

临床决策支持系统技术现状及质量评价思路初探

李澍, 王浩, 任海萍* (中国食品药品检定研究院, 北京 102629)

摘要 目的:我国监管部门对临床决策支持系统(CDSS)的法律责任和监管模式正在探索之中,如何更好地规范临床决策支持系统,使其发挥临床辅助作用,是监管科学的重要研究方向。本文主要探索临床决策支持系统的质量评价方法和检测手段。**方法:**通过调研临床决策支持系统的发展现状以及分析现有的具有一定影响力的临床决策支持系统的技术特性,在此基础上,讨论临床决策支持系统在医疗领域的发展趋势和所面临的挑战。并对临床决策支持系统在质量评价过程中需重点关注的内容进行比较全面的总结,包括分类方法、评价方法和伦理问题。**结果与结论:**临床决策支持系统正处于蓬勃发展期,相应的技术和业态也在不断形成中,目前来说,绝大多数临床决策支持系统并不能直接给出相应的决策,多数还是处于启发和帮助医护人员进行思考的阶段。对于监管和评价来说,面临的挑战主要来自于算法适用性、决策透明性、界面可用性、系统鲁棒性和可移植性等因素的评价。

关键词: 临床决策支持系统; 人工智能; 算法; 伦理; 可移植性

中图分类号: TP181; TH77 **文献标识码:** A **文章编号:** 1002-7777(2019)09-1015-07

doi:10.16153/j.1002-7777.2019.09.009

Preliminary Study on Current Technique Status and Quality Evaluation of Clinical Decision Support System

Li Shu, Wang Hao, Ren Haiping* (National Institutes for Food and Drug Control, Beijing 102629, China)

Abstract Objective: Regulatory authorities are exploring the legal responsibility and supervision model of the Clinical Decision Support System (CDSS) in China. How to better standardize the CDSS and play its clinical assistant role is an important research direction of regulatory science. This paper mainly discusses the quality evaluation methods and detection methods of CDSS. **Methods:** Development trend and challenges of CDSS in the medical field were discussed based on investigation of the development status of CDSS and the technical characteristics of existing influential clinical decision support system were analyzed. Moreover, the key points of CDSS in the process of quality evaluation, including classification methods, evaluation methods and ethical issues were summarized. **Results and Conclusion:** The CDSS is developing vigorously, and the corresponding technologies and formats are developing constantly as well. At present, most of the CDSS which are still in the stage of inspiring and helping medical staff to think can not directly make decisions. As far as monitoring and evaluation are concerned, the challenges mainly come from the evaluation of algorithm adaptability, decision transparency, interface availability, system robustness and portability.

Keywords: clinical decision support system; artificial intelligence; algorithms; ethics; portability

基金项目: 国家重点研发计划(编号 2017YFC0111203)

作者简介: 李澍, 博士, 副研究员

通信作者: 任海萍, 博士; 研究方向: 主要从事有源医疗器械质控与评价; E-mail: renhaiping@nifdc.org.cn

随着医疗复杂程度的不断加深,在对患者进行相关的诊断和治疗等活动时需要获取和使用的医疗信息越来越大,同时知识新旧替代的速度也加快,这样的知识增长和更新的速度远远超出了个体学习的能力,知识与实践之间的差距成为了影响医疗健康服务质量的关键问题^[1-2]。1997年的一项关于冠心病治疗的调查表明,美国、加拿大和欧洲国家在阿司匹林、钙离子拮抗剂、 β 阻滞剂和抗缺血药物的使用上存在很多差异,并未遵循最佳的临床证据。1998年关于万古霉素使用的一项调查发现,86%的临床用药不符合临床指南,存在滥用情况。鉴于此,美国国立医学研究院发表了一篇“人是会犯错误的”报告,表明大部分的医疗错误是可以通过计算机系统避免的^[3]。因此,依赖计算机来做出更智能的临床决策支持系统(Clinical Decision Support System, CDSS)一直是医学信息学研究的核心。而在大数据环境下,机器学习正在成为一种极具潜力的智能服务形式。机器学习可以对医学数据进行自动学习,并提取其中隐含的规则或模型,进而对病情做出智能决策。因此,机器学习技术的发展为疾病智能诊断系统提供了一条崭新的途径^[4]。这些基于机器学习技术的疾病智能诊断系统主要以数据集、神经网络、贝叶斯分类器、主成分分析、演化计算、支持向量机、集成学习等关键理论和方法为核心。

