

• 思路与方法 •

基于临床科研信息共享系统开展 中医临床研究的探索

余海滨 符宇 李卓 李真* 王丹妮 胡金亮

(河南中医学院第一附属医院, 河南省郑州市人民路 19 号, 450000)

基金项目: 2012 年国家中医药管理局中医药行业科研专项资助项目(HY12079675)

* 通讯作者: jidiban311@126.com, 13503817158

[摘要] 临床科研信息共享系统是开展真实世界中医临床研究的基础。通过构建临床数据采集、数据仓库、数据挖掘三大平台, 结合河南中医学院第一附属医院重点研究病种艾滋病和慢性阻塞性肺疾病(COPD)的临床研究问题, 充分利用结构化电子病历中的海量数据, 通过智能分析挖掘平台, 在真实世界中开展以艾滋病、COPD 为示范中医临床研究, 实现中医临床医疗科研信息共享。结果形成了“以数据为导向”的真实世界中医临床科研新模式; 且重点病种艾滋病、COPD 已充分利用多种数据挖掘工具开展中医证候分布规律和名老中医辨证用药规律的示范研究。证明基于临床科研信息共享系统在真实世界中利用大规模数据进行临床研究是形成临床决策和理论创新的高效途径。

[关键词] 临床科研信息共享系统; 真实世界研究; 数据采集; 数据仓库; 数据挖掘

“临床科研信息共享系统”是国家中医临床研究基地业务建设重要内容之一, 是应用临床实际诊疗数据开展真实世界中医临床研究的关键技术平台, 是开展真实世界中医临床研究的前提和基础。临床科研信息共享系统可以从不同方面、层次对临床数据进行研究, 使数据转化成新的信息与知识, 使知识进一步变成支持决策的智慧, 是真实世界开展中医临床经验整理、传承和学术创新的有效途径。临床科研信息共享系统和一体化应用技术体系的建立, 可以从根本上解决制约中医学快速发展的技术瓶颈问题, 有利于解决医疗信息孤岛, 实现临床、科研、教学和社会的资源共享。

2010 年 12 月, 河南中医学院第一附属医院在中国中医科学院的指导和支持下, 完善了医院信息系统(hospital information system, HIS)、医院检验管理信息系统(laboratory information management system, LIS)、图像存储与传输系统(picture archiving and communication systems, PACS)等信息系统, 提出以临床医疗数据流的结构化采集为核心, 以数据质量管理体系建设和术语的规范化研究为前提; 利用现代信息技术, 全面采集临床诊疗信息, 形成可分析的中医临床诊疗数据; 利用数据仓库、数据挖掘等方法, 基于海量临床实际数据开展中医证候规律、疗效评价、诊疗方案优化等研究; 不断总结和

验证中医临床诊疗规律, 支持中医理论的传承与创新, 开展真实世界的中医临床研究, 提高中医防病治病的能力和水平。

1 临床科研信息共享系统的构建

围绕中医药防治重大疾病艾滋病和慢性阻塞性肺疾病(COPD)的临床研究支撑体系, 充分运用计算机、数据库、数据挖掘等方法和技术, 利用结构化的电子病历采集系统, 充分采集临床医生大量的诊疗信息, 建立满足中医临床和科研需求的医疗服务平台、数据管理平台和临床研究平台, 形成中医临床科研信息一体化平台, 将从根本上改变中医“从临床中来, 到临床中去”的效率和能力, 从而推动中医个体化诊疗实践和研究水平^[1]。临床科研信息共享系统具体工作流程见图 1。

1.1 建立重点病种临床数据采集规范

制定医院重点病种临床数据采集模板技术规范并建立模板集, 完善其临床术语集并形成疾病术语字典, 形成两个重点病种中医结构化电子病历系统方案; 构建重点病种中医临床数据仓库和中医临床信息知识库和辅助决策系统; 制定临床数据共享标准与规范、临床数据汇交模式与方法; 修订医疗与临床科研信息共享研究的管理制度和质量控制规范; 形成“以数据为导向”的真实世界中医临床科研范式。

