DOI:10.3969/j. issn. 1671-3982. 2017. 09. 003

· 研究与探讨 ·

循证医学数据库信息模型的语义结构构建

朱妍昕,徐 维

[摘要]通过分析医务人员需求、调研循证医学相关标准和数据库,建立了循证医学数据库信息模型的文献特征、诊疗过程和证据强度三维顶层结构模块,并在顶层模块下表达了更加细化概念的数据组和数据元,规范了数据元表达的概念属性等等级语义结构。结果证明,采用模块化的方法构建循证医学数据库信息模型语义结构,能更好地管理和组织临床文献,使医务人员能更好地检索和利用循证医学信息。

[关键词]循证医学;语义;本体;信息模型

[中图分类号]R-58;TP311. 13 [文献标志码]A [文章编号]1671-3982(2017)09-0012-08

Construction of semantic structure for information model of evidence-based medicine database

ZHU Yan-xin, XU Wei

(Teaching and Research Section of Medical Informatics, Second Military Medical University, Shanghai 200433, China)

Corresponding author: XU Wei

[**Abstract**] A 3-dimensional top level structure module of literature characteristics, diagnosis and treatment process, and evidence level was established respectively for the information model of evidence-based medicine (EBM) database by analyzing the demands of medical staff and investigating EBM-related standards and databases. The data sets and elements with their concepts subclassified were expressed under the top level module, and the semantic structure of data elements-expressed concept attributes was standardized, which showed that the semantic structure of information model for EBM database that was constructed with modularized method can help medical staff to manage and organize the clinical literature, retrieve and use the EBM information more effectively.

[**Key words**] EBM; Semantics; Ontology; Information model

循证医学数据库是临床医务人员更新临床知识、解决临床问题的重要信息源。循证医学资源因提供的临床证据的直接性和针对性不同而分为不同的等级。Brain Haynes 将循证医学资源分为 5 个等级,即循证医学资源的“5S”模型[1]。从模型的最底层——以临床医学期刊文摘和全文为主的“研究(studies)”层到模型最顶层——能整合重要临床证据的“系统(systems)”层[2]。发达国家正在这 5 个层面在开展数据库建设并逐步完善,如位于模型顶层的知名循证医学数据库 Dynamed、Uptodate 等。我国循证医学数据库建设仍在起步阶段,虽有循证医学证据评价的语义模型与应用研究[3]等相关研究,但并没有根据用户需求调查构建的循证医学数据库,目前只有处于模型最底层——“研究(studies)”层的资源,即生物医学期刊文摘和全文数据库,如 SinoMed、CNKI 等。这些数据库一般只提供包括名称、作者、出版、摘要等外部特征以及关键词、

[基金项目]国家社科基金“构建基于本体的循证医学知识管理体系”(13BTQ012)

[作者单位]第二军医大学医学情报教研室,上海 200433

[作者简介]朱妍昕(1987-),女,江苏昆山人,硕士,馆员,发表论文 6 篇。

[通讯作者]徐 维(1971-),女,重庆市人,博士,副教授,主要研究方向为医学信息学。E-mail: zhuizhui@ smmu. edu. cn

主题词等内部特征检索。然而通过用户需求调查,深层次挖掘了医务人员循证医学信息需求后,我们可以确定仅仅通过这些常用的文献特征元数据是无法满足其需求的。因此,建设我国高级别的循证医学数据库势在必行。

本文构建了循证医学数据库信息模型的顶层宏观结构、中层语义结构和底层属性结构,为建立高级别的循证医学数据库语义结构提供了一种思路。

1　循证医学数据库信息模型的顶层宏观结构

医务人员对于循证医学文献的需求是多维的,而这些需求与其获取信息的目的或出发点密不可分。我们从文献调研、知名循证医学数据库分析以及循证医学相关文献标准 3 个方面归纳提取医务人员信息需求,并通过问卷调查以及专家咨询的方法进行验证。

