

# A Research on the Methodology for Regulation of Diagnostics and Treatment From Clinical Terms

Lu Yubo<sup>1</sup>, Liang Zhiwei<sup>1</sup>, Yang Zhimin<sup>1</sup>, Li Xiaohua<sup>1</sup>, Chen qian<sup>1</sup>, Lin Yingzi<sup>1</sup>, Ou Aihua<sup>1</sup>,  
Cheng Yiyu<sup>2</sup>, Liu Baoyan<sup>3</sup>

(1. The 2nd Hospital Affiliated to Guangzhou University of Traditional Chinese Medicine, Guangzhou 510405;

2. Zhejiang University, Huangzhou 310027; 3. China Institute of Traditional Chinese Medicine, Beijing 100700)

**Target:** Terms from Original data of the patient record collected during regular clinical practice for the specific disease or the specific department-disease were Sorted out by computer with intelligent technique so as to discover regulation of diagnostics and treatment as well as to set up a useable technical platform. **Method:** 50 patient records were randomized selected from 600 times-cases of those patient were diagnosed and treated as hysteromyoma in Guangdong Province Hospital of Traditional Chinese Medicine. The word-frequency of the clinical terms written in the case report was counted. **Result:** There exists in a certain extent similarity among the terms of diagnostics and medical management for the specific disease of hysteromyoma. The dictionaries about both the rational terms and the irrational terms for the disease were compiled and a set of effectively and easily used software was developed. **Conclusion:** Regulation may be discovered after the statistics being taken with computer and intelligent technique for the word-frequency of the clinical terms in the patient record. The software designed with such function is helpful to scientifically sort out the clinical terms and discover the regulation of diagnostics and treatment.

**Key Words:** medical information; medical record, computerized; clinical terms; the regulation of diagnostics and treatment

## 从临床诊疗术语发现诊疗规律的方法学研究

吕玉波<sup>1</sup>, 梁志伟<sup>1</sup>, 杨志敏<sup>1</sup>, 李小华<sup>1</sup>, 陈倩<sup>1</sup>,  
林英姿<sup>1</sup>, 欧爱华<sup>1</sup>, 程翼宇<sup>2</sup>, 刘保延<sup>3</sup>

(1. 广州中医药大学第二附属医院, 广州 510120; 2. 浙江大学, 杭州 310027; 3. 中国中医研究院, 北京 100700)

**摘要:**【目的】运用计算机智能技术处理,对源自于专科专病临床实践原始资料进行名词用语整理,以期从中发现诊疗规律,并建立可操作的技术平台。【方法】从广东省中医院收治的子宫肌瘤患者共600例次的病历中抽样出50份,将其中诊疗术语的出现频率(简称词频)进行统计处理。【结果】发现病历中存在一定程序的共性诊疗用语,中医专病专科住院病历中存在一定程度共性的诊疗用语言和治疗手段,并以此建立了相关专病的理性词库和非理性词库,研制出简便有效的软件系统。【结

**论】**将散在于病历中的诊疗用语以计算机智能技术对术语出现的频率统计，可以从中发现其中的诊疗规律。为以研制而成系统软件，有助于科学地整理和统计病历中诊疗语言及发现诊疗规律。

**关键词：**医学信息学； 病案系统，计算机化； 诊疗术语； 诊疗规律

**中图分类号：**R197.324      **文献标识码：**A

目前医院病历绝大部分以文本纸质或非自由结构化的平面数码介质保存。病历的分析需要采用人工方法进行标引标注提取信息，这种病历回顾性分析手段难以满足多中心大样本的病历分析。如何从信息化医院保存的海量级的非结构化文档中，多角度、高效率地钻取有益的临床信息是当前医学临床信息学的一个热点<sup>[1,2,3]</sup>。在医学临床分析需要引入多种相关规范或标准。国外标准有疾病分类代码、医学系统术语表等，国内颁布有中医病证分类和中医临床诊疗术语方面的国家标准<sup>[4,5,6,7]</sup>。病历信息进行分析研究除了一般性的病历研究以外，还有在重大疾病研究、名老中医临证经验整理与承传工作方面，都可以利用信息技术挖掘和发现有益的临床诊疗信息或诊疗规律<sup>[8]</sup>。

