



# 中药上市后再评价 HIS“真实世界”集成数据仓库的设计方法探讨

庄严<sup>1</sup>, 谢邦铁<sup>1</sup>, 翁盛鑫<sup>1</sup>, 谢雁鸣<sup>2\*</sup>

(1. 中国人民解放军海军总医院, 北京 100048;

2. 中国中医科学院, 北京 100700)

**[摘要]** **目的:**探讨利用医院信息系统数据的集成构建涵盖广泛的用于药物上市后再评价的大型数据库系统的可行性和必要性,为中药上市后再评价 HIS(hospital information system)集成数据仓库的构建提供总体设计思路和方法。**方法:**通过对国内外利用电子信息系统开展基于真实世界的临床试验设计的分析和对比,结合我国医院信息系统的特性,从设计思路、设计特点、存在问题和解决办法等几个方面对总体构架进行设计和探讨。**结果:**提出了中药上市后再评价 HIS 集成数据仓库的设计方案,指明了建设的途径和方法。**结论:**提出的中药上市后再评价 HIS 集成数据仓库的设计方案具有高内聚低耦合、安全、通用、高效、可维护性好的特点,可以有效解决多家医院 HIS 数据集成过程中的问题和挑战,具有实用价值。

**[关键词]** 真实世界;医院信息系统;数据仓库;数据挖掘;抽取转换加载;可操作数据存储

虽然随机对照试验(randomized control trials, RCTs)是评价未上市新药疗效的金标准,但在解决中药上市后安全性和有效性再评价问题上存在不足。在广大人群中开展真实世界研究(real world study, RWS)可以较真实地收集药品安全性和有效性相关信息,从而为评价药品的受益/风险及采取相应措施提供重要依据<sup>[1]</sup>。RWS 是开展上市后中药临床再评价研究的一种新理念,必将在中药上市后再评价实践工作中得到充分的应用和检验。

近年来,随着对临床试验设计方法的逐渐深入探索和医疗实践的迫切需要,RWS 受到越来越多医学研究人员的关注。在欧洲和北美,研究人员和相关组织开展了了一些 RWS,这些研究在设计上不尽相同,包括观察性设计、横断面设计和队列设计等,其中以观察性设计为主<sup>[2]</sup>。推荐等级的评估、制定与评价(the grading of recommendations assessment, development and evaluation, GRADE)标准认为在强调随机对照试验的同时,观察性研究也具有其重要作用。在 GRADE 中,如果观察性研究中某干预措施疗效显著,证据显示存在剂量效应关系,存在各种可能导致疗效显著性降低的偏倚时,观察性研究证据的等级将可能提高<sup>[3]</sup>。然而,在医学研

究中 RWS 也并非完美,在实际研究中需要大样本量的数据和较长的随访时间,开展 RWS 的成本会很高。目前,随着医疗信息技术的发展,利用医疗电子信息系统开展 RWS 的研究越来越多。国际上,电子数据获取(electronic data capture, EDC),临床试验管理系统(clinical trial management system, CTMS),交互语音和网络应答及网络门户系统(interactive voice and web response system and web-based portals, IVRS-IWRS)等技术层出不穷,Gale C P 等人利用英国心肌梗死国家监测数据库(myocardial infarction national audit project database, MINAP)在真实世界人群中评价 ST 段抬高心肌梗死(ST-elevation myocardial infarction, STEMI)患者院内死亡率的预测因素是比较有代表性的研究案例<sup>[4]</sup>。在国内,上海市药品不良反应监测中心首次将 RWS 方法应用于中药注射剂安全性的研究中,但国内尚缺乏涵盖广泛的用于药物上市后再评价的大型数据库系统。使用临床电子数据库进行基于 RWS 的中药上市后再评价研究,可以很容易获得巨大样本量的数据,并能有效降低成本,将具有广阔的发展空间和研究价值。