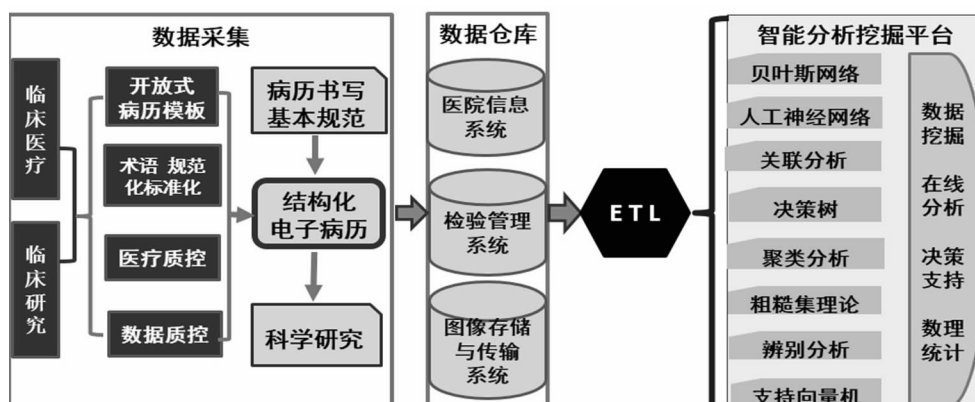


图1 临床科研信息共享系统工作流程图

1.2 构建数据采集平台

中医临床术语标准及其规范是建设中医医院、实现中医临床与科研工作及医院管理工作信息化的基础和保障,是结构化中医临床信息采集和数据挖掘、分析利用的基础^[2]。结合重点病种艾滋病和COPD临床特点、研究需求,梳理归集与本病种相关的疾病的中西医诊断、证候分类、临床分期、疗效评价、用药规范等名词术语,界定术语的内涵与外延,更新重点病种相关的术语字典,形成本地化术语词条,补充到统一的临床标准术语集,不仅提高了临床工作效率,而且为后期的数据存储、转化、挖掘分析奠定坚实的基础。同时根据医疗业务流程和中医辨证论治的临床思维模式,规范重点病种艾滋病、COPD的数据采集过程,进行病历模板的设计,建立符合临床科研需求的病种模板研究。

1.3 构建临床数据仓库

根据数据仓库概念的创始人 W. H. Inmon 的定义,数据仓库就是面向主体的、集成的、稳定的、不同时间的数据集,用于支持经营管理中的决策制定过程^[3]。数据仓库的构建是一个长期、复杂、细致的过程,需要把多个系统的数据按照临床研究主题数据结构进行整理、规则定义,根据主题设计数据模型,通过该模型可以对所研究的内容状况进行全面、清晰的信息描述和统一数据标准。数据仓库的设计以数据模型为基础,而数据模型建立的依据是临床研究的主题。从数据库中抽取适合科研分析的数据,设定抽取和代码转换规则,实现对来源病历数据的筛选、数据仓库数据的自动转换、大规模病历数据到数据仓库的批量导入与汇总处理。本院数据仓库平台的构建采用中国中医科学院的技术框架^[4],并根据本院的病种数据和信息系统特点进行了完善和修改。

1.4 建立智能分析挖掘平台

数据挖掘是近年来随着数据库和人工智能技术的发展而出现的一种全新信息技术,是指从大量的、不完全的、有噪声的、模糊的数据中提取隐含在其中的人们事先不知道的但又潜在有用的信息的过程^[5]。数据挖掘是一个多步骤过程(它需要为数据挖掘算法访问和准备数据),包括挖掘数据、分析结果和采取行动。数据挖掘的目的在于使用所发现的模式帮助解释当前的行为或预测未来的结果^[6]。构建智能分析挖掘平台,进行数据挖掘分析,将医院现有的系统和数据整合为一个数据平台,通过多维度交叉索引技术实现信息的关联。以结构化的电子病历系统为核心,同时将这些结构化信息转化导入数据仓库,通过长期积累形成蕴含着丰富临床经验和临床规律的海量数据,通过智能分析挖掘平台,实现中医临床医疗科研信息共享和开展临床实际信息的科学研究。

