

本文引用:匡浩铭,杨张琪,匡建军,邓宇,张尚华.基于古今医案云平台分析中医治疗骨质疏松症的治疗规律[J].湖南中医药大学学报,2022,42(10):1757-1761.

基于古今医案云平台分析中医治疗 骨质疏松症的治疗规律

匡浩铭¹,杨张琪¹,匡建军²,邓宇^{2*},张尚华^{2*}

(1.湖南中医药大学,湖南长沙410208;2.湖南省中医药研究院,湖南长沙410006)

[摘要] **目的** 探索中医医案文本术语识别与抽取的方法,系统地总结分析中医治疗骨质疏松症的治疗规律。**方法** 基于湖南省中医药研究院中医药传承创新知识共享平台所收集的中医治疗骨质疏松症的医案文本,运用 Python 3.7 软件实现医案文本术语自动化抽取和分类,并形成结构化数据,最后通过古今医案云平台进行数据分析。**结果** 共收集得到 447 篇中医治疗骨质疏松症的医案,从医案中中共标注 1812 个实体,其中症状 179 项、脉象 20 项、穴位 10 项、中医证候 12 项、治法 27 项、中药 279 项,最常见的证型、治法、中药分别是肝肾阴虚证、补肾和杜仲。**结论** 运用古今医案云平台,对形成的结构化数据进行分析,为中医医案文本研究提供了新的手段,并发现了杜仲、肉苁蓉和菟丝子的不同配伍规律,为中医治疗骨质疏松症的现状提供了有价值的信息。

[关键词] 古今医案云平台;中医;骨质疏松症;医案;知识抽取

[中图分类号] R274

[文献标志码] A

[文章编号] doi:10.3969/j.issn.1674-070X.2022.10.027

TCM laws in treating osteoporosis based on the cloud platform of ancient and modern medical cases

KUANG Haoming¹, YANG Zhangqi¹, KUANG Jianjun², DENG Yu^{2*}, ZHANG Shanghua^{2*}

(1. Hunan University of Chinese Medicine, Changsha, Hunan 410208, China;

2. Hunan Academy of Chinese Medicine, Changsha, Hunan 410006, China)

[Abstract] **Objective** To explore the terminologies extraction and coding methods in the text of TCM medical cases, and to systematically summarize and analyze the laws of TCM in treating osteoporosis. **Methods** The knowledge sharing platform of TCM inheritance and innovation in Hunan Academy of Chinese Medicine was used to collect the medical cases of osteoporosis treated by TCM. Python 3.7 software was then applied to automatically extract and classify the terminologies of the medical cases, and the structured data was thus established to analyze via the cloud platform of ancient and modern medical cases. **Results** A total of 447 osteoporosis cases treated by TCM were collected, with 1812 tagged entities, including 179 symptoms, 20 pulse manifestations, 10 acupoints, 12 TCM syndromes, 27 therapies, and 279 Chinese material medicas. Among the data, the most common syndrome was yin deficiency of both liver and kidney, the most common therapy as tonifying the kidney and the most common Chinese material medica as Duzhong (Cortex Eucommiae). **Conclusion** We applied the cloud platform of ancient and modern medical cases to analyze the structured data, which has provided the new perspective in text research of TCM cases. The different compatibility rules of Duzhong (Cortex Eucommiae), Roucongrong (Herba Cistanches) and Tusizi (Semen Cuscutae) were found, which may provide valuable insight for the current TCM treatment of osteoporosis.

[Keywords] cloud platform of ancient and modern medical cases; TCM; osteoporosis; medical cases; knowledge extraction

[收稿日期] 2022-03-28

[基金项目] 长沙市科技计划项目(kq2202360);湖南省中医药科研计划项目(2021231);湖南省中医药研究院院级课题(202004)。

[第一作者] 匡浩铭,男,硕士,研究方向:中医医史文献。

[通信作者] *张尚华,男,研究员,E-mail:zshzc1997@126.com;邓宇,男,高级工程师,E-mail:5137096@qq.com。

骨质疏松症是最常见的全身性骨病之一,其发病特征是骨结构恶化和骨量慢性减少^[1],通常由于独立诱发因素(包括衰老和持续钙流失)导致骨质疏松性骨折。随着人口老龄化日益明显以及饮酒、饮食不平衡、吸烟和一些无意识因素日益普遍,这些因素都会改变骨形成和骨吸收之间的平衡,导致骨质疏松症发病率、骨折率越来越高^[2],并且骨折后有一半患者因其并发症而导致不能独立活动,严重降低了患者的生活质量,不仅增加了医疗负担,死亡风险还会持续升高^[3-4]。