本研究从专科专病入手，对本医院采集和保存的非自由结构数字化病历信息，运用现代计算机智能技术，进行探索性的分析归纳，旨在从临床实践原始资料中充分而客观地进行名词术语整理，探索与诊疗规律相关信息，建立一种适用于专科专病术语整理和诊疗规律辅助发现的有效研究方法和可操作的技术平台。

## 1 研究资料与方法

从广东省中医院于2004年1月至2005年5月已收入院的子宫肌瘤患者共600例次中，随机抽样出其中的50份病历，作为研究对象样本。

**1.1 病历的诊断标准** 住院病案中，出院诊断以国际疾病分类代码ICD10标准编码为D25（子宫肌瘤）的临床诊断和疾病编码<sup>[7]</sup>。

**1.2 病历的纳入标准** （1）出院诊断中第1诊断的ICD10编码为D25.001（子宫粘膜下肌瘤）、D25.201（子宫浆膜下肌瘤）、D25.901（子宫平滑肌瘤）的所有入院时间为2004年1月至2005年5月的病历。（2）选择采集出院中医诊断部分、中医诊断依据部分、出院记录之中治疗经过部分和出院医嘱部分共取4个部分的原始病历信息。

**1.3 病历的排除标准** 出院诊断中第1诊断的ICD10编码为D26.001（子宫颈良性肿瘤），以及D26.901（子宫良性肿瘤）的病历不予采纳。

**1.4 研究病历的预处理** 对经过上述纳入标准和排除标准的住院病历随机抽样出50份，并对之按纳入标准第2点所述的各部分相同的内容进行汇总，整理形成纯文本文档以作为源文件，用于进一步分析处理。

**1.5 病历分析方法** 本课题组研制开发的住院病历数据挖掘系统——科专病术语整理与规律发现子系统（图1），本系统具有分析纯文本病历的词频和甄别医用或专用术语的功能。首先，对该子系统多个分析参数调配，包括词长、词频域等。然后，启动系统运行并对每次的运行处理结果加以人工判断，根据人工判断结果再对各个相关的词库加以完善（图2）。之后利用这些词库对需要分析的文档或进行分析，或再进行反复的分析训练，以使系统的词库会不断地得到丰富和完善。当满意系统的运算结果时，可以停止系统的学习训练，确定分析参数并加以保存。利用此训练后的系统可以处理更多、样本量更大的同类文档。