2 数据分析与挖掘利用

进行中医药的数据挖掘是中医药现代化研究的重要组成部分^[7],利用结构化电子病历数据进行医学研究已经成为当前国际转化医学研究的切入点和热点,其基本原理是真实世界的临床数据挖掘分析。真实世界的中医临床科研范式,即以人为中心,以数据为导向,以问题为驱动,医疗实践与科学计算交替,从临床中来到临床中去的临床科研一体化的科研范式^[8]。在真实世界中利用大规模数据进行临床研究是形成临床决策和理论创新的高效途径,能够发现中医潜在、隐含的医学规律、诊疗特点。

数据分析与挖掘方案的研究:构建智能化分析挖掘平台,进行数据挖掘分析,将医院现有的系统 HIS、LIS、PACS 导入数据提取、转换和加载(extraction-transformation-loading, ETL),根据临床研

究主题需求,ETL 从不同的数据来源中抽取数据,再进行数据清洗、整合、转换等处理。确定研究主题、准确地定义数据挖掘任务是数据挖掘成功的关键要素;明确临床研究任务与建模工具(Builder)建立数据模型,同时通过增量模型和主次模型的应用来逐步丰富整个数据集;通过分析挖掘工具来分析挖掘数据,对科学决策和临床研究提供信息支持。目前重点病种艾滋病、COPD 临床研究和名老中医医案已充分利用多种数据挖掘工具开展研究。常用的数据挖掘方法主要有贝叶斯网络、人工神经网络、关联分析、决策树、聚类分析、粗糙集理论、判别分析、支持向量机等。

3 重点病种利用临床科研信息共享系统开展临床研究的初步探索

3.1 艾滋病并发胃脘痛证候分布研究

采集艾滋病并发胃脘痛的结构化电子病历 372 份,基于 Crystal Report 软件分析艾滋病并发胃脘痛证候分布。结果显示,胃脘痛辨证分型主要为脾胃虚寒证(频次 98,占 26.1%,下同),肝气犯胃证(74,19.7%),饮食伤胃证(53,14.1%),气滞证(42,11.2%),中焦气虚证(42,11.2%);其次为肝胃阴虚证(35,9.3%),湿热中阻证(14,3.7%),瘀血停胃证(9,2.4%),胃阴亏耗证(5,1.3%)。

3.2 艾滋病并发咳嗽名老中医用药规律研究

采集名老中医治疗艾滋病并发咳嗽的处方 523 个,使用平台中医临床复杂网络分析系统(Liquorice)进行用药规律研究。其核心药物配伍结果如图 2 所示,甘草、款冬花、黄芩、半夏、黄芪、黄连、防风、麻黄、冬瓜仁处于处方配伍网络的核心节点。每一味药物与周围药物的边表示配伍关系。药物节点之间的连接边的权重即药物配伍次数在一定程度上表现了药物之间同时配伍应用的强度。

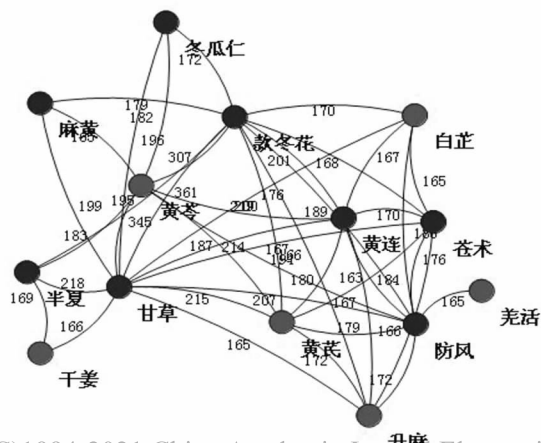


图 2 艾滋病并发咳嗽核心药物及配伍网络图

3.3 COPD 急性加重期证候分布研究

采集 COPD 急性加重期结构化电子病历 5000 份,基于 Crystal Report 软件分析其证候分布情况。结果显示,COPD 急性加重期证候分布主要为痰热壅肺证(频次 1521,占 30.42%,下同),肺肾气虚证(1491,29.82%),痰瘀阻肺证(1273,25.45%),痰浊阻肺证(755,15.11%),肺脾气虚证(417,8.35%),气阴两虚证(358,7.16%),痰湿阻肺证(308,6.16%),痰饮内阻证(219,4.37%)。