中医在治疗骨质疏松症方面历史悠久,疗效确切,历代医者的学术思想和辨证经验都蕴含在其代表医案中。近几年,名老中医经验总结成为当下研究热点,但每位医者都会有自身用药习惯和特点,从中难以找寻中医治疗骨质疏松症的规律,所以运用大数据对历代医家治疗骨质疏松症的医案文本进行挖掘和整理治疗规律是有必要的。本课题组希望通过 Python 软件建立自动中医医案文本抽取方法或模型,自动将文本按照个案、诊次进行分割,再将“语料库”与单个医案文本进行自动匹配,并对这些信息进行结构化储存管理,从中分析出中医治疗骨质疏松症的治疗规律,拟解决中医学领域语言文本知识自动抽取问题和总结中医对骨质疏松症的治疗规律。

1 材料与方法

1.1 数据收集与整理

依托湖南省中医药研究院中医药传承创新知识共享平台对关键词“中医治疗骨质疏松症”“骨质疏松”的医案进行检索,通过阅读检索后的标题,初步剔除非相关性的医案,再将余下医案进行摘要阅读,剔除非中医治疗、非骨质疏松症状等不符合纳入条件的医案。最终将纳入的医案进一步全文阅读,筛除掉不完整的医案,把剩下医案进行全文下载。

1.2 中医语料库的建立

语料库构建工作的核心理念是制定规范和依据规范标注,《中医药学》第二版^[5]的疾病名词定义较全面规范,因此选择其作为基础语料库,进行中医临床症状信息抽取实验^[6]。由于每种疾病的症状不同,且每个人的症状表述用语也有差别,在进行具体医案研究时,还需要补充新的术语到语料库中,用于构建适用于当前研究的信息抽取模型。将数据自动采集至 Excel 中作为原始语料,用于标注中医医案信息。参照《中医临床基本症状信息分类与代码》的信息属性分类^[6],将命名实体分为 9 类,分别为中医病

名、症状、脉象、舌象、穴位、证型、治法、方剂、中药。

1.3 抽样验证及完善

运用 Python 3.7 软件对医案文本进行反向关键词标识,通过 SPSS 22.0 软件对医案文本进行随机抽样,抽出的文本通过人工检验的方式予以验证,将错误、遗漏的实体信息补充完善至语料库。并给出精确率(P)、召回率(R)和 $F1$ -Score($F1$),其中 $P = \frac{\text{正确识别的实体数}}{\text{识别的实体总数}} \times 100\%$ 、 $R = \frac{\text{正确识别的实体数}}{\text{所有实体总数}} \times 100\%$ 、 $F1 = \frac{2 \times P \times R}{P + R}$ 。

1.4 信息提取

基于 Python 3.7 正则表达式函数构建中医药文本知识抽取模型,从自然语言书写的中医医案文本中,自动抽取所需信息。正则表达式通常被用来检索、识别那些符合某个模式(规则)的文本,是对字符串操作的一种逻辑公式,本课题组构建了八大类(包括中医病名、症状、证型、治法、舌苔、脉象、方剂、饮片)中医学术语词典,并组成了用来识别中医医案术语信息的“规则字符串”。该模型首先依据诊次对医案文本进行“切割”,再通过术语词典匹配、识别并抽取字符串中的术语,依据中医医案语法规则,从不同诊次的文本中识别出类似阳性症状、改善症状、饮片加减等变化内容,最终形成可直接分析、利用的结构化数据。

1.5 数据分析

将由 Python 3.7 软件抽取结构化数据,批量导入古今医案云平台,运用其分析池进行症状、证型、中药的词频分析、复杂网络图分析和知识图谱可视化分析。

2 结果

2.1 数据收集与整理结果

首先运用湖南省中医药研究院中医药传承创新知识共享平台搜索关键词“骨质疏松症”相关医案,得到 688 篇,通过阅读标题初步剔除非治疗骨质疏松症的医案 140 篇,将筛选出的医案进行摘要阅读后剔除非中医治疗和重复医案 69 篇,后将纳入的医案进行全文阅读,剔除掉关键数据不完整的医案 32 篇,最后将余下 447 篇医案进行归档整理。