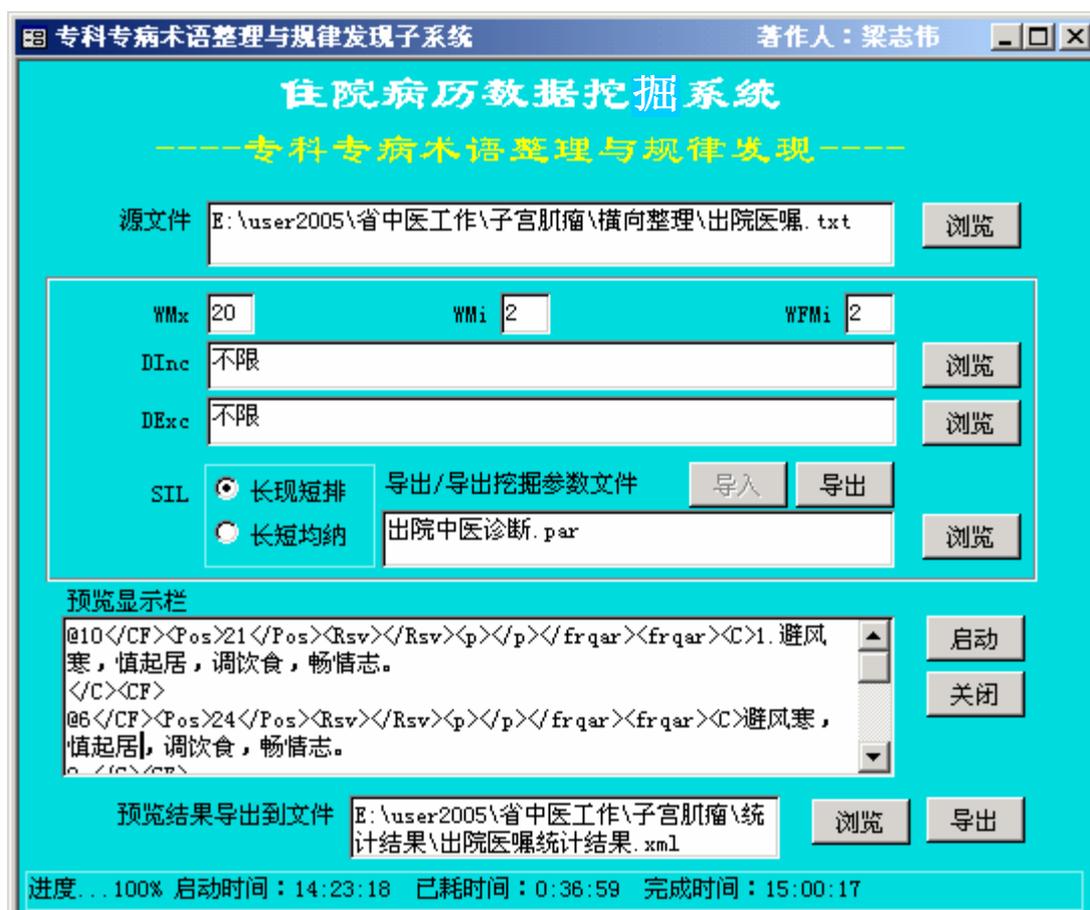


图 1. 专科专病术语整理与规律发现子系统操作界面

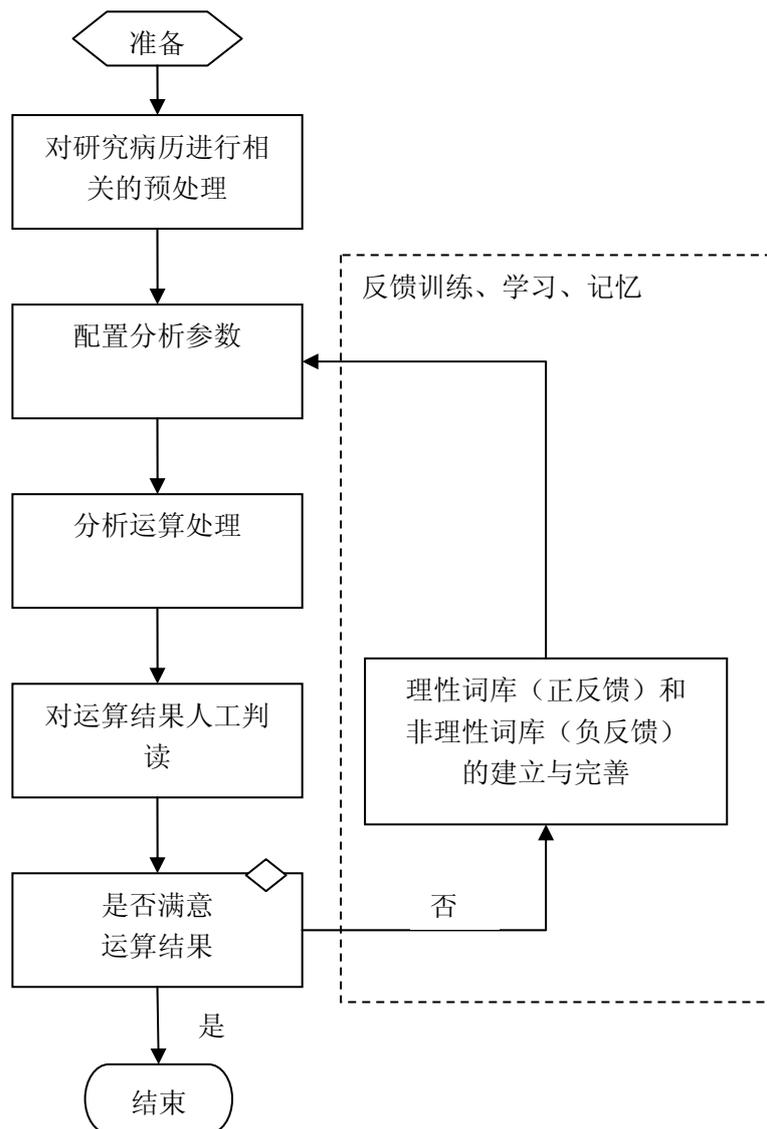


图 2. 病历分析系统的学习训练与分析流程

## 2 结果与分析

系统对上述 4 个部分的原始病历信息处理结果如表 1 至表 2。

将 50 份病历中出院中医诊断中词频数不低于 9 的出院中医诊断用语词频分布情况结果（表 1）与国家标准 GB/T 16751 中医临床诊疗术语（简称诊疗术语标准）进行对照<sup>[7]</sup>，发现大部分的中医诊断所使用的名称均符合诊疗术语标准用语，只有词频为 11 的“症瘕”一诊断用词未见于诊疗术语标准。

表 2 显示了对中医诊断依据统计后所得词频数不低于 9 的中医诊断依据用语词频分布情况。全部名词术语均可见于诊疗术语标准。

从表 1 - 2 还可以发现，血瘀、肾虚、石瘕、气虚、肾虚血瘀词频较高。词频次之的有：症瘕、气滞血瘀、年近七七、腰酸、腹痛、舌淡、舌暗、苔薄白、脉象呈脉弦细或脉细或脉弦或脉沉。这些术语反映了收治入院的子宫肌瘤病人的中医病证特征，其中医病名被诊断为石瘕或症瘕，其证候以肾虚血瘀和气滞血瘀为主，其症状可有腰酸、腹痛，或有经量增多，或有阴道出血。

表 3 和表 4 分别显示了出院记录之中治疗经过词频分布情况和出院医嘱及其出院带药词频分布情况。在住院期间的诊疗经过的统计中显示，预防感染、排除手术禁忌、子宫切除术等词频出现较高。从出院医嘱情况分析显示，全休、避风寒、禁房事、盆浴、调饮食、畅情志、慎起居、复查等词频较高，出院后的中药治疗使用频数较高的有：白术、云苓、太子参、甘草等。提示治疗方面有一定规范性，并揭示出治疗的规律、共性和个性情况。