1.1　文献调研

通过检索 PubMed、中国生物医学文献数据库(SinoMed)以及重庆维普数据库,对国内外医务人员的信息需求进行文献调研,筛选出采用问卷调查法、访谈法、观察法等分析医务人员信息需求的研究型论文,发现医务人员对医学信息的需求主要包括诊断、治疗、药物、症状特点、最佳证据等。

1.2　循证医学数据库分析

循证医学专题综述知识系统已经成为国外提供医学知识服务的重要形式之一,而国内目前尚无运行比较成熟的循证医学数据库。因此,我们选择 4 个国外知名的循证医学数据库(两个收费数据库 UpToDate 和 DynaMed,以及两个免费使用的网络数据库 MedSkills 和 CISMeF),从目的、涵盖的主要内容、检索入口、证据评级标准、特色等方面进行分析比较。除了文献调研获取的需求元素外,我们还获取了疾病(主题)名称、作者、出处、摘要、关键词、ICD-10 编码、病因学、预后等需求元素。

1.3　循证医学相关文献标准

本文认为一个好的循证数据库能够直接提供证据分级和推荐强度信息,能够提供对证据强度方面的检索或分类。在 SIGN、GRADE 以及中国循证医学中心分级标准等国际上使用范围较广的循证医学证据标准中,随机对照试验及与其相关的系统性综述、Meta 分析均属于最高等级证据类型,深受医务人员及科研人员的重视。因此我们分析了随机对照试验的质量评估标准 CONSORT 2010[4] 以及系统性综述/Meta 分析报告标准 PRISMA 2009[5],从中提取出了相应的需求元素。在考虑证据的表达性、标引和可读性等因素后,从 CONSORT 2010 中选用了试验设计、干预、结局、顺序产生、分配隐蔽机制、盲法和流程图等项目元素,从 PRISMA 2009 中选用了数据来源、检索策略、研究选择、数据采集过程、结论、干预等项目元素。

对应这 3 方面的需求进行归类并建立循证医学数据库信息模型的宏观结构。因此,循证医学数据库信息模型的宏观结构包括文献特征模块、诊疗过程模块、证据强度模块。文献特征模块提供最基本的题名、作者等文献外部特征和关键词主题词等文献内部特征;诊疗过程模块主要提供疾病、诊断、治疗、药物、症状等诊疗内容;证据强度模块依据 CONSORT 2010 声明及 PRISMA 2009 中主要关于临床试验方法的核心元素,包括试验设计、干预、结局、顺序产生、分配隐蔽机制、盲法、流程图等,可以为医务人员提供对临床研究文献质量判断的一定依据。本文提出的循证医学文献信息本体的宏观结构为三维结构,未来还可以根据需求增加语义表达维度(图 1)。

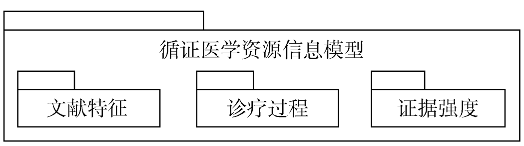

图 1　循证医学文献信息模型的宏观结构

1.4　验证

为保证这一信息模型顶层宏观结构的合理性和科学性,采用了问卷调研法和专家咨询法进行验证。调研对象为某三甲医院的医务人员及来自该院全国各地的进修医生,共 80 人;调研内容涉及其信息获取行为以及信息需求。结果表明,被调研的医务人员最关注的信息主要包括疾病的诊断方法、治疗方法、药物信息等方面,与文献调研的结果一致;也有调研对象提到"高质量的随机对照试验研究文献""临床研究的患者入选标准""样本数量/病历数"等。因此,我们在证据强度模块中新增了"样本量"这一元素。

此外,采用专家咨询法对本信息模型进一步验证,10 名来自某医科大学图书馆及附属医院的专家教授对信息模型进行了修改前后的论证。根据专家的意见,在诊疗过程模块中加入了 ICD-10 数据元素,以增强数据库与其他系统的互操作。

2　循证医学数据库信息模型的中层语义结构

中层语义结构是对顶层宏观结构的进一步细化。中层语义结构模型在构建时采用模块化方法,顶层宏观结构在细化时,形成相对独立的概念类别和元素,每一个概念类别或元素可以与其他类别或元素组合起来表达一个更复杂的概念。本文将概念类别称为数据组,概念元素称为数据元。

