

数据分析与挖掘技术在医疗质量管理中的应用

余元龙

郭 茜

(中山市人民医院 中山 528400) (美国泓维软件有限公司广州分公司 广州 510145)

〔摘要〕 结合中山大学附属中山医院实践, 介绍医院在医疗质量管理中的困难, 从数据分析与挖掘技术角度来探索如何帮助医院提高医疗质量管理水平, 阐明系统设计过程, 包括技术基础、数据基础等方面。

〔关键词〕 医疗质量管理; 多维数据分析; 指标监控; 单病种质量管理; 手术质量管理

Application of Data Analysis and Mining Technology in Medical Quality Management YU Yuan-long, People's Hospital in Zhongshan City, Zhongshan 528400, China; GUO Qian, Dimensional Insight, Inc., Guangzhou 510145, China

〔Abstract〕 Combining with the experiences of Zhongshan Hospital Affiliated to Sun Yat-sen University, the paper introduces the difficulty of hospital in medical quality management, explores how to help hospitals to improve medical quality management level from the perspectives of data analysis and mining. It also elaborates the system design process, including technology basis and data basis, etc.

〔Keywords〕 Medical care quality management; Multi-dimensional data analysis; KPI monitoring; Single disease quality management; Operation quality management

医疗质量与广大患者的生命健康息息相关, 是一个医院的立足之本, 提高医疗质量、加强医疗质量管理是医院工作的核心和永恒的主题, 是医院自身发展的需求, 也是开展各项工作的出发点。国家卫生部 2008 年提出的“以病人为中心, 以提高医疗服务质量为主题”将医疗质量管理要求提上新的高度。然而由于医疗质量管理涉及的范围广、因素多、过程长, 环节复杂多变等原因, 医疗质量管理既成为医院工作的重点, 亦成为难点^[1]。中山大学附属中山医院是中山市唯一一家三甲医院, 作为全国 20 家最先进数字化医院的示范站点之一, 从 2001 年开始实施信息化建设, 迄今基本形成了覆盖全院业务流程、多个系统高度集成的一体化信息系

统。为探索构建一套切实可行的医疗质量管理体系奠定了现实基础, 医院于 2008 年开始了对医疗质量管理体系的研究与实践。

1 医院医疗质量管理的难点

1.1 数据质量问题

目前许多数据都是分散在不同系统中, 从病案首页无法连接到医嘱、病情记录等, 如果发现一个指标异常, 必须手工找到该医嘱或病情记录, 逐份查找病历。同时由于原始数据填写不全、不正确, 分析中有很多数据信息缺少或错误的情况。

1.2 手工查找和分析

医院质量指标的分析每月由统计室人员专门负责统计, 消耗大量时间与人力。而且常常由于分析

〔收稿日期〕 2010-06-29

〔作者简介〕 余元龙, 主任医师, 发表论文 127 篇。

的力度和能见度有限，无法多维度地去了解问题，了解是哪些原因造成某个指标异常，从而无法制定解决问题的方法。

1.3 信息需要二次加工

由于数据的孤立，造成很多信息无法从各自系统里自动关联起来，需要经过二次加工。例如在医嘱里只能看到药的名称，没有分类，在分析抗菌药的使用中无法了解到某类药物的使用情况、统计指标。因此在提供报告给管理者时需要二次手工加工，数据量庞大时处理非常棘手。

1.4 环节指标监控分析困难

只能在每月统计分析异常的质量指标，对环节指标进行监控分析困难^[2]。

2 利用数据分析与挖掘技术加强医院医疗质量管理

2.1 数据清洗、整合、关联

上述困难按性质归类主要可分成 3 大问题：系统间数据没关联、数据质量有问题；数据的分析能力弱；指标的生成与监控难。数据分析与挖掘技术恰恰是解决这些问题的有效工具。为能够获得客观的数据、有效的信息、高效的分析与监控奠定基础。在整合多个系统数据时，多数据源的数据格式、类型、存储方式会不同，在做数据清洗时需要首先对数据按照定义进行标准化。其次对于系统中无效数据或无法归类数据，将数据字典合理归类，按照数据字典和标准格式剔除无效数据。对于由于多系统描述相同的对象造成的数据记录重复，通过多维度分析，模糊动态匹配的技术剔除单数据源单一信赖数据。在经过初步的清洗、整合和关联后，数据由原来的孤立状态变成了整合状态，数据中多数的遗漏、重复、错误、不匹配也得以发现和解决，数据通过关联可直接利用，避免手工二次加工。

2.2 建立自由灵活的多维分析模式

2.2.1 自定义模型 模型主要由“维度”、“摘要

(指标)”、“信息（动态维度）”构成。维度是自定义模型中的重要概念，通俗说就是看问题的角度，在医疗质量管理中，所有与之相关的人、事、物在系统中有相应的字段都可以被制定在维度模型中。

2.2.2 多维分析 在自定义模型中，强大的分析功能体现在多维分析，它包括横向维度的交叉分析，以及纵向维度的潜入分析。要了解一个问题，可以通过宏观层面发现问题，也可通过横向交叉对比与纵向层层潜入来寻找问题的根源。

