

临床数据中心建设助力转化医学研究

罗晶,李劲松,黄丽丽,周天舒,翁盛鑫,田宇

[摘要] 建立融合生物样本库信息管理、临床研究设计、生物统计及临床数据分析功能于一体的临床数据中心,为转化医学研究提供信息支持服务。在所有临床数据互联互通的前提下,采用符合国际标准方式对各系统进行集成,实现统一界面集成展示,建立集实验室科研样本数据及门诊、住院、体检、随访业务数据于一体的临床数据中心。在此基础上,建立面向不同临床专科需求的专病数据库,运用各种数据挖掘技术,为临床科研提供信息支持服务。通过临床数据分析,可跟踪疾病、症状、治疗过程及康复情况,对改善医疗质量、效率,建设健康档案管理,改善医疗服务合作有很重要的意义。

[关键词] 临床数据中心;转化医学;生物样本库;数据挖掘

[中图分类号] R197.324

[文献标志码] A

[文章编号] 2095-3097(2013)02-0106-03

doi: 10.3969/j.issn.2095-3097.2013.02.014

在医院里,不同岗位的工作人员在日常业务的应用系统中生产大量数据。既往对这些数据的应用主要围绕医疗活动展开,随着我院建立一流研究型军种总医院目标的确立,成立临床转化医学中心,将基础研究的成果转化成为患者提供治疗手段作为临床科研发展方向。在此情况下,科研对实验室和临床数据的依赖,使得针对临床数据挖掘的应用需求日益迫切,急需进行融合生物样本库、临床研究设计、生物统计及临床数据分析功能于一体的临床数据中心(clinical data repository, CDR)建设^[1]。

1 现状与问题

转化医学倡导从临床工作中发现和提出问题,由基础研究人员进行深入研究,然后再将基础科研成果转向临床应用,基础与临床科技工作者密切合作,提高医疗总体水平。要从临床工作中发现和提出问题,除了要依赖临床医师的临床实践经验,更需要来自于大量临床实践中取得的数据。经过十几年的医院信息化建设实践积累,我院硬件设施、软件应用和临床数据的积累已经初具规模。数字化技术为我们带来的最大益处就是对数据收集、保存和处理的高效率。这些数据中蕴含许多潜在的未知的规律等着我们去探索和发掘。传统的“数据查询统计”已不能满足当前医学研究的需求,整合临床数据与基础研究数据,在基础研究与临床应用之间架起信息桥梁,是转化医学实现的关键基础^[2]。

各临床科室针对科研需要,进行独立的病例管理系统的建设虽有其内在的驱动力,但各科室重复建设的方法并不可取,无法实现对现有的大量临床数据资源的利用和共享。经主观筛选出的病例,往

往带有研究者固有的研究模式和研究思路,手工二次录入数据无法实现基于大数据量的、隐含未知规律的数据挖掘,影响临床科研有效进行回顾性研究和前瞻性研究,无法实现生物基因样本库中基础研究成果与临床工作更好的结合。为更好利用现有临床数据,将其转化为有用的临床科研信息,必须通过对临床数据中心、专科数据库、科研样本库等数据的综合分析,才能为转化医学研究提供回顾性研究和前瞻性研究的强有力数据支持。

2 建设思路

在临床数据中心的建设上,总体思路是“先整合后专业划分”,即保证所有临床数据从医院层面实现数据互联互通,采用符合国际标准方式对各系统进行集成,实现统一界面集成展示,建立集实验室科研样本数据及门诊、住院、体检、随访业务数据于一体的临床数据中心。在此基础上,建立面向不同临床专科需求的专病数据库,运用各种数据挖掘技术,为临床科研提供信息支持服务。

2.1 关键技术 在关键技术上以数据挖掘优先,采用决策树、关联规则、聚类分析等多种算法,从大量的数据中抽取潜在的、有价值的知识或隐含规律,用于统计、发现、预测。

2.1.1 决策树算法 决策树算法是应用最广的归纳推理算法之一,其基本原理是构造一棵决策树,在树的每个结点通过属性测试来划分事例数据集,结点的每个后续分支都对应于属性的某个可能状态。使用决策树算法创建的模型易于理解,预测效率也很高。如以“餐后2h血糖”、“空腹血糖”、“年龄”和“体质量”作为各结点属性分类依据,建立“糖尿病类型”分型决策树算法。

2.1.2 关联规则算法 关联规则算法用于从数据集中找出数据项之间相互关联的规则,例如从患者

[基金项目] 国家863课题(2009AA045300)

[作者单位] 100048 北京,海军总医院计算机管理中心(罗晶,黄丽丽,翁盛鑫);310027 浙江 杭州,浙江大学化学工程与生物工程系(李劲松,周天舒,田宇)

临床数据中找出最有可能引发致病因素或筛选具有疾病特异性的基因与临床表型相关,证实并完善疾病特异性的生物标志物等^[2]。关联规则算法的核心是相关性计数,即找出数据集中出现频率高于指定阈值(即支持度)的项集,而后从这些频繁项集中生成关联规则。如果该信任度的值大于预定的阈值,那么就认为关联规则成立。