2.1　文献特征模块

目前,大部分循证医学原始研究文献数据库均收集了表达文献特征的元素,通常分为文献外部特征信息和文献内部特征信息。外部特征信息包括题名、作者、出处、摘要等,内部特征信息包括关键词、主题词等。本文采用都柏林核心元数据集(Dublin Core Metadata Element Set,简称"DC 元数据集")组织和描述文献特征的属性信息。DC 元数据集是都柏林核心元数据计划(Dublin Core Metadata Initiative,DCMI)负责维护的,由 15 个描述资源的基本且通用的信息元素组成的元素集[6]。它是目前被广泛认可的元数据标准,多用于表达网络资源信息。采用 DC 元数据集的组织结构方式和描述方式,将为今后建立循证医学资源网站、公开和共享信息奠定良好基础。

在文献特征模块中,除了对循证医学文献资源进行所有资源都适用的名称、描述、来源、语种、主题、创建者等数据元的常规描述以外,还针对循证医学文献特点,设立新的数据元表达特定循证医学信息。这些新的数据元也可纳入 DC 元数据集,作为自定义的修饰词。目前,我们在数据元"类型"下设置修饰词"循证医学文献类型",数据元"标识符"下设置修饰词"临床试验注册号"和"系统综述研究方案注册号",数据元"时空范围"下设置修饰词"临床试验招募国家",数据元"其他责任者"下设置修饰词"经费资助来源"(图 2)。

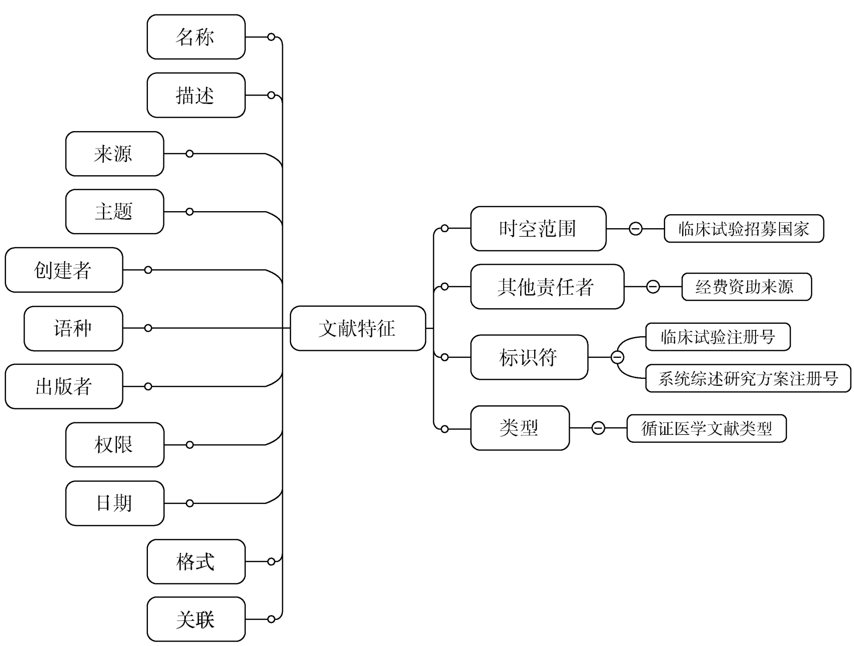

图 2　基于 DC 元数据集的文献特征本体框架结构

2.2　诊疗过程模块

诊疗过程是临床文献中临床内容的具体体现,具有实质性的循证信息,因而是医务人员极为宝贵的数据资源。诊疗过程模块中的需求元素根据文献需求调研结果,从现有的多个国外知名循证医学数据库提取,包括疾病、ICD-10、诊断、治疗、药物、症状、病因学、预后、患者教育在内的多个数据元。为了与电子病历系统对接,在诊疗过程模块的本体构