2.2.3 手术质量分析举例 在手术分析中可以充分使用多维分析和自由潜入功能来了解一个问题的始末细节。选定一个手术，查看死亡率最高的手术，潜入其他维度寻找死亡率高的原因，最终可以了解到死亡率最高或较高的手术术式、手术医生、手术时间、麻醉医生、麻醉类型、切合、感染、用药、诊断（病种）、诊断符合等等情况。将死亡率指标与实际医疗行为中的人、事、物关联，找到问题的原因，帮助医院规范医疗行为，提高效率。从某个手术的死亡率来分析问题，寻找某手术的死亡率最高的切口、费用等各种指标及手术相关人员工作量，见图 1。

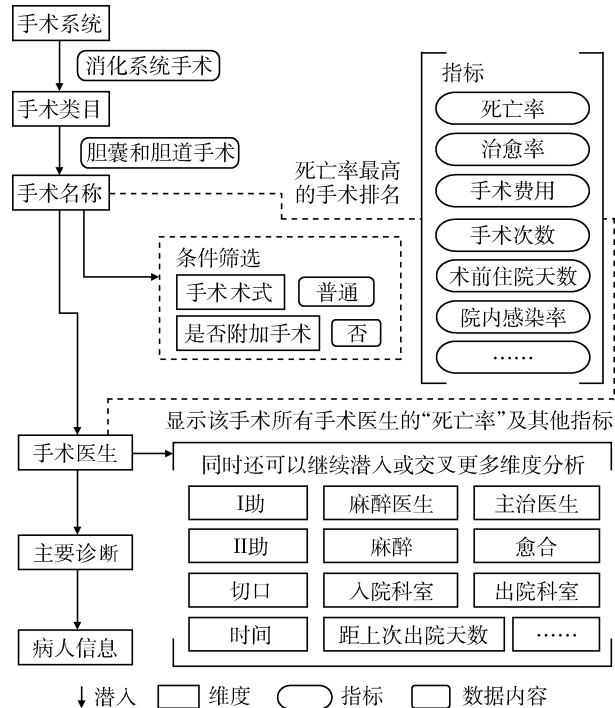


图 1 单个手术与病种死亡率等医疗质量指标分析

2.3 指标的生成、监控与分析

2.3.1 质量指标类型 将涉及到与医疗质量相关的经济指标分成 3 大类: 费用性指标, 比如住院费、各种类别费用; 经营性指标, 比如会诊次数、诊断符合率; 结果性治疗指标, 这个也是长期被忽略的可独立分析比较的重要指标, 如 24/48/72 小时重返手术室率、与单病种诊断相结合的死亡率、各种死亡类指标和负性事件指标^[3]。

2.3.2 核心指标监控方法 按照指标需要被监控的不同需求, 通过制定成仪表盘来监控最核心的指标。根据指标的类型, 可以制定出费用性质量指标仪表盘、运营性指标仪表盘或者治疗结果指标仪表盘。医院可以通过对仪表盘的监控, 了解每天每月的重要指标的变化。仪表盘具有提醒功能, 当发现问题时, 管理人员可以直接点击看问题的细节。也可以通过制定成实时报表或图表来监控一些时效性高的指标, 对异常情况提出报警或提示。

2.3.3 过程质量指标分析 过程质量指标很多属于第 3 类结果性治疗指标。将每条流程划分到最细致的点, 对每个点相关的人、事、物进行分析与监控。

3 系统设计

3.1 系统的技术基础——联机分析处理与交叉索引技术

将有关医疗行为、过程、因素等海量数据清洗、整合、关联、自由分析, 是系统设计的出发点。其中最主要运用到以下技术。1993 年诞生的联机分析处理 (On-line Analysis Processing, OLAP) 技术成为了系统技术支持的基础。联机分析处理的定义为: 使分析人员、管理人员或执行人员能够从多种角度对从原始数据中转化出来的、能够真正为用户所理解的, 并真实反映企业维特性的信息进行快速、一致、交互地存取, 从而获得对数据的更深入了解的一类软件技术^[4]。OLAP 的目标是满足决策支持或多维环境特定的查询和报表需求, 它的技术核心是“维”这个概念, 因此 OLAP 也可以说是

多维数据分析工具的集合^[5]。交叉索引 (Crossing-index) 技术是一项美国专利技术。由美国泓维软件 (Dimensional Insight) 的创始人 Frederick A. Powers 于 1995 年主导发明, 应用在数据模型的建立中, 通过对数据的自由交叉索引处理, 实现任意“维”之间的关联和潜入来获取最需要的信息, 也是通常说的多维分析。此外, ETL (Extract, Transform and Load) 工具 Integrator 在数据的清洗、整合与关联过程中发挥巨大作用。