医院信息系统(hospital information system, HIS)作为最主要的临床电子数据库已经积累了大量的临床数据,具有较高的研究价值<sup>[5]</sup>。上市中药 HIS 集成数据仓库设计采用真实世界的方法,利用信息采集系统和数据仓库技术,开展以上市中药再评价为基础的,基于真实世界的大样本的观察性研究设计,采集多家综合性医院 HIS 中的中药使用信息和相关患者信息,整合构建统一的数据仓库。上市中药 HIS 集成数据仓库可以为以中药上市后的适应症、剂量疗程、联合用药、合并疾病以及不良反应等为主要内容的临床再评价研究提供数据支撑,为将大样本的观察性研究(真实世界)应用于

**[稿件编号]** 20110806034

**[基金项目]** 国家“重大新药创制”科技重大专项(2009ZX09502-030)

**[通信作者]** \*谢雁鸣,研究员,博士生导师,研究方向为中医临床评价方法研究, Tel: (010) 64014411-3302, E-mail: zhinanb2010@ yahoo.com.cn

**[作者简介]** 庄严,硕士,工程师,研究方向为数据仓库与系统开发, Tel: 15910790248, E-mail: joyear2008@ yahoo.com.cn

药物上市后评价、建立符合中药临床观察特点的数据库提供范例,同时,所形成的研究结果和临床应用可以进一步为临床安全、合理、有效使用中药提供依据,为医院医生临床使用中药提供帮助,为企业评价药品疗效和安全性提供方法,为国家制定相关药品政策提供参考<sup>[6]</sup>。

### 1 系统总体设计

HIS 集成数据库主要围绕患者住院主记录、患者住院医嘱、患者疾病诊断以及检验指标和经济指标几个主题进行构建。将多家医院的数据整合成统一的数据仓库,关键在于将数据的提取过程工程化,将提取数据的结构和内容标准化与规范化。其总体设计思路基于可复用构件思想,采用面向多元异构数据集成的抽取、转换和加载(extract, transform, load, ETL)技术,将来自于各 HIS 的底层数据库离散标准化、顶层设计视图统一标准化,使用标准接口和可复用思想构建 ETL 过程,并将 ETL 工具与标准 SQL 语句相结合的方法完成 HIS 数据的集中和海量数据仓库的构建。

为了得到高质量标准化的数据,必须对抽取出来的原始数据做一系列复杂转换处理,最后装载到数据仓库中。实现 ETL 过程的效率和质量很大程度上决定了数据仓库系统的构建效率和质量。目前研究 ETL 过程都是集中于个案的研究,强调 ETL 系统的可扩展性和灵活性,对于如何创建可复用的标准化的 ETL 过程的研究则很少,如何在一系列相似或相近的 ETL 过程中发现其共同特征、知识和需求,使得 ETL 过程可以在这些项目中被反复使用而无需修改或少量修改,大幅度提高实现 ETL 过程的效率,从而提高数据仓库构建的效率,研究该问题具有一定理论意义及实用价值。正是基于这种考虑,在 HIS 数据的提取与标准化过程中使用的设计思路见图 1。

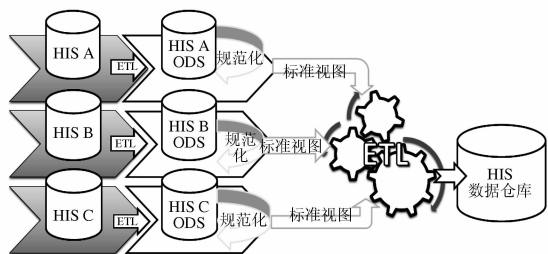


图 1 HIS 数据的提取与标准化总体设计思路示意图

总体设计思路是以构建各家 HIS 的可操作性数据存储(operational data store, ODS)数据库为纽带,连接 2 个 ETL 过程,实现从分散的各家医院的 HIS 数据到集成统一的数据仓库的转换。医院 HIS 数据通过 ETL 过程进入 ODS 之中,由于 HIS 的异构性,此步骤的 ETL 过程需要分别开发。ODS 为各医院分别建立数据库,并在各自的数据库中针对各维度表与事实表进行结构和数据的标准化并构建标准视图,使得各 HIS 数据在进入数据仓库构建的 ETL 过程之前已经具有相