建中参考了我国颁布的《电子病历基本架构与数据标准(试行版)》中的《电子病历临床文档数据组与数据元》。根据其中的数据组分类,将疾病、ICD-10 归入诊断下位类,并表述成"疾病名称""疾病代码",与"诊断方法"一同构成"诊断"数据组。疾病代码则采用 ICD-10 编码表示属性。目前,ICD-10 这一疾病编码系统已被公认为疾病的分类标准且被广泛应用,国内也有医院将 ICD-10 编码库加入其电子病历系统中使用[7]。因此,为适应统一化管理趋势,在诊疗过程模块的本体框架结构中加入 ICD-10 编码,以提高循证医学数据库信息模型与其他系统的互操作和兼容性。如通过统一的 ICD-10 编码与电子病历系统准确对接,为医务人员提供即时的床旁循证文献辅助功能,提高临床决策效率。

另外,将需求元素中的"治疗"以及"药物"进行结构处理,构建"干预"数据组,其中包含"手术名称"以及"药物名称"这两个较为常用的数据元。在今后的研究中,可按需要加入其他干预方法数据元,如"穴位名称""膳食名称"等。诊疗过程模块的本体结构如图 3 所示。

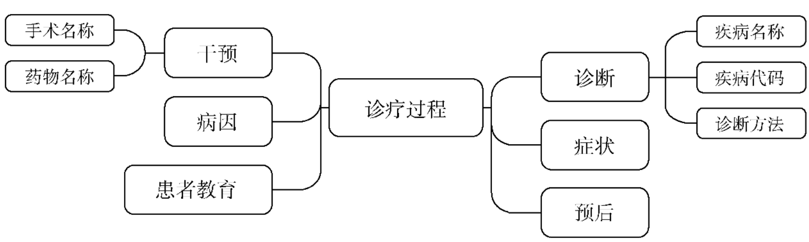

图 3　诊疗过程模块的本体框架结构

2.3　证据强度模块

考虑到证据的表达性、标引和可读性等因素,我们从 CONSORT 2010 中选用了试验设计、干预、结局、样本量、顺序产生、分配隐蔽机制、盲法和流程图等项目元素[4],从 PRISMA 2009 中选用了数据来源、检索策略、研究选择、数据采集过程、结论、干预等项目元素,构成了证据强度模块[5],其本体结构如图 4。对纳入数量、纳入类型、DOI、筛选流程图这几个数据元进行结构处理,构成研究选择数据组。由于诊疗过程模块中已经构建了"干预"数据组,根据模块化方法,该数据组也可以在证据强度模块中复用。