3.2 数据基础——基于病案首页信息搭建平台

将医嘱系统接入平台, 计划连接手术、ICU、药品、电子病历等各个系统。从病案首页入手搭建分析平台, 是由它的性质和状况决定的。从它入手建立以医疗质量安全管理为核心的标准化应用体系成本低、可行性大。其次, 为了分析监控单病种的质量管理, 将医嘱系统、手术、药品、ICU 等系统数据关联到病案首页的数据上, 陆续整合在一个数据平台上。数据将不再是孤单的某个系统中的数据, 而是成为一个数据网络, 数据可以相互交流对话。多系统的数据集成也为今后对手术室医疗质量分析、门急诊 (包括 ICU) 医疗质量分析、电子病历医疗质量分析等打好基础。

3.3 系统设计

目前系统基于病案首页的数据, 已拥有近 100 个维度, 40 多个原始指标。维度之间能任意交叉、潜入 (下钻) 分析, 并能随时生成数据中所需要的, 包括二次计算的指标, 扩展后的系统基本能涵盖医疗质量监控分析的所有指标。维度的建立是将病案首页所有的字段作为可分析的维度, 在模型中成为直接点击就能看到数据信息的分析角度。维度可以相互任意交叉分析, 从而也就是多角度地看问题、聚焦问题, 当发现一个问题时依然可以多层次的潜入, 找寻到问题根源。模型维度建立, 见图 2。

指标的建立是抽取医院关注的与医疗质量管理有关的指标, 形成一套丰富的指标集合在模型中。对关键指标, 可以实时监控, 观测指标的变化, 也可以了解某个指标在一段时间的变化趋势。当指标

维度分类	维度	维度分类	维度		
时间信息	年月	手术信息	手术码		
	年		手术系统分类		
	月		手术类目分类		
	出院日期		手术名称		
	入院日期		愈合		
	上次出院日期		麻醉		
	手术日期		是否附加手术		
			I 助		
住院信息	系统（ICD10）		II 助		
	类 目（ICD10）		III 助		
	亚目（ICD10）		术式 1		
	系统（ICD9）		术式 2		
	类 目（ICD9）				
	亚目（ICD9）	医生信息	手术医生		
	出院科室		主任医生		
	出院科室代码		主治医生		
	出院科室分类		住院医师		
	入院科室		实习医生		
	入院科室代码		科主任		
	入院科室分类		进修医生		
	转科可是		研究生医生		
	距上次出院天数		质控医生		
			质控护士		
诊断信息 ICD10	主要诊断编码（ICD10）	病人信息			
	主要诊断治疗结果		病案号		
	主要诊断（ICD10）		住院号		
	并发症治疗结果		住院次数		
	附加诊断 ICD10		年龄		
	附加诊断治疗结果		职业		
	附加诊断		性别		
	外因码治疗结果		病人来源		
	感染治疗结果		省		
	其他诊断 1		市		
	其他诊断 2		镇		
	其他诊断 3		区		
	其他诊断 1 治疗结果		入院情况		
	其他诊断 2 治疗结果		医疗付款方式		
	其他诊断 3 治疗结果				
					
	诊断信息 ICD9		* 和 ICD10 基本一样，用于分 析 2003 年之前数据		

图 2 模型维度建立

出现异常或需要了解更细节的原因时, 可以通过不同维度聚焦或潜入分析。通过这样的分析, 所有指标从静止的数值变成动态的、可分析、可挖掘的信息。

4 结语

医疗质量管理是一片广阔的领域, 随着技术的发展与医院管理思想的进步, 两者的结合才能真正实现医疗质量管理水平的提高。数据分析与挖掘技术将数据的价值前所未有地展现与利用, 通过对医院数据本身的分析了解问题、做好管理, 是帮助医院建立切实可行的管理系统的出发点。这将是任重道远的过程, 通过正确的思路与技术支持, 医疗质量管理水平将达到一个新的高度。

参考文献

1 中国医院协会. 单病种质量管理手册 [M]. 北京: 科技文献出版社, 2008, 12.

2 赵刚明, 王锡宁, 舒婷, 等. 以医疗质量安全为核心的医院信息系统研究 [J]. 中国数字医学, 2009, 4 (1): 35 - 36.

3 王锡宁, 武广华, 班博, 等. 新型 CHQIS 医疗质量检测技术研究 [J]. 中国医院院长, 2009, (23): 62 - 69.

4 孙微微, 刘才兴. 数据仓库与数据挖掘 [J]. 农业网络信息, 2005, (1): 19 - 21.

5 汤效琴, 戴汝源. 数据挖掘中聚类分析的技术方法 [J]. 微计算机信息, 2003, (1): 52 - 54.

2011 年《医学信息学杂志》征订启事

《医学信息学杂志》是国内医学信息领域创刊最早的医学信息学方面的国家级期刊。主管：卫生部；主办：中国医学科学院；承办：中国医学科学院医学信息研究所。被美国《化学文摘》、《乌利希期刊指南》及 WHO 西太区医学索引（WPRIM）收录，中国生物医学核心期刊，并收录于国内 3 大数据库。

2011 年《医学信息学杂志》国内外公开发行，每册定价：15 元（月刊），全年 180 元。邮发代号：2-664，全国各地邮局均可订购。也可到编辑部订购：北京市朝阳区雅宝路 3 号（100020）医科院信息所《医学信息学杂志》编辑部；电话：010-52328673，52328674，52328671。

《医学信息学杂志》编辑部