同的关联结构和维度表与事实表的数据字典,这样就可以保证新加入 HIS 数据或已有 HIS 数据的更新对 ETL 过程以及以后的数据仓库处理不产生影响,进而实现 ETL 过程的复用。然后将形成的标准视图整合后通过第 2 个 ETL 过程形成统一的 HIS 数据仓库以供进一步处理和科学研究的需要。

### 2 系统设计特点

**2.1 高内聚度与安全性** HIS 数据在进行标准化的过程中,需要得到相关业务人员的协作。可能需要将数据库提供给相关业务人员在其所在医院完成标准化,那么其他 HIS 数据的安全性成为一个需要解决的问题。将数据及其结构的标准化放在各自的 ODS 来执行,使医院个性化的数据分散到各自的数据库,在提高内聚度的同时保证了其他 HIS 的数据的安全性得到保障。

**2.2 低耦合度与通用性** 在各自 ODS 进行标准化后的数据通过标准化视图的方式形成统一的数据源进入到数据仓库构建的 ETL 过程中,标准视图由于构建在结构和数据标准化之后,其代码可以重用,通过极少量的改动即可实现新加入 HIS 系统的处理,而对于不同 HIS 来说标准视图以后的过程更是无需调整,因此具有低耦合度和很好的通用性。

**2.3 复用构件与高效性** 为了使 ETL 的维护以及开发成本最低,笔者将不同 HIS 系统里面的相应的数据逻辑存储结构采用统一的标准结构,这样使得所有的 HIS ODS 到 HIS 数据仓库的 ETL 都是同一份 ETL 过程,大大降低了 ETL 的开发及维护成本,提高了复用性;同时对于 ETL 工具效率低下的部分采用 SQL 语句替代的方法,提高了 ETL 过程的效率。

**2.4 可维护性** 提供了专门的程序对数据集中的各步骤进行创建和维护。标准化视图和可重用 ETL 过程的设计使系统后端基本固定,开发维护成本及其风险降到最低。同时只需提供标准结构,不同的 ODS 便可结合相应的开发人员及业务人员进行数据及结构的标准化,并且在 ODS 的视图层完成,对开发人员和维护人员的能力要求较低。

### 3 讨论

由于采集自临床实时发生的数据具有更真实、更自然的特性,数据量大且成本低,因此基于 HIS 数据的中药上市后再评价和中药分析数据仓库的构建,可以从另一个角度完善中药上市后再评价体系,为中药上市后再评价提供新的分析手段,有利于推动中药上市后再评价的进程,具有广阔的研究空间。上市中成药在 hospital 的应用进行分析和数据挖掘,首先要构建上市中成药的数据仓库,数据量越大单个上市中成药病例数越多挖掘出的信息价值越高。单一医院的上市中成药数据量有限,必须多家医院数据集中在一起形成较大的数据仓库才有较高的数据挖掘价值。理想的数据来源是所有 HIS 数据来自于中医医院,这样中医的元素比较丰富,但一个区域中医医院数量明显少于综合性医院数量,数据来源有限,要在全中国范围内采集中医医院的 HIS 数据,时间和费用成本会很高,综合性医院中也大量使用上市中成药,纳



入西医医院的 HIS 数据会大量增加数据来源。虽然我国医院信息化建设起步不晚,整体建设水平也不低,但因为经费原因、对医院信息化整体建设的认知原因等,我国的医院信息化建设水平亦参差不齐,还有行政管理方面的原因等,要想获得合格的数据需要对提供数据的医院进行沟通、数据质量进行实地考察,以免取不到数据或数据不合格达不到分析要求。