表 1 50 份子宫肌瘤病历出院中医诊断用语词频分布情况

描述语	词频
血瘀	48
肾虚	28
石瘕(肾虚血瘀)	14
石瘕(肾虚血瘀)	14
气滞血瘀	13
气滞	13
症瘕	11

表 2 50 份子宫肌瘤病历中医诊断依据词频分布情况

描述语	词频						
血瘀	138	证属肾虚	22	血瘀之象	14	瘀型	
肾虚	100	发现子宫		脾虚	12	肾虚血瘀	
石瘕	71	增大	22	脉沉	12	型	10
肾虚血瘀	71	病机为肾		瘀阻胞宫	12	虚实夹杂	10
气虚	66	虚血瘀	22	隐痛	12	经量增多	10
瘀血	59	证属肾虚		气虚无力		舌淡红	10
气滞	54	血瘀	21	推动	12	证属肾虚	
苔薄	53	脉细	20	气虚无力		血瘀型	10
苔薄白	41	舌暗	20	推动血行	12	血瘀型	10
病位在胞		肾气渐虚	20	经期延长	12	日久结为	
宫	33	舌淡暗	19	血瘀之征	12	包块	10
脉弦	31	无力推动		病性属虚	12	均为肾虚	10
舌淡	30	血行	18	结为包块	12	均为肾虚	
子宫增大	30	瘀血阻滞	18	阻滞胞宫	12	血瘀	10
脉弦细	25	肝郁	16	瘀血阻络	11	崩漏	10
腰酸	24	胀痛	15	苔薄黄	11	阴道出血	10
病机为肾		病性属本		阻络	11		
虚	22	虚标实	15	瘀血阻滞			
不通则痛	22	气虚无力	14	胞宫	10		
		腹痛	14	属肾虚血	10		
		年近七七	14				

