皮细胞之所以能诱导 DC 形成,其机制可能与 HUVEC 膜表面表达的多种膜受体和分泌多种细胞因子有关,这种诱导 DC 形成的方法为 DC 的产生途径提供了新思路,怎样更充分的发挥其临床应用价值,尚需进一步深入探讨。

## 参考文献:

- [1] 祝和成,吴尚辉,黄柏英. 人 DC 与黑色素瘤细胞融合疫苗体外诱导特异性抗肿瘤 CTL[J]. 中南大学学报,2008,33(11):1672-7347.
- [2] 欧阳清,魏玉英,金伯泉. 肿瘤疫苗的作用机制及研究进展[J]. 第四军医大学学报,2008,29(19):1821-1824.
- [3] 袁婷婷,刘艳荣. 树突状细胞生物学研究进展[J]. 中国实验血液学杂志,2010,18(4):1074-1078.
- [4] 羨鲜,王永祥,孙树汗. 以树突状细胞为基础的肿瘤疫苗设计策略的研究进展[J]. 临床荟萃,2010,25(2):179-182.
- [5] 朱传东,郑勤,张金安. 树突状细胞在肿瘤免疫治疗中的应用与进展[J]. 实用癌症杂志,2011,26(1):108-110.
- [6] Wu YG, Wang L, Zhang YY. Dendritic cells as vectors for immunotherapy of tumor and application (下转第 361 页)