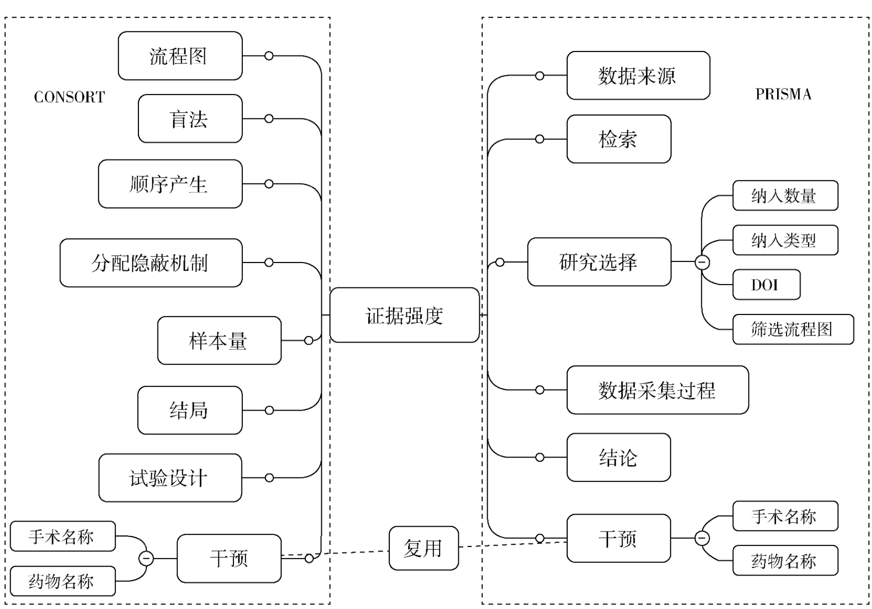

图 4　证据强度模块本体框架结构

3 循证医学数据库信息模型底层属性结构

确定了顶层和中层语义结构之后,对其中的数据元进行微观属性描述。本文参考 WS363.1 标准制定了循证医学数据库信息模型概念属性(表 1)。WS363.1 标准是医学信息领域数据元微观属性描述的顶层标准,该标准遵循 ISO/IEC 11179-3,而且也是中国医学信息领域各专业数据元属性描述的元标准。因此,建立在 WS363.1 标准上的属性描述,不仅与国际标准兼容,也为以后与医学信息领域其他标准和信息系统的映射和互操作提供依据。

表 1 以“疾病名称”为例的循证医学数据库信息模型概念属性

属性类型	属性名称	属性值
标识类	数据元标识符(Data Elemnt Identifier)	diseases Name
	数据元名称(Data Element Name)	疾病名称
	同义名称(Synonymous Names)	Diseases
定义类	定义(Definition)	A definite pathologic process with a characteristic set of signs and symptoms. It may affect the whole body or any of its parts, and its etiology, pathology, and prognosis may be known or unknown.
	来源(Source)	MESH 词表
关系类	映射关系(Mapping)	
	等级关系(Hierarchy)	诊断 ……疾病名称 ……疾病代码 ……诊断方法
	相关关系(Relevance)	手术名称、药物名称
管理类	(略)	
表示类	编码标准(Scheme)	MESH 词表
	数据类型(Data Type)	字符型
	表示格式(Data Format)	AN..40
使用类	使用条件(Condition Of Use)	
	必备条件(Obligation)	必备
	重复次数(Occurrence)	N 次
	实例(Example)	支气管炎

本文在 WS363.1 描述数据元的 5 类属性(标识类、定义类、关系类、管理类、表示类)基础上扩充了一个使用类。对于 WS363.1 的管理类、标识类和关系类中的公用数据元属性,主要表达有关数据元注册、版本等信息,通用于整个数据元集合标准,本文不做进一步讨论。

标识类中增加了同义名称属性,主要指容纳该数据元名称的其他同义表达,包括英文表达,为该数据元与其他信息模型和系统中对应的元素提供参照。

定义类中增加了来源属性,因为很多医学信息学领域的相同术语和词汇有不同的定义和定义来源。明确定义来源有助于了解该定义是否适合和准确。

关系类的属性是公用数据元属性“分类模式”,该属性是描述整个数据元集合或标准采用的分类法标准,而不表达特定数据元的等级结构和关系。因此在循证医学数据库信息模型概念属性中的关系类增加了表达概念等级、相同和相关关系的描述属性。

使用类包括描述数据元在实际使用时的条件、是否是必须具备的、是否可以重复使用、著录的实例表达等。以“疾病名称”这一数据元属性为例著录。

4 循证医学数据库信息语义模型

4.1 循证医学数据库 Protégé 语义模型

Protégé 由斯坦福大学开发,是目前国内外各领域技术人员使用最广泛的本体构建工具之一。

Protégé 能够较好地实现对本体和知识库的构建和编辑,对用户免费开放源码,用户不需要掌握具体的本体描述语言,就能够非常方便地构建本体模型[8]。Protégé 中可以实现添加或编辑类(class)、子类(subclass)、属性(properties)及实例(individuals),并生成本体图(OntoGraf)。

Protégé 中默认存在的顶级大类为“Thing”,所有的客观客体都是它的子类[9]。因此我们在“Thing”下面选择“Creat subclass”,建立循证医学文献子类,并按照其本体结构进行逐级构建。

循证医学数据库信息模型概念属性,可根据 Protégé 中对类的属性定义通过“object properties”实现,并且可定义类与类之间的上下级关系,形成一个本体属性树形结构(图 5)。