2.2 中医语料库的构建

根据语料库的规范性,以《中医药学》第二版^[5]作为基础语料,增加研究数据后构建中医医案术语词库分类及数据量统计共有 40 959 个实体,其中

症状8096个、证型4570个、治法10866个、穴位393个、脉象302个、舌象496个、方剂14708个、中药1528个。

2.3 抽样验证及完善结果

利用 Python 3.7 软件对医案文本进行反向标识,通过 SPSS 22.0 软件随机抽取 30 篇医案后人工校对。发现精确率、召回率、F1 值分别为 98.48%、95.31%、95.51%,具体数值见表 1。

表 1 分类统计精确率、召回率、F1 值

实体类型	精确率/%	召回率/%	F1 值/%
全部实体	98.48	95.31	95.88
症状	98.60	98.60	98.60
脉象	100.00	100.00	100.00
舌象	100.00	100.00	100.00
穴位	100.00	100.00	100.00
证候	98.10	97.72	97.91
治法	98.40	96.20	97.29
方剂	80.00	73.80	76.80
中药	96.76	96.13	96.44

2.4 信息提取结果

利用 Python 3.7 软件搭建好的模型抽取 447 篇医案文本的症状信息。发现刻下症状有疼痛、乏力、压痛、头晕、麻木、膝酸、畏寒、晨僵、背疼等 179 项;刻下脉象有脉弦、脉沉细等 20 项;刻下舌象有苔薄白、苔白等 42 项;治疗穴位有三阴交、绝谷、阳陵泉等 10 项;中医证候有肾阳虚证、肝肾阴虚证、血瘀气滞证等 12 项;刻下治法有补肾、健脾益气、活血化瘀等 27 项;中药有杜仲、肉苁蓉、菟丝子等 279 项。

2.5 数据分析

2.5.1 证候分析 在 12 项证候当中,肝肾阴虚证和脾肾阳虚证最为常见,其次是血瘀气滞证、脾肾两虚证、肾虚血瘀证、肾阳虚证等,其中肝肾阴虚证和脾肾阳虚证占到总比的 39.74%。具体证候频数前 10 位分布情况见图 1。截取频数前 10 位数据结合对各种症状频数的密切观察,发现骨质疏松症发病主要与肝、肾、脾、气血等息息相关,其病因病机主要是肾亏、脾虚、痰瘀阻脉 3 个因素^[7]。

2.5.2 治法频数分析 在 12 项证候当中,补肾治法最为常见,其次是生精补肾、健脾益气、强筋骨、滋阴益肾、益阴填髓等,其中补肾法占到总比的 20.42%,具体频数分布见图 2。综合治法频数前 10 位数据可知中医针对治疗骨质疏松症多从补肾、补肝、健脾、补气血等方面着手。