虽然各种机器学习方法在CDSS诊断领域的使用取得了一定的效果,但已有智能诊断模型大多依赖于现有的机器学习模型,其泛化能力还有待进一步改善。另外,随着机器学习理论和方法的不断完善,各种新方法被不断提出,整个医疗诊断领域期待强泛化能力的机器学习方法。综合利用各种机器学习、自然语言理解和知识库系统成为将来智能健康服务最为核心的内容。但是临床决策支持长期以来并没有在临床实践中“有意义的使用”是由于自身的瓶颈和技术限制。Sittig D F等人^[5]把这些挑战总结为1)提高临床决策支持干预效率问题,具体包括如何改善人机交互接口、如何总结病人信息、如何对于用户的建议进行分优先级和过滤、如何对于具有并发症的患者进行建议合并、如何使用自由文本信息来驱动临床决策支持等。2)构建新的临床决策支持干预方式问题,具体包括如何重点

推进临床决策支持的内容建设和实施、如何挖掘临床大数据来建立新的临床决策支持。3)有效传播临床决策支持知识和干预的方法的问题,具体包括如何传播临床决策支持设计、开发和实施的经验,如何构建可以共享的临床决策支持模块和服务的架构,如何建立可以通过互联网访问的临床决策支持库等。

总体来说,伴随着信息技术的飞速发展和老龄化社会的不断邻近,CDSS逐渐成为医学信息学发展的焦点之一,这也为相关的检测、评价和标准体系工作带来了新的挑战。在这一领域中,国内外专家都逐渐认识到了目前CDSS系统发展所面临的瓶颈,但还缺乏统一认可的评价规范,随着更多CDSS系统走向医院,面临应用,如何在促进产品发展的同时,保证相关CDSS系统基本安全有效,成为了亟待解决的问题。本文旨在系统性介绍CDSS的技术分类、历史发展及国外监管现状,以期为我国CDSS系统监管提供思路和方法。

1 CDSS的技术分类和典型产品代表

根据CDSS知识表达特点,可以将CDSS系统划分为基于知识决策和非知识决策。基于知识决策的CDSS源自于早期的专家系统。这些系统试图复制人类决策者的逻辑和推理,基于现有知识做出基本决策。基于知识的CDSS的兴起是由于人们直觉地认为在医疗过程中会产生大量的经验知识。而计算机可以模拟现实临床医生的思维过程,然后根据已有的经验知识给出最终的诊断。然而,21世纪以来,CDSS不再试图做出严格的临床决定,而是倾向于提供各种可能的诊断/治疗选择,然后让临床医师自己做出最终决定^[6]。产生变化有多种原因。其中包括对计算机本身容易出错的潜在担忧,以及对医生/患者关系以及法律伦理的考虑等等。总之,计算机模拟现实临床医生的知识和推理技能还有很长的路要走。因此,目前的CDSS为临床医师提供了多种诊断/治疗选择,允许他们评估患者的临床症状,同时,利用系统作为可能诊断的参考点,典型的产品代表见表1。

本文通过检索ISI Web of Knowledge数据库进行典型CDSS研究。检索文献时限:1990年1月至2018年12月。检索关键词:CDSS or Clinical Decision Support System。检索文献数据库:web of science核心合集。检索文献量:相关文献9521篇。纳入标

准：1) CDSS综述文献，2) 文献被引用次数大于 100。排除标准：低质量或数据不全的文献。文献数量：经过纳入标准和排除标准筛选，共有近50篇有关CDSS的研究和综述被纳入。通过阅读并归纳所有文献的主要研究内容，共提炼出具有代表性的CDSS产品9项。