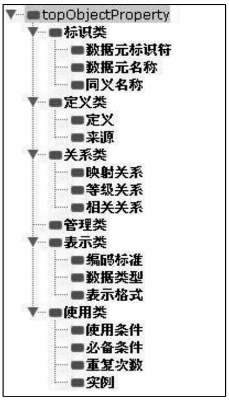

图 5　基于 protégé 的循证医学文献信息模型概念属性结构

本文选择 OntoGraf 中的 radial 生成本体结构图。由于 Protégé 的本体结构图对中文的兼容性较低,因此我们选择本体的英文表达进行构图(图 6)。

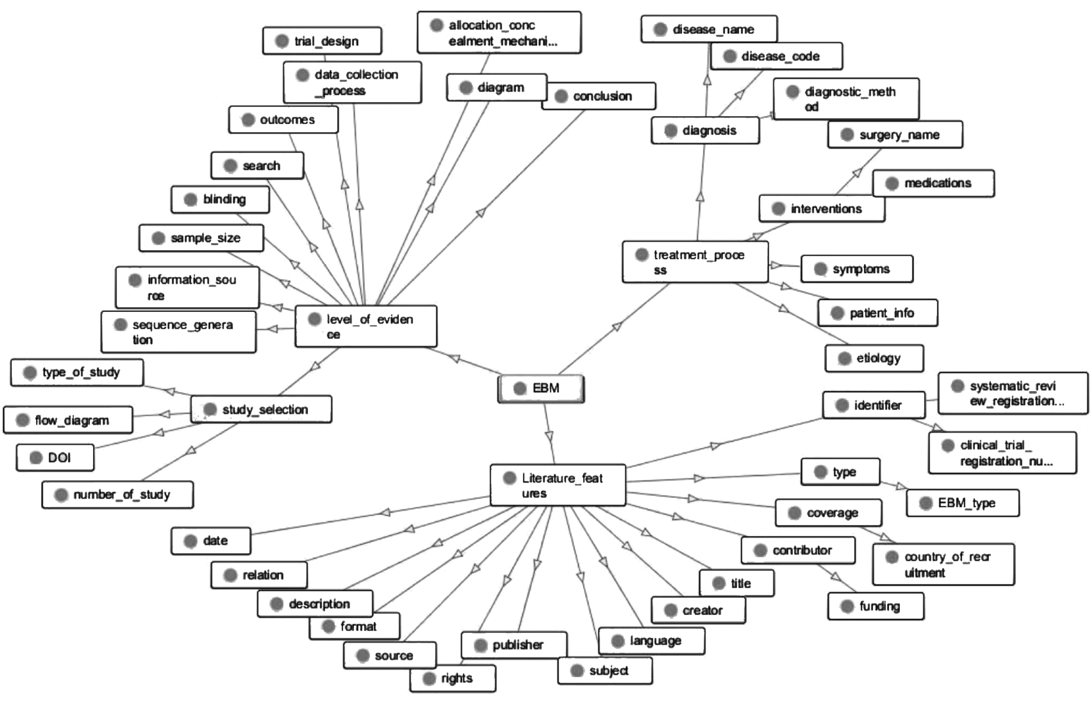

图 6　基于 protégé 的循证医学文献信息模型本体结构

4.2　XML 著录实例

PubMed 的题录文摘信息中已含有丰富的文献特征信息,在 PubMed 中称作为“字段”,经提取后成为循证医学数据库信息模型的数据元。而关于诊疗过程和证据强度的数据元通常表达文献更深层次的内容,因此需要通过文献全文提取其内容。

我们在 PubMed 数据库中选取了一个实例进行著录,用以验证该循证医学数据库信息模型的直观易读性。为了保证选取的实例著录元素较为丰富,我们精心挑选了 1 篇发表于《植物医学》(*Phytomedicine*)上的关于大蒜治疗高血压的 Meta 分析——“Garlic for hypertension: A systematic review and meta-analysis of randomized controlled trials”作为著录对象[10](图 7)。