表 3 50 份子宫肌瘤病历出院记录之中治疗经过词频分布情况

描述语	词频						
预防感染	49	腹式全宫		术		盆腔粘连	12
防感染	49	切除术	18	外观正常	14	右输卵管	
切除术	43	子宫增大		血尿	14	系膜	12
平滑肌瘤	43	如孕	17	剔除术	13	予出院	11
宫切除	40	术后安返		会诊	13	复查血尿	11
宫切除术	35	病房	17	完善相关		全麻	11
排除手术		甲硝唑	16	检查	13	病情稳定	11
禁忌	35	粘连	16	血压	13	复查血尿	
送手术室	29	益气	16	入院后完		常规	11
送手术	29	右输卵管	16	善相关检		健脾	11
全宫切除	26	子宫肌瘤	16	查	12	附件外观	
子宫增大	25	右输卵	16	全子宫切		正常	11
化瘀	24	行腹式全		除术	12	查血尿常	
排除手术		宫切除术	16	腹式全子		规	11
禁忌症后	24	恢复良好	16	宫切	12	虚则补之	10
腹式全宫		多发性平		腹式全子		腰硬联合	
切除	22	滑肌瘤	15	宫切除术	12	麻	10
活血	21	活血化瘀	15	输卵管系		中医辨证	10
全宫切除		子宫多发		膜囊肿	12	腹式全宫	
术	21	性平滑肌		入院后完		切除术+	10
增生期	21	瘤	15	善相关检	12		
子宫平滑	20	宫切除术		腹式全子			
子宫平滑		+	15	宫切除	12		
肌瘤	20	宫颈炎	15	全宫切除			
增大如孕	18	血尿常规	14	术+	12		
大如孕	18	子宫切除	14	全子宫切			
		子宫切除	14	除	12		

表 4 50 份子宫肌瘤病历出院医嘱及其出院带药词频分布情况

描述语	词频	太子参	25	淮山	20	布渣叶	13
全休	42	甘草	25	陈皮	18	北芪	12
避风寒	41	门诊随		赤芍	17	丹皮	11
白术	37	诊	24	泽泻	16	定期门	
云苓	35	七剂	23	党参	16	诊复查	11
调饮食	33	中药七		丹参	15	夜交藤	11
盆浴	32	剂	22	枳壳	14	定期门	
复查	32	不适随		后下	14	诊	11
禁房事	32	诊	21	坐盆	14	鸡血藤	10
畅情志	32	门诊复		砂仁	14		
慎起居	31	查	21				

### 3 讨论

在临床实践的信息采集方面,目前大多数医院采集和保存的是纸质病历信息,部分医院已保存了数年的非自由结构数字化病历信息<sup>[9,10]</sup>,运用现代计算机智能技术,进行探索性的分析归纳,是充分利用好这些宝贵临床信息资源的手段。

专科专病术语整理与规律发现子系统是课题组自行研究开发的系统。通过对某一病种主题的病历内容进行词频分析,建立并不断地修正和完善理性词库和非理性词库,从而使系统拥有分析处理的正负反馈信息控制能力,系统的判断能力会逐渐得以校正与增强。

诊断标准的采纳与应用方面,本研究直接从临床病案库中按 ICD10 标准采集病例,不特别强调必须统一采用某种制订过的专业诊断标准,而是以学术包容与开放式以及“数据为先导”的研究模式、旨在通过大量临床数据分析归纳出辅助诊疗决策的结果,强调尊重住院病历的原始性和法律性、信任住院医疗小组的诊疗水平与常规临床医疗实践行为。