4 讨论

随着大数据时代的到来,科学数据的产生和积累呈指数级增长,中医药在几千年防治疾病的过程中积累了丰富的临床经验,形成了许多行之有效的方法和技术,必须主动利用这一变化来进行战略性创新,满足需求,创造未来。随着对临床医学研究方法的逐渐深入和医疗实践的迫切需要,真实世界研究(real-world study,RSW)越来越受到关注^[9]。中医的临床研究往往是症状改善和综合性干预的实际效果,尽管实用性随机对照试验(pragmatic randomized controlled trial,PRCT)的结论被视为制定临床实践指南的金标准,但采用严格设计的 PRCT 来评价中医干预措施的效果往往不太合理,而 PRCT 的结论在实际推广中也存在不足^[10]。RSW 是样本量大、纳入标准相对宽泛、不预定和限设实际临床过程,结合实际情况开展的研究,应用 RSW 的方法进行中医药治疗方案和药物疗效评价是符合中医学发展规律的。临床科研信息共享系统使医疗实践和计算机结合起来,在真实世界中通过海量数据挖掘解决实际问题。

通过应用临床科研信息共享系统,结合重点病种艾滋病和 COPD 的临床研究主题,以临床医疗数据流的结构化采集为核心,以数据质量管理体系建设和术语的规范化研究为前提,利用现代信息技术,全面采集临床诊疗信息,实现以患者为中心的所有临床信息的纵向和横向整合,形成了“以数据为导向”的真实世界中医临床科研新模式。尽管这种新的中医临床科研模式已建立,但要形成一个可持续发展的临床研究创新技术平台,需加强医疗质量管理模块的构建和更新,注重优化后台数据解析和分析模块的功能。重点病种艾滋病、COPD 已利用多种数据挖掘工具开展中医证候分布规律和名老中医辨证用药规律的示范研究。这一平台的实现必将为

(下转第 2109 页)

为佐药;甘草、大枣缓急补气为使药;全方共奏补气活血、疏肝解郁之效。

本研究结果显示,治疗后治疗组对轻度抑郁患者疗效明显优于对照组,对中度抑郁患者疗效与对照组相当,其中对抑郁情绪、睡眠不深、早醒和胃肠道症状的改善尤为突出。治疗组治疗后患者每日服用盐酸舍曲林平均剂量小于治疗前,组间比较小于对照组;治疗组治疗后 NIHSS 评分显效、mRS 残障评级 0~2 级和 BI 评分 ≥ 85 分的病例数均多于对照组。说明神宁方联合西药可以改善脑梗死恢复期伴轻中度抑郁患者的抑郁症状,减少患者每日服用盐酸舍曲林平均剂量,同时改善患者的神经功能,降低其残障评级,提高其日常生活能力。

由于患者抑郁症状改善后服药依从性有所下降,所以本研究取 30 天为试验结点,未能观察到中医药对轻中度抑郁患者的更长期疗效,是本研究的不足,中医药的更长周期使用是否对改善轻中度抑郁患者的症状有益,有待今后进一步观察研究。

参考文献

- [1] Pohjajärvi T, Leppavuori A, Siira I, et al. Frequency and clinical determinants of poststroke depression[J]. Stroke, 1998, 29(11): 2311-2317.
- [2] Kotila M, Numminen H, Waltimo O, et al. Depression after stroke: results of the FINN-STROKE Study [J]. Stroke, 1998, 29(2): 368-372.
- [3] 周洋, 王德刚, 陆征宇, 等. 神宁颗粒治疗缺血性卒中后轻中度抑郁临床研究[J]. 中国中医急症, 2013, 22(3): 439-441.
- [4] 中华医学会全国第四次脑血管病学术会议. 各类脑血管

疾病诊断要点[J]. 中华神经科杂志, 1996, 29(6): 379-380.