循证医学数据库信息语义模型XML著录项	PubMed 文献题录信息

循证医学数据库信息语义模型XML著录项

……………………**文献特征信息**……………………

1. <名称>大蒜治疗高血压：随机对照试验的系统性综述及 meta 分析</名称>
2. <主题词>抗高血压药/*药理学；*大蒜；人类；高血压/*药物疗法；随机对照试验（主题）</主题词>
3. **<经费资助来源>**National Natural Science Foundation of China (no.81403375) </经费资助来源>
4. <描述>
   <摘要>背景:在过去的十年里，大蒜已经成为高血压病人中最流行的血压辅助疗法之一……</摘要>
   </描述>
5. <来源>Phytomedicine</来源>
6. <创建者>
   <第一作者>Xiong XJ</第一作者>
   </创建者>
7. <日期>2015.03.15</日期>
8. **<临床试验招募国家>**日本；南澳大利亚；中国台湾；俄罗斯</临床试验招募国家>
9. <标识符>
   <临床试验注册号> ACTRN12609000151235</临床试验注册号>
   </标识符>
10. <类型>
    <循证医学文献类型>系统性综述； Meta 分析</循证医学文献类型>
    </类型>

……………………**诊疗过程信息**……………………

11. <诊断>
    <疾病名称>高血压</疾病名称>
    <疾病代码>I10. X02</疾病代码>
    <诊断方法></诊断方法>
    </诊断>
12. <干预>
    <手术名称></手术名称>
    <药物名称>大蒜</药物名称>
    </干预>
13. **<症状>**收缩压≥140mmHg 和/或舒张压≥ 90mmHg
    </症状>

……………………**证据强度信息**……………………

14. **<检索>**
    (“garlic”OR“garlicin”OR“garlicoil”OR“garlicoilmacerates”OR“garlicextract”OR“garliccloves”OR“garlicpowder”OR“agedgarlicextract”OR“Alliumsativum”OR“alliinase”OR“allicin”)AND(“hypertension”OR“bloodpressure”OR“highbloodpressure”)AND(“clinicaltrial”OR“randomizedtrial”OR“randomizedcontrolledtrial”)
    </检索>
15. **<数据来源>**Pubmed(1995-2014), Cochrane Library(1996-2014), EMBASE(1995-2014)
    </数据来源>
16. **<研究选择>**
    <纳入数量>7</纳入数量>
    <纳入类型>随机对照试验</纳入类型>
    <DOI> 10.3892/etm.2012.819；
    10.1016/j.maturitas.2010.06.001；
    10.1038/ejcn.2012.178；
    10.1038/hr.2009.36；
    </DOI>
    <筛选流程图>有</筛选流程图>
    </研究选择>
17. **<试验设计>**治疗组使用大蒜；对照组使用与大蒜气味、颜色、形态一致的安慰剂
    </试验设计>
18. **<结论>**目前的综述表明大蒜是治疗高血压的有效和安全的方法。然而，在大蒜可以被推荐治疗高血压患者之前，仍然需要更严格设计的随机对照试验进行长期随访。
    </结论>

PubMed 文献题录信息

Phytomedicine. 2015 Mar 15;22(3):352-61. doi: 10.1016/j.phymed.2014.12.013. Epub 2015 Feb 4.

Garlic for hypertension: A systematic review and meta-analysis of randomized controlled trials.

Xiong XJ[1], Wang PQ[2], Li SJ[3], Li XK[4], Zhang YQ[5], Wang J[6].

Author information

Abstract

BACKGROUND:

In the past decade, garlic has become one of the most popular complementary therapies for blood pressure (BP) control used by hypertensive patients. Numerous clinical studies have focused on the BP-lowering effect of garlic, but results have been inconsistent. Overall, there is a dearth of information available to guide the clinical community on the efficacy of garlic in hypertensive patients.

AIM:

To systematically review the medical literature to investigate the current evidence of garlic for the treatment of hypertension.

METHODS:

PubMed, the Cochrane Library and EMBASE were searched for appropriate articles from their respective inceptions until August 2014. Randomized, placebo-controlled trials comparing garlic vs. a placebo in patients with hypertension were considered. Papers were independently reviewed by two reviewers and were analyzed using Cochrane software Revman 5.2.