基于 HIS 数据的中药上市后再评价和数据仓库构建的重点在于规范化,在于如何更好地统一多家医院 HIS 数据的数据规范。HIS 的规范化包含术语规范化、代码规范化、接口规范化、数据交换规范化等。要实现多家医院数据集中、数据共享,必须使用统一的规范,并执行实际可操作的规范化推行办法,这是真正实现医疗信息资源数据集成共享的关键前提。规范化过程的设计与建立是 HIS 集成数据仓库建设的难点和重点。数据仓库的建设涉及的多家 HIS 由不同 HIS 开发商建设,有些医院 HIS 的不同模块还由几家开发商独立建设,数据没有集中在一起,需要分别进行提取,再整合成完整的 HIS 数据;有些医院 HIS 没有提供数据结构等相关文档,只能通过具体数据反复研究理解其含义。各个医院的 HIS 设计方法不一样,数据结构不一样,录入数据的医疗业务员的素质也不一样,这些都对规范化过程的设计有很大的

影响,是需要具体考虑的问题,其复杂性程度不亚于区域医疗数据共享,需要构建一个统一的标准化平台并制定完善的标准化办法。本文提出的提出的设计方案能初步满足研究的需要,是建设中药上市后再评价 HIS 集成数据仓库的有益探索,其设计方法还需要进一步论证和探讨。

#### [参考文献]

- [1] 谢雁鸣,毛平,田峰. 真实世界研究在中药上市后临床再评价中应用前景的探讨[J]. 中药新药与临床药理, 2010, 21(3): 324.
- [2] Cazzoletti L, Marcon A, Janson C, et al. Asthma control in Europe: a real-world evaluation based on an international population-based study[J]. J Allergy Clin Immunol, 2007, 120(6): 1360.
- [3] 陈耀龙,李幼平,杜亮,等. 医学研究中证据分级和推荐强度的演进[J]. 中国循证医学杂志, 2008, 8(2): 503.
- [4] Gale C P, Manda S O M, Batin P D, et al. Predictors of in-hospital mortality for patients admitted with ST-elevation myocardial infarction: a real-world study using the myocardial infarction national audit project (MINAP) database[J]. Heart, 2008, 94: 1407.
- [5] 杨长明,张晓伟,王徐东. 我国 HIS 的发展及未来趋势[J]. 经济师, 2010(4): 223.
- [6] 王永炎,吕爱平,谢雁鸣. 中药上市后临床评价关键技术[M]. 北京:人民卫生出版社, 2010: 193.

## Designs and thoughts of real world integrated data warehouse from HIS on re-evaluation of post-marketing traditional Chinese medicine

ZHUANG Yan<sup>1</sup>, XIE Bangtie<sup>1</sup>, WENG Shengxin<sup>1</sup>, XIE Yanming<sup>2\*</sup>

(1. PLA Navy General Hospital, Beijing 100048, China;

2. China Academy of Chinese Medical Sciences, Beijing 100700, China)

[Abstract] **Objective:** To discuss the feasibility and necessity of using HIS data integration to build large data warehouse system which is extensively used on re-evaluation of post-marketing traditional Chinese medicine, and to provide the thought and method of the overall design for it. **Method:** With domestic and overseas analysis and comparison on clinical experiments' design based on real world using electronic information system, and with characteristics of HIS in China, a general framework was designed and discussed which refers to design thought, design characteristics, existing problems and solutions and so on. **Result:** A design scheme of HIS data warehouse on re-evaluation of post-marketing traditional Chinese medicine was presented. **Conclusion:** The design scheme was proved to be high coherence and low coupling, safe, Universal, efficient and easy to maintain, which can effectively solve the problems many hospitals have faced during the process of HIS data integration.

[Key words] RWS; HIS; data warehouse; data mining; ETL; ODS

doi:10.4268/cjcm20112033

[责任编辑 马超一]