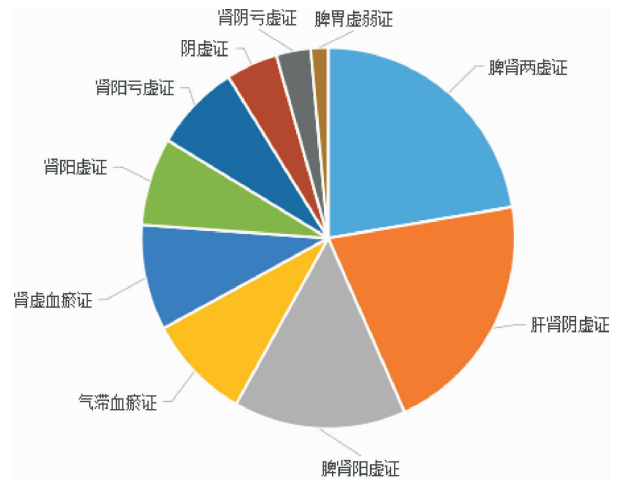


图 1 证候频数前 10 位分布情况图

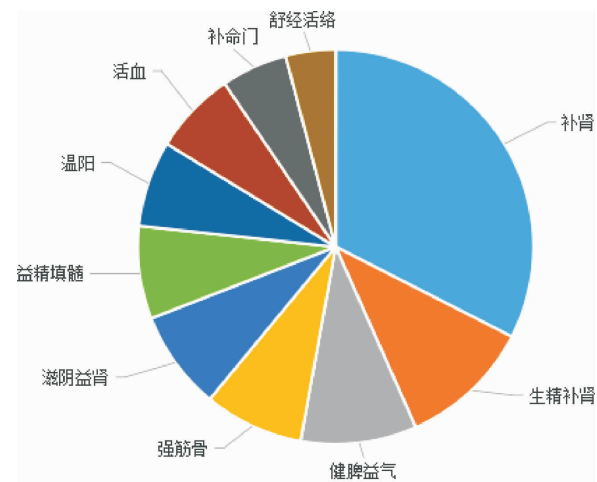


图 2 治法频数前 10 位的分布情况图

2.5.3 中药频数分析 本研究共纳入 667 个处方,其中中药 279 种,667 个处方中杜仲、肉苁蓉和菟丝子最为常见,其次是黄芪、当归、茯苓、骨碎补、甘草、补骨脂、白术。在处方中出现的频率,杜仲为 30.37%,肉苁蓉为 28.27%,菟丝子为 27.75%。截取频数前 10 位数据结合对中药药性分析,发现药物归肾经最为常见,频率为 49.24%,五味统计分析结果发现甘味、辛味、苦味等最为常见。四气分析发现多以温、平为主。具体药物属性分析见图 3。

2.5.4 中药对骨质疏松症的聚类分析 通过本研究医案中的药物信息聚类形成了 4 个聚类,第一类主要为补肾强筋骨类药,其代表中药有淫羊藿、菟丝子、肉苁蓉、骨碎补、补骨脂等;第二类主要是补气活血类药,其代表中药有当归、黄芪等;第三类主要是温通经脉类药物,其代表中药有桂枝、甘草、天麻、茯苓等;第四类主要是祛风止痛、镇静安神类药物,其代表中药有川芎、茯神、夜交藤、丹参等。具体聚类情况见图 4。