RESULTS:

A total of seven randomized, placebo-controlled trials were identified. Compared with the placebo, this meta-analysisrevealed a significant lowering effect of garlic on both systolic BP (WMD: -6.71 mmHg; 95% CI: -12.44 to -0.99; P = 0.02) and diastolic BP (WMD: -4.79 mmHg; 95% CI: -6.60 to -2.99; P < 0.00001). No serious adverse events were reported in any of the trials.

CONCLUSION:

The present review suggests that garlic is an effective and safe approach for hypertension. However, more rigorously designed randomized controlled trials focusing on primary endpoints with long-term follow-up are still warranted before garlic can be recommended to treat hypertensive patients.

Copyright © 2015 Elsevier GmbH. All rights reserved.

KEYWORDS:

Blood pressure; Garlic; Hypertension; Randomized controlled trial; Systematic review

PMID:25837272 DOI:10.1016/j.phymed.2014.12.013

Publication types

Meta-Analysis

Research Support, Non-U.S. Gov't

Review

MeSH terms

Antihypertensive Agents/pharmacology*

Garlic*

Humans

Hypertension/drug therapy*

Randomized Controlled Trials as Topic

图 7 循证医学数据库信息语义模型 XML 语言著录实例

5 结语

本文从语义结构入手建立了循证医学数据库信息模型,采用模块化的方法构建了文献特征、诊疗过程以及证据强度这 3 个多级、可组配的原型模块,每个模块均可重复使用,使信息模型在建立和更新维护时更加容易,而且局部修改不影响其他部分的结构和内容,形成了完整的基于本体的循证医学数据库信息语义模型。采用相对简单的表达循证医学特定信息的元素,从而充分反映循证医学证据文献的关键信息,能够更好地管理和组织临床文献这一类重要的临床证据,使医务人员能更好地检索和利用循证医学信息。

本文也还存在着一定的局限性,如对于患者的信息需求尚未进行仔细研讨;对于除临床文献以外的信息对象尚未明确标准化,而且尚处于概念验证阶段,还未构建实际的循证医学数据库。这是我们下一步的研究方向,即构建出适合医务人员以及患者使用的循证医学数据库,并且通过某一种具体疾病的信息录入及其实际临床使用情况,对该数据库进行信息一致性和模型一致性验证,从而进一步完善该数据库。

【参考文献】

[1] 李幼平. 循证医学(第 2 版)[M]. 北京:高等教育出版社, 2009.

[2] Haynes B. Of studies, syntheses, synopses, summaries, and systems: the "5S" evolution of information services for evidence-based healthcare decisions[J]. Evid Based Nurs, 2007, 10(1): 6-7.

[3] 万　毅. 循证医学证据评价的语义模型与应用研究[D]. 西安:第四军医大学, 2009.

[4] Begg C, Cho M, Eastwood S, *et al*. Improving the quality of reporting of randomized controlled trials. The CONSORT statement[J]. The Journal of the American Medical Association, 1996, 276(8): 637-639.

[5] Moher D, Liberati A, Tetzlaff J, *et al*. Preferred reporting items for systematic reviews and meta-analyses: the prisma statement[J]. International Journal of Surgery, 2010, 8(5): 336-341.

[6] 国家质监局国家标准化管理委员会. ISO 15836:2009. 信息与文献 都柏林核心元数据元素集[S]. 2010.

[7] 李少玲. 基于电子病历系统的 ICD-10 在线编码[J]. 中国病案, 2013, 14(7):30-31.

[8] 林　喆, 刘　艺. 基于 Protégé 的大学英语教学资源本体构建[J]. 广西民族大学学报:自然科学版, 2013, 19(2):59-62.

[9] 吴煜煌. 基于 Protégé 的课程知识本体的构建[J]. 电脑知识与技术, 2008, 1(4):707-710.

[10] Xiong XJ, Wang PQ, Li SJ, *et al*. Garlic for hypertension: A systematic review and meta-analysis of randomized controlled trials[J]. Phytomedicine International Journal of Phytotherapy & Phytopharmacology, 2015, 22(3): 352-361.

[收稿日期:2017-08-25]

[本文编辑:黄思敏]