从本研究所采集的病历分析结果来看,大多数的临床描述用语用词除遵循 ICD10 标准外,大都与 GB/T1675 标准相符。提示这一套中医临床诊疗术语的国家推广标准具有普适性,并可能与其相应的宣讲、普及教育以及病案采集医院的医务管理有关。另外,还发现本病种有某些不少用的临床实际用语未被纳入 GB/T1675。对于这种潜在的、尚未被纳入标准或规范的临床实际高频用词,需考虑在标准修订或完善过程中是否采纳。

标准化与规范化是一个长期发展、修正完善、深化细化的过程,尤其新学科领域、欠发达学科领域的发展需求更为具前景,加上国内标准正处于与 ISO 接轨的过程,需调整和颁布新一轮的国家标准<sup>[11]</sup>。中医药标准化与规范化工作将会迎来更多的发展新机遇<sup>[12,8]</sup>。

在中医名词术语整理相关的研究工作中,过去曾采用分类分工地进行专家依据经验提议、多轮讨论,反复审核,然后制订相关术语规范,如 GB/T1675 的制定与颁发。近年有学者注意采用大范围问卷采集共性知识<sup>[13]</sup>,王永炎院士提出中医名词术语可探讨采用“小、大、小”范围原则开展研究工作,即首阶段在小范围由专业委员会主任副主任收拢意见并定初稿,形成术语集的问卷,第二阶段大范围征询各类各型各方面意见加以修订,最后阶段由擅辞书的少数专家审定<sup>[14,15]</sup>。

本研究的方法可高效地归纳临床实际常用语,若通过全国多中心临床病案的采集,整理出相关的名词术语,供用于新一轮规范或国标制订与修订,尤其可作为名词术语研究初期和中期研究阶段一种重要素材。本方法与其他的术语整理研究比较具有几个方面的优越性:首先,证据可以直接来源于受法律约束的海量临床实践数据而不是单纯地来源于领域专家的经

验；其次，能够高效而广泛地归纳产生有关中医术语的首次咨询问卷，或作用其他方法的征询问卷的修订辅助判别工具。第三，可客观地、全面地发现高频用词，避免了人为整理工作中因为考虑不周而疏漏了某些在临床实践当中广泛而经常使用的术语。

与此同时，从分析结果还发现子宫肌瘤这一专科专病的诊治有一定的共性和规律，包括中医病名诊断、中医证候分型、医嘱和用药等。提示系统在本研究范围和地点具有一定的准确度和精确度。系统可作为日后采用多中心协作，并以期外延推广和验证，为寻找学术共识和临床规律，提供了新的思路和可操作技术。

此外，也注意到有研究开始采用结构化甚至自由结构化的电子病历进行中医临床信息的采集<sup>[16, 17, 18]</sup>。结构化的单元信息，可作为基本统计单元，经过一定的数据整理后供统计分析。而自由结构化文档能够让信息的数理维度不受限制、用字用词自由、灵活方便，更贴近临床自然过程。其基本单元信息可内含图文声光等丰富而且可具有个性化的临床信息。自由结构化文档可能是最理想的一种保存和研究中医临床诊疗过程的文档。

通过控制若干诊疗干预措施进行随机对照临床评价方法，是一种公认可行的、可作演绎推理的临床科研手段，尤其适用于研究以辩证唯物主义哲学体系为基础的现代医学。然而，这并非开展科学研究和认知自然规律的唯一方法和手段。其中，现阶段中医临床具有明显个体化诊疗特征、采用多因素干预治疗、并且现代科技理论内涵相对地仍处薄弱，大量中医临床疗效暂时尚未必完全地甚或唯一地适合采用随机对照临床评价方法加以评价<sup>[19, 20]</sup>。中医临床更需要从专科或专病、甚或专家的临床业务角度，尤其是可着先考虑针对重大疾病防治、围手术期中医诊疗、以及名老中医传承工作等方面，进行大样本量的临床信息采集与计算机存储管理。通过对海量信息的分析处理，将会归纳和发现某些诊疗规律，从而为其后进行具有可演绎推理的随机对照研究提供更充实、更规范的研究素材。