表 1 代表性的基于人工智能的 CDSS 对比分析

名称	研发机构	应用领域	技术特点
MYCIN ^[7]	美国斯坦福大学	细菌感染性疾病诊断和治疗咨询系统；使用人工智能领域中专家系统的建模方法	第 1 个功能较为全面的 CDSS，因出现时间早于个人台式机及网络等因素，没有应用于临床
DXplain ^[8]	美国哈弗大学	解释型辅助诊断系统。知识库包括 2400 个病种、500 多个临床症状及病症关系	医生输入基于医学词汇的病人信息，系统从知识库中输出一个可能诊断结果列表
HELP ^[9]	美国犹他州大学	基于医学知识库和文献引擎的综合性医院信息系统；决策逻辑由“事件—条件—行为”规则表示	包括呼吸系统疾病治疗方案合理性检查系统、实验室异常检查处理系统、用药合理性检查报警系统以及传染病监控系统
Iliad ^[10]	美国犹他州大学	内科诊断临床决策支持和教学，包含 1500 个诊断数据和 12000 个相关病种	使用贝叶斯模型、引入集群概念
QMR ^[11]	美国匹兹堡大学	早期基于人工智能的 CDSS 系统之一；包含 750 种内科疾病、5000 个临床症状，以及 50000 多个病症关系	允许临床医生以任何方式修改知识库和推理引擎，同时允许医生建立自己的知识库和推理引擎
VisualDX ^[12]	/	包含 6000 多种疾病的内置数据库	由症状和其他视觉线索为基础，将特定患者的异常图像与预先存在的图像进行视觉匹配
Isabel ^[13]	/	基于 Web 系统，设计时考虑到了医生的易操作性。以交互式系统的方式启发医生对患者陈述进行多层次思考	包含两个子系统：诊断清单子系统和知识迁移子系统
My Cancer Genome ^[14]	范德比尔特－英格拉姆癌症中心	癌症决策支持工具，针对不同癌症的基因突变提供专门治疗措施，同时与电子病历系统整合	数据来源于护理、医学文献及网络，包括来自美国、欧洲、澳洲和亚洲不同机构的 59 个临床和科学专家，定期更新
Waston Health ^[15]	IBM	基于专家知识库和人工智能技术；通过 150 万份病历和诊断图像、200 万页文本记录、文献等语料来构建肿瘤识别模型	基于自然语言处理、深度学习等技术进行大规模证据搜集、分析和评价

1.1 基于知识决策的CDSS

基于知识的CDSS是具有内置决策表，其中包含有关不同疾病、治疗等的信息。它们使用传统的逻辑判据（如条件逻辑）来决定治疗过程。基于知识决策的CDSS有三个主要部分。它们是知识库、推理引擎和用户交互，见图1。

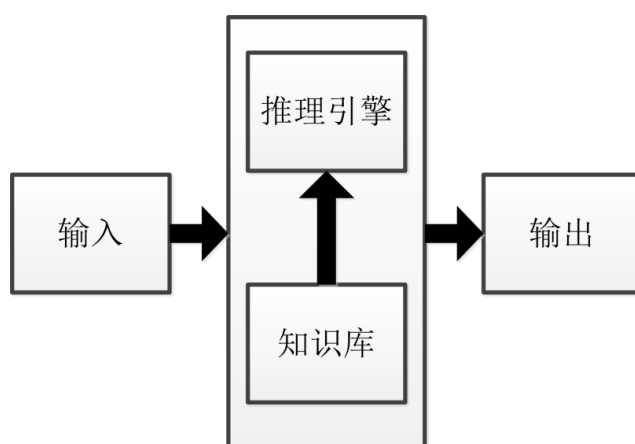


图1 基于知识的临床决策支持系统框图

知识库本质上是一个已编译的信息集，每一条信息都以if-then规则进行结构化。简单地说，用户输入患者信息，然后系统通过其知识库搜索匹配的疾病或可能的治疗方法^[16]。推理引擎则将逻辑系统应用于知识库，通过建立新的和/或更新的知识，使其“变得更智能”。它包含将知识库中的规则与患者可用数据相结合的规则，允许系统根据特定患者的病史和当前状况水平创建特定患者的规则和条件。因此，从本质上说，推理引擎和知识库是相互排斥同时又互相补充的。用户交互是指由临床医生如何将患者数据输入CDSS并接收结果的过程。在某些CDSS中，需要手动输入患者数据。然而，大多数情况下，患者数据是通过基于计算机系统提供的。基于知识的CDSS通常的格式是要求临床医生提供一定数量的输入，然后通过系统的知识库和推理引擎进行处理并输出一系列可能的诊断或治疗选项。

1.2 非知识决策的CDSS

非知识决策与知识决策的CDSS不同之处在于，它们不是由用户定义知识库，而是通过机器学习等人工智能形式进行学习。该系统不是查阅预先编制好的“if-then”等知识库，简单地“逻辑化”过去的经验；而是将学习内容固化到整个系统中。目前，有两种流行的非基于知识的CDSS类型：人工神经网络和遗传算法^[6]。