2.1.3 聚类分析算法 主要用于发现紧密相关的观测值组群,使得与属于不同簇的观测值相比,属于同一簇的观测值相互之间尽可能类似(图1)。聚类分析在医学领域中主要用于DNA分析、医学影像数据自动分析以及多种生理参数监护数据分析、疾病危险因素等方面。如通过将临床检查结果与体征和人类肿瘤组织来源的生物信息学信息结合起来,认识临床症状和体征、疾病发展以及治疗策略,探究与临床检查、病理、生化分析、影像和治疗的关系。

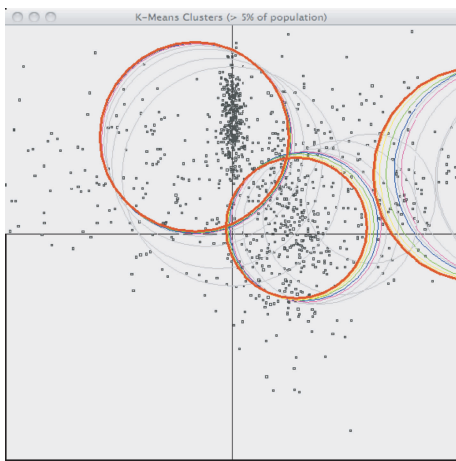


图1 聚类分析

2.2 数据挖掘的展现 采用 Visualization 即可视化技术,实现数据挖掘结果的展现,对于数据挖掘分析中最困难的一部分就是数据的展示。准确解读数据之间的关系,才能清晰有效地传达沟通数据信息。visualizing.org 网站上有几类有用的形式: Chart、Time series、Map、Flow、Matrix、Network、Hierarchy、Info-graphic。具体采用何种形式进行数据展现,还要看数据之间的关系而定^[5]。

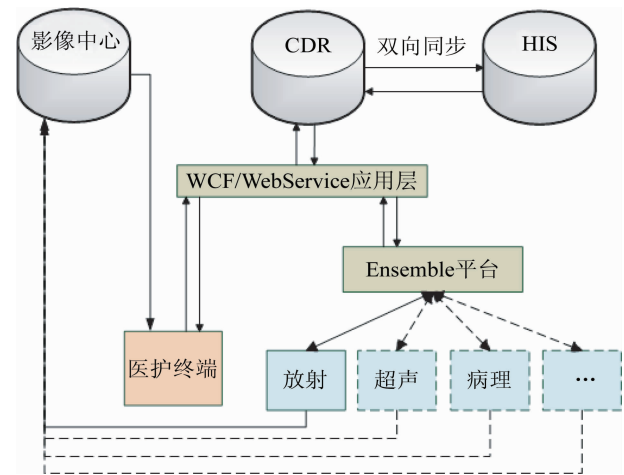
3 关键步骤

3.1 新版电子病历系统部分实现数据整合 目前海军总医院运行的新版电子病历系统(electronic medical record system, EMRS),其最显著的特点之一就是在集成界面集中展现患者病例资料、影像资料、检查检验资料,实现快速浏览,极大地提高了工作效率;特点二是以时间轴连续展现患者体征、临床数据及诊疗过程,各种数据因果对应关系和趋势一目了然,

符合医务人员临床逻辑思维习惯;特点三就是结构化的病历文书格式,便于按需求进行文本检索和数据挖掘。

3.2 集成平台的建设实现临床数据整合 根据医院实际建设情况,为提高数据集成的安全性和可扩展性,在各外围系统与医院信息系统(hospital information system, HIS)之间采用集成平台作为 CDR/HIS 连接外部各种系统之间的主要通道,为业务流程的顺畅进行提供统一的保障。

临床信息集成用于对各临床信息系统进行信息层面的集成事务处理。事务的定义参照集成医疗企业协会制定的规范(integration healthcare enterprise, IHE)执行,消息的交互标准参照卫生信息交换标准(health level 7, HL7) V2.4、影像消息内部传递医学数字成像和通讯(digital imaging and communications in medicine, DICOM)方式执行。集成平台支持不同系统之间的医疗数据整合、业务整合与数据共享,传递和展现整个医疗过程中的相关信息,为临床数据中心的数据来源提供技术基础和保障,对业务系统提供标准的信息交换服务,确保数据交换过程的安全性、可靠性,实现数据在系统平台范围内自由、可靠、可信的交换(图2)。



WCF: windows communication foundation (windows 通讯接口)

图2 集成平台

3.3 生物样本信息库建设 生物样本信息库的建设是将生物样本库与科研电子病例的有机结合,对样本供体、样本本体和样本衍生物进行综合管理,科研电子病例的管理是对样本主体信息的管理。为生物样本库打造的信息共享平台是一个以样本库信息管理系统为核心的、科研电子病例管理系统、随访管理系统等一系列其他系统为辅助的平台^[4]。

结构化的科研电子病例极大地方便了研究人员对生物样本的筛选;科研电子病例的在线功能还在很大程度上减少了不同机构在病例采集上的沟通时

间,改善了工作效率。将临床研究对象入组,进行样本搜集。通过生物信息学和系统生物学,分析样本的基因或蛋白质特征,探索基因和蛋白质间动态网络和相互联系。结合了详细的科研电子病例报告表和丰富的生物样本信息是转化医学研究的坚实基础。