### 小结

利用现代信息技术从专科专病临床实践原始资料中充分而客观地进行名词术语整理，探索归纳与诊疗规律相关信息，可能是一种适用于当前中医药专科专病研究的有效技术手段。本研究体现科学方法学中逻辑归纳法科研的思路，为多中心进行术语规范、标准的编写提供了纳词辅助工具，为名老中医临证经验整理与承传提供了新技术，为发现临床诊疗规律及其科学演绎夯实理论基础，丰富了“数据为先导”的临床实践型科研模式的内涵。

## [参考文献]

- [1] 周雪忠, 吴朝晖, 刘保延. 生物医学文献知识发现研究探讨及展望[J]. 复杂系统与复杂性科学, 2004, 1(3): 45-55.
- [2] 刘晋平, 黄宇虹, 陆小左. 数据挖掘在中医脉诊中的应用[J]. 天津中医学院学报, 2003, 22(3): 9-10.
- [3] 贺宪民, 吴骋, 于长春, 等. 数据挖掘技术在医学领域中的应用[J]. 第二军医大学学报, 2003, 24(11): 1250-1252.
- [4] 世界卫生组织(北京协和医院世界卫生组织疾病分类合作中心编译). 疾病和有关健康问题的国际统计分类(ICD-10) [M]. 人民卫生出版社, 第一版, 1996.
- [5] 陈惟昌, 李思生. SNOMED——的结构特点及其在医学信息系统中的意义[J]. 医学信息(云南), 1998, 11(1): 17-19.
- [6] 国家技术监督局. 中华人民共和国国家标准 GB/T 15657-1995 中医病证分类与代码[S]. 中国标准出版社, 第一版, 1995.
- [7] 国家技术监督局. 中华人民共和国国家标准 GB/T 16751.1-1997 中医临床诊疗术语(疾病部分、症候部分、治法部分) [S]. 中国标准出版社, 第一版, 1997
- [8] 未名. 中医药疗效及安全性研究全面展开重点挖掘名老中医学术思想. 中医药通报[J]. 2005, 4(1): 11.
- [9] 金文涛 蔡道章 高璐. 电子病历在骨科应用的一年总结[J]. 中华医院管理杂志, 2004, 20(4): 213-215.
- [10] 张广清, 刘玉珍, 王影, 等. 中医整体护理电子病历系统的创建与应用[J]. 护理学杂志: 外科版, 2005, 20(2): 12-14.
- [11] 沈同. 标准编写必读 GB/T 1.1-2000 应用指南[M]. 海潮出版社, 第一版, 2001;
- [12] 未名. 中医术语规范化取得成就. 世界科学技术: 中医药现代化[J], 2005, 7(1): 46.
- [13] 赖世隆, 杨小波, 温泽堆, 等. 证候宏观诊断标准基本框架的探讨[J]. 中国中西医结合杂志, 2005, 25(6): 552-555.
- [14] 王永炎, 朱建平. 中医药学名词审定工作的探讨[J]. 科学技术语研究, 2002, 4(3): 3-6.
- [15] 白茅, “中医药基本名词术语规范化研究”项目通过验收[J]. 中华医史杂志, 2004, 34(2): 124.
- [16] 刘乐宝. 结构化的智能电子病历系统[J]. 中国医院, 2004, 8(2): 58-60.
- [17] 梁志伟, 刘保延, 吕玉波, 等. 临床实践科研一体化新技术——自由结构电子病历[R]. 第三届数字中医药学术研讨会. 北京 2005; 10
- [18] 肖强, 吴伟斌, 陈联忠, 等. 自由结构录入法在电子病历系统中的应用[J]. 解放军医院管理杂志. 2005; 12(3): 222, 228
- [19] 王群红, 李宏建, 王介明, 等. 补充和替代医学研究中的问题及解决方法[J]. 国外医学: 中医中药分册, 2004, 26(1): 9-11.
- [20] 胡镜清, 刘保延, 王永炎. 中医临床个体化诊疗信息特征与数据挖掘技术应用分析[J]. 世界科学技术: 中医药现代化, 2004, 6(1): 14-16.