1.2.1 人工神经网络

人工神经网络（Artificial Neural Network, ANN）通过评估并最终从现有实例/事件中学习来

模拟人类思维。基于神经网络的结构与基于知识的CDSS非常相似。然而，与基于知识的CDSS不同，ANN没有预先定义的知识库。相反，神经网络研究患者数据中的模式，然后发现患者的体征/症状与可能的诊断之间的相关性。另一个显著的区别是，基于知识的CDSS通常比神经网络覆盖更广泛的疾病范围。为了使神经网络正常工作，首先必须对神经网络进行“训练”，将大量临床数据输入神经网络，对其进行分析，然后假设正确的输出。再将这些有根据的猜测与实际结果进行比较，并相应地调整权重，将正确的结果赋予更多的权重。通过持续迭代地运行这个过程，直到做出大量正确的判断。使用人工神经网络的优点是它不需要手动编写规则和寻求专家输入。神经网络还可以通过推断数据应该是什么来分析和处理不完整的数据，随着更多患者数据的研究，分析质量不断提高。不幸的是，人工神经网络也有一些缺点。由于它们的迭代性，训练过程非常耗时。更重要的是，此过程产生的公式/权重不容易读取和解释。因此，由于系统无法描述其使用某些数据的方式，由此，系统可靠性是需要着重关注的问题。

1.2.2 遗传算法

另一个非知识系统的典型例子是遗传算法。遗传算法是基于达尔文的自然选择和适者生存理论。正如物种为了适应环境而改变一样，遗传算法也会定期“复制”自己，以便更好地适应手头的任务。遗传算法通常先通过使用随机生成的解决方案来解决问题，下一步是通过使用“适应度函数”来

评估所有可用解决方案的质量（即“适应度”）。这些解决方案根据其适应度得分进行排名，更适合的解决方案通过彼此之间的相互交流，培育新解决方案的可能性更大。这些新的解决方案的评估与它们的父解决方案类似，并且该过程重复进行，直到找到最优的解决方案。

2 基于人工智能的CDSS监管面临的挑战

2.1 CDSS的分类

医疗器械分类是医疗器械科学监管的基础，而关于CDSS是否为医疗器械，如何分类的争论一直没有停止，2016年，美国通过《21世纪治愈法案》（21st Century Cures Act），作为1938年《联邦食品、药品和化妆品法案》（Federal Food Drugs and Cosmetic Act, FD&C Act）的补充。《21世纪治愈法案》规定了某些低风险的医疗设备将不再作为FDA的监管范围。对于医疗器械软件来说，《21世纪治愈法案》中3060(a)节对《联邦食品、药品和化妆品法案》进行了增补，明确了联邦法案中某些符合医疗器械软件功能的产品可以不作为医疗器械进行监管。具体而言，需要同时符合以下4个要求：

1) 不用于获取、处理或分析来自体外诊断设备的医学图像或信号，或来自信号采集系统的模式或信号[section 520(o)(1)(E) of the FD&C Act 136];

2) 用于显示、分析或打印有关患者的医疗信息或其他医疗信息（如同行评审的临床研究和临床实践指南）[section 520(o)(1)(E)(i) of the FD&C Act];

3) 用于旨在支持或向医护人员提供有关疾病或疾病预防、诊断或治疗的建议[section 520(o)(1)(E)(ii) of the FD&C Act 142];

4) 用于旨在使医护人员能够基于提供的信息独立地判断相关软件所提供的建议，从而使医护人员不主要依赖于任何软件提供的建议来对单个患者做出临床诊断或治疗决定[section 520(o)(1)(E)(iii) of the FD&C Act]。

一般来说，根据CDSS的定义，CDSS是指符合上述第1、第2和第3条要求的判据。但CDSS是否是被治愈法案排除在监管设备定义之外取决于第4条。只有当CDSS功能也符合第4条标准[section 520(o)(1)(E)(iii) of the FD&C Act]（该标准涉及对建议依据进行独立审查）时，CDSS功能才被排除在医疗器械定义之外。

根据以上定义，判断CDSS是否是医疗器械的

核心在于评价预期用户是否能够在不依赖于软件的情况下自己独立做出相同的诊断结论。同时，用户做出诊断结论的支持信息应是易于预期用户访问，易于用户理解的，并公开的信息（例如，临床实践指南、已发表的文献）。