- [5] 中华医学会神经科分会. 中国精神障碍分类与诊断标准 (CCMD-3) [M]. 3 版. 济南: 山东科学技术出版社, 2001: 87-89.
- [6] 郑筱萸. 中药新药临床研指导原则 (试行) [M]. 北京: 中国医药科技出版社, 2002: 101.
- [7] Goldstein LB, Bertels C, Davis JN. Interrater reliability of the NIH stroke scale[J]. Arch Neurol, 1989, 46(6): 660-662.
- [8] van Swieten JC, Koudstaal PJ, Visser MC, et al. Inter-observer agreement for the assessment of handicap in stroke patients[J]. Stroke, 1988, 19(5): 604-607.
- [9] Collin C, Wade DT, Davies S, et al. The Barthel ADL index: a reliability study[J]. Int Disabil Stud, 1988, 10(2): 61-63.
- [10] Frühwald S, Löftner H, Baumhackl U, et al. Depression after cerebrovascular injury. Review and differentiation from other psychiatric complications[J]. Fortschr Neurol Psychiatr, 1999, 67(4): 155-162.
- [11] 贾艳滨, 周廷璋. 首发脑卒中后抑郁相关因素的 I 临床研究[J]. 中国神经精神疾病杂志, 1998, 24(2): 66-69.
- [12] Herrmann M, Bartels C, Schumacher M, et al. Post-stroke depression. Is there a pathanatomic correlate for depression in the postacute stage of stroke? [J]. Stroke, 1995, 26(5): 850-856.
- [13] 龙洁, 刘永珍, 蔡焯基, 等. 卒中后抑郁状态的发生率及相关因素研究[J]. 中华神经科杂志, 2001, 34(3): 145-148.

(收稿日期: 2013-08-12; 修回日期: 2013-08-30)

[编辑: 叶晨]

(上接第 2094 页)

中医临床研究提供新思路和新技术, 促进中医学术思想的升华和创新, 推动发掘新的疾病证治规律和深化名老中医经验的内涵和外延。

参考文献

- [1] 宋观礼, 刘保延, 王映辉, 等. 基于中医临床科研信息一体化技术平台的冠心病诊治规律研究[J]. 中华中医药杂志, 2013, 28(5): 1247-1251.
- [2] 刘保延, 周雪忠, 李平, 等. 个体诊疗临床科研信息一体化平台[J]. 中国数字医学, 2008, 2(6): 11-13.
- [3] Inmon WH. 数据仓库[M]. 王海志, 译. 北京: 机械工业出版社, 2006: 27-58.
- [4] Zhou X, Chen S, Liu B, et al. Development of traditional Chinese medicine clinical data warehouse for medical knowledge discovery and decision support[J]. Artif In-

tell Med, 2010, 48(2-3): 139-152.

- [5] 胡波. 并行计算机技术与数据挖掘[J]. 计算机与网络, 2009, 35(8): 65-68.
- [6] 章兢, 张刚. 数据挖掘算法及其工程应用 [M]. 北京: 机械工业出版社, 2006: 35-45.
- [7] 乔延江. 中药 (复方) KDD 复方研究开发的意义[J]. 北京中医药大学学报, 1998, 21(3): 15-17.
- [8] 刘保延. 真实世界的中医临床科研范式[J]. 中医杂志, 2013, 54(3): 451-455.
- [9] 谢雁鸣, 毛平, 田峰, 等. 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(5): 324-327.
- [10] 田峰, 谢雁鸣. 真实世界研究: 中医干预措施效果评价的新理念[J]. 中西医结合学报, 2010, 4(8): 301-306.

(收稿日期: 2013-08-12; 修回日期: 2013-10-08)

[编辑: 崔京艳]