3.4 专病数据库建设 面向临床科研项目管理、数据采集和统计分析的临床专病数据库建设,内容涵盖临床科研中科研项目管理、病例报告表单(case report form, CRF)设计、科研数据采集、查询和导出等主要过程,支持临床科研的回顾性和前瞻性两种类型的研究。实现对整个医疗机构的临床科研的集中统一管理以及单个科研项目的个性化支持;基于规范的临床试验业务流程开发;支持针对不同科研项目的需要,自行定义数据采集 CRF,支持多种数据录入方式和数据质量校验;支持从 HIS、EMRS、实验室信息管理系统(laboratory information management system, LIS)和放射信息系统(radiology information system, RIS)等第三方系统中自动化采集数据,不需用户重复查找和录入,极大地减轻了科研工作量,更能充分利用医院既有信息系统中的数据,避免产生临床科研的“信息孤岛”。针对不同学科制订不同的 CRF,符合目前国内针对重大疾病、慢性病等课题收集样本的习惯,集成流行病学调查、检查检验、随访等多种信息,极大地丰富了样本的附加信息,避免了纸质病例报告表错填、漏填、不规范的缺点,方便进行数据的统计、分析。

2013年美国医疗行业八大信息技术预测中,大数据的临床数据分析和“有意义地”使用数据位列其中^[6]。

通过研究建立临床数据中心重新规划临床研究中数据的获取和应用方式,运用临床分析,医生能发现最普遍的疾病和状况、不同治疗过程的康复率。就医院而言,也为其提供了运用数据发现罹患慢性病如糖尿病、哮喘和高血压患者的方式。临床分析提供的信息帮助更好地应对疾患,降低昂贵的就诊随访费用,为转化医学研究提供强有力的数据支撑,更有助于提升医疗质量。

对医疗机构来说,临床分析都是第一要务,海量数据正开始从研究步入主流。有意义使用数据(meaningful use)是所有医疗机构电子健康档案的应用达到以下要求:改善质量、安全、效率,降低健康差异;将患者和其家庭成员纳入健康管理;改善医疗

服务合作;改善人口与公共健康;维护隐私和安全。按要求,医疗机构须实现让患者在36 h内能够浏览住院信息并下载相关数据。只有真正让这些数据得到最有意义的使用,才能将临床问题与实验室研究更好地结合,从而促进科研成果向临床应用的转化。

【参考文献】

- [1] 时占祥. 洛克菲勒大学医院附属临床和转化科学中心:建设和发展理念[J]. 转化医学杂志, 2012, 1(2): 124-126.
- [2] 刘雷. 生物医学信息学是连接科研与诊疗的桥梁[EB/OL]. (2012-11-16) [2012-12-18]. <http://md.tech-ex.com/2012/interview/22124.html>.
- [3] 吴益飞, 刘雷, 杨远, 等. 建立整合科研电子病例管理的生物样本库信息管理系统[EB/OL]. (2011-09-01) [2012-12-18]. http://md.tech-ex.com/html/2011/article_0901/19678.html.
- [4] 周洁白, 王向东. 肿瘤生物信息学:系统临床医学的新手段[J]. 转化医学杂志, 2012, 1(1): 41-43.
- [5] 佚名. 数据挖掘:用可视化效果展现你的数据[EB/OL]. (2012-12-11) [2012-12-18]. <http://www.chinaz.com/manage/2012/1211/285296.shtml>.
- [6] Rich H. 8 IT Healthcare Trends for Tech Job Seekers in 2013 [EB/OL]. (2013-02-15) [2012-12-18]. <http://www.networkworld.com/topics/it-careers/>.
- [7] 钱志明, 徐海澎. 基于 HL7 V3 建立临床数据中心[J]. 中国数字医学, 2008, 3(5): 67-69.
- [8] 杨静芳, 陈凌, 费晓璐, 等. 利用智能计算技术整合医院临床信息系统[J]. 中国病案, 2012, 13(9): 39-41.
- [9] 龙凤舞. 在医院实现信息化集成平台[J]. 中国现代医学杂志, 2010, 20(6): 938-942.
- [10] 周攀, 王开正. 数据挖掘的研究进展及在临床医学中的应用[J]. 国际检验医学杂志, 2012, 33(10): 1230-1232.
- [11] 张晓祥, 张蕾. 医院数据中心及决策管理平台的建设[J]. 医疗卫生装备, 2012, 33(1): 79-81.
- [12] 王国强, 阚红星, 王宗殿. 综合多种数据挖掘技术的糖尿病诊断系统[J]. 电脑知识与技术, 2012, 8(23): 5547-5551.
- [13] 孟哲. 集成化医院信息资源平台(HRP)的设计与实现[D]. 大连:大连海事大学, 2009.
- [14] 张侃怀, 臧国华, 程远, 等. 浅议医院数字化平台中信息资源的有效利用[J]. 西南国防医药, 2011, 21(11): 1244-1245.

(收稿日期:2012-12-23 本文编辑:徐海琴)