2.2 CDSS质量评价面临的挑战

从评价的角度，知识库的科学扩充、推理引擎的复用性和可移植性，以及用户交互界面的友好性、易用性是需要重点考虑的内容。鉴于CDSS开发的耗时性和复杂性，一个好的推理引擎应该允许操作者通过现有的知识库构造新的知识库和逻辑规则。然而，这种持续动态变化的过程给监管提出了新的要求和挑战。

总体来说，质量评价的主要挑战来自如何对算法适应性（即算法在丢弃过时知识的同时“学习”新医学知识的能力）、决策透明性（即计算机基于输入进行判断的流程、依据及结果重复性的问题）、界面可用性（在不干扰患者/临床医师交互的情况下如何熟练地提出建议）、系统鲁棒性和可移植性（即不同区域、参数条件下CDSS系统结果的可靠性）等因素进行评价。具体包括：

1) CDSS自我升级问题：开发人员很难在开始就建立出持续有效的数据库及规则/条件集。因此，如何在使用过程中保证CDSS自我进化，且持续符合设计人员的预期要求是CDSS可用的关键问题。然而，自我进化对于质量评价来说，需要建立一套针对自我升级的规范性、科学性的评价方法。

2) CDSS的移植性问题：可移植性差一直是困扰CDSS发展的一个问题。具体表现为某一环境下非常成功的临床决策支持应用往往很难成功移植到其他临床机构中。分析其原因，涉及数据标准化、知识库需求差异性以及临床工作流程复杂性等多方面因素。与影像识别和分类人工智能软件不同，CDSS的输入除了包含客观的临床诊断输入，还包含很多医生的诊断、建议等。不同的医生、医院的诊断习惯、处方习惯还是有差异的。因此，需要针对CDSS的移植性设计相应的评价思路和规范。

3) CDSS的可靠性问题：和所有的计算机算法一样，CDSS也会出现软件bug或者故障。这对试图将其用于临床决策支持的临床医师构成了重大的判断风险。因此，在评价中，如何评价CDSS的可靠性及鲁棒性，也是评价部门面临的一个问题。

2.3 CDSS伦理问题

鉴于医疗领域信息的敏感性,很自然地,临床医生和CDSS设计师都会非常关注道德伦理问题和合法性。自从CDSS诞生以来,人们就提出了许多方法来规范这一领域中允许清单和禁止清单。例如,允许计算机辅助医生进行逻辑判断,禁止计算机违反医生的自主决策等。计算机不能取代人类的决策者已被社会所接受。从伦理的角度来看,计算机不应该被用来代替人类的基本决策。

2.4 我国对CDSS的监管现状

我国CDSS起步较晚,但发展迅速,目前已从理论和实验室研究阶段向产业化迈进;适用范围也从单一病种或单学科诊断向多模型、多病症决策系统发展。目前,我国监管部门对CDSS的法律责任和监管模式正在探索之中,如何更好地规范CDSS,发挥其临床辅助作用是一个长久的命题。总体来说,如何持续推动和规范CDSS规则编制,构建可得到的、高质量的,基于医学基础、临床指南、医学知识图谱、循证CDSS知识数据库和评价数据库,是CDSS监管的核心要求^[1]。

同时,由于CDSS涉及领域较广,卫生部门、药监部门、信息部门等监管机构均应承担起相应的责任。从卫生部门的角度来说,目前与CDSS直接相关的是《电子病历系统功能应用水平分级评价方法及标准(试行)》。该方法及标准将电子病历系统应用水平划分为8个等级。每一等级的标准包括电子病历系统局部的要求和整体信息系统的要求;从第3级开始,就要求具备初级医疗决策支持;从5级开始,对CDSS的标准就需要达到系统能够参与决策,而非简单的信息查询。从药监部门的角度来说,2018版的医疗器械分类目录中已包含决策支持软件类,明确将药物计算软件、计算机辅助诊断/分析软件和中医诊疗软件定义为医疗器械。但是从前文的分析可以看出,CDSS的范畴是大于目前的定义范围的。因此,对于专科/全科诊断治疗方案辅助设计等具体软件,还需要一事一议,通过国家药品监督管理局医疗器械标准管理中心的医疗器械分类界定程序来判断是否为医疗器械。国家药品监督管理局医疗器械技术审评中心于2019年发布基于深度学习的医疗器械软件的审评指导原则。同时,鉴于CDSS的特殊性,还需要针对前文中所述质量评价的难点进行深入研究。从信息部门的角

度来说,需要尽快出台相应术语标准,从而促进医疗机构实现跨系统信息共享和互连互通,从而提高CDSS数据集的可移植性和可规范性。

3 结论

CDSS为临床医生提高医疗诊断的准确性和治疗的可靠性提供了巨大的机遇。目前,有许多CDSS正在临床实践中使用,每个系统都为临床医生提供了一个独特的视角,能够使医生根据自己的个人偏好调整系统,从而按照自己的方式诊断疾病。CDSS除了能提供比单个临床医师更详细的患者病情描述之外,它还可能为临床医师提供被忽略的治疗建议,甚至组合治疗建议等。然而,CDSS系统鲁棒性和灵活性有待提高,并容易出现诸如可靠性差和导致缺乏经验的临床医生误用等问题。同时,系统的可用性和透明性对于保证患者的人身安全和临床医生对真实情况的了解都是必要的。总体而言,从科学评价的角度,需要针对算法自我学习性、决策透明性、界面可用性、系统鲁棒性和可移植性等因素进行综合评价。

参考文献:

- [1] 王浩,孟祥峰,李澍,等.数据集在人工智能医疗器械质控中的角色与要求[J].中国医疗器械杂志,2019,43:54-57.
- [2] 王浩,孟祥峰,刘艳珍,等.医疗器械中人工智能方法检验策略的研究[J].中国医疗设备,2016,31:67-70.
- [3] Medicine I O. To Err is Human: Building a Safer Health System[J]. Front Health Serv Manag, 2006, 18(6): 453-454.
- [4] 孟祥峰,王浩,王权,等.影像类人工智能医疗器械评价方法研究[J].中国医疗设备,2018,33:14-17,21.
- [5] Sittig D F, Wright A, Osheroff J A, et al. Grand Challenges in Clinical Decision Support [J]. Journal of Biomedical Informatics, 2008, 41: 387-392.
- [6] Berner E S. Clinical Decision Support Systems: Theory and Practice[J]. Men Am J Matern Child Nurs, 2007, 144(1): 375-391.
- [7] Melle W V. MYCIN: A Knowledge-based Consultation Program for Infectious Disease Diagnosis[J]. International Journal of Man-Machine Studies, 1978, 10(3): 313-

- 322.
- [8] Barnett G O. DXplain: An Evolving Diagnostic Decision-Support System[J]. JAMA The Journal of the American Medical Association, 1987, 258 (1) : 67-74.
- [9] Haug P J, Rocha B H S C, Evans R S. Decision Support in Medicine: Lessons from the Help System[J]. International Journal of Medical Informatics, 2003, 69: 273-284.
- [10] Turner C W, Lincoln M J, Haug P, et al. Iliad Training Effects: A Cognitive Model and Empirical Findings[C]// Symposium on Computer Application. Proc Annu Symp Comput Appl Med Care, 1991.
- [11] Lemaire J B, Schaefer J P, Martin L A, et al. Effectiveness of the Quick Medical Reference as a Diagnostic Tool[J]. CMAJ: Canadian Medical Association Journal, 1999, 161 (6) : 725-728.
- [12] Tleyjeh I M, Nada H, Baddour L M. VisualDx: Decision-support Software for the Diagnosis and Management of Dermatologic Disorders[J]. Clinical Infectious Diseases, 2006, 43 (9) : 1177-1184.
- [13] Graber M L, Mathew A. Performance of a Web-Based Clinical Diagnosis Support System for Internists[J]. Journal of General Internal Medicine, 2008, 23 (1 Supplement) : 37-40.
- [14] Carney P H. Information Technology and Precision Medicine[J]. Seminars in Oncology Nursing, 2014, 30: 124-129.
- [15] Doyle-Lindrud S. Watson Will See You Now: A Supercomputer to Help Clinicians Make Informed Treatment Decisions[J]. Clinical Journal of Oncology Nursing, 2015, 19: 31-32.
- [16] Garg A X, Adhikari N K J, McDonald H, et al. Effects of Computerized Clinical Decision Support Systems on Practitioner Performance and Patient Outcomes: A Systematic Review[M]. UK: Centre for Reviews and Dissemination, 2005.

(收稿日期 2019年6月24日 编辑 王雅雯)