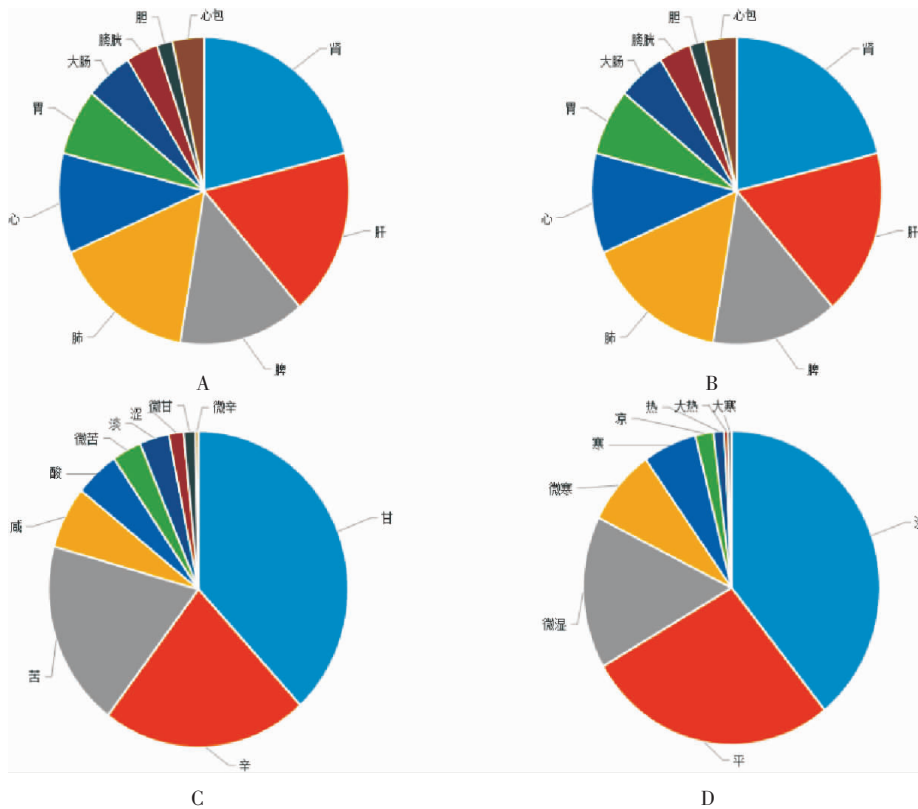


图 3 药物属性频数前 10 位的分布图

注:A.处方中常出现的中药;B.中药归经 C.中药五味;D.中药四气

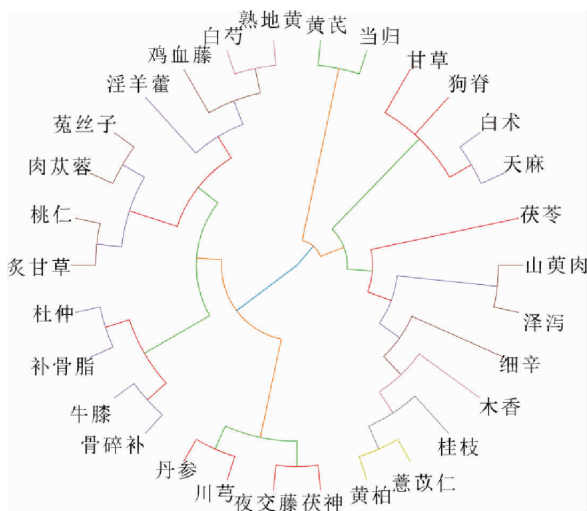


图 4 中药对骨质疏松症的聚类分析图



图 5 中药复杂网络图

注:黑色是权重≥100的中药,黄色代表60≤权重<100的中药。

2.5.5 知识图谱及复杂网络分析 将病例最多的肝肾阴虚证的治法、方剂、中药构化数据重新导入软件 Python 3.7,再将中药-中药进行复杂网络分析,由于医案基数较大,为提取出更具代表性的中药关系,现将边权重设置为≥60,提取出常用的 25 味核心中药。详见图 5。

3 讨论

骨质疏松症作为与人口老龄化相关的世界性健康问题,不仅严重降低了患者的生活质量,而且增加

了医疗负担^[8]。由于中医在治疗骨质疏松症方面历史悠久,疗效确切,现有越来越多的人尝试使用中医来治疗骨质疏松症。中医学作为历史悠久的学科,历代医者的学术思想和辨证经验都蕴含在代表医案当中,但由人工检索提取医案费时费事,若要基于中医医案数据挖掘研究,整理数据通常需要耗费整体研究时长的 70%,所以 20 世纪 90 年代语言文本知识自动抽取概念被提出,并在 2010 年美国国家集成生物与临床信息学研究中心(Informatics for Integrating Biology & the Bedside)首次发布了英文电子语料数据集^[9-10]。从此命名实体识别技术已被广泛应用于各个领域^[11]。但在中医领域,专业术语具有其特定属性

如歧义词、一词多义、多词一义等比较棘手的文法现象,且古今语法构成灵活多变,缺乏统一的规范,其表达差异无法被计算机识别,所以该项数据难以在中医领域实现^[12-13]。课题组设想并不试图全面理解整篇医案,只需对医案包含关键信息关键词进行提取和识别。现已建立中医属性的语料库,算法模型技术相应成熟,精确率、召回率、F1值分别为96.76%、96.13%、96.44%,可在前期保证大量中医术语的受控,从而做出提取和分类。

课题组通过此项研究从四诊角度分析,认为骨质疏松症患者多为虚证,舌苔主要呈现薄、白的现象,脉象主要呈现为弦、沉、细,根据证候数据分析发现骨质疏松症发病主要与肝、肾、脾、气血等息息相关。且病变的性质主要集中在代谢性骨病变上,病变部位集中于人体中轴骨及四肢长骨骨干,疼痛是其最常表现形式,通过具体医案分析发现腰痛、背疼、腰细酸软、下肢痉挛等发生概率与骨密度紧密相关,骨密度越低则以上症状发生概率就越高。从治疗经脉闭塞不通而发病的一系列临床症状出发,运用传统中医学“痛则不通”的思想理念,中医常针对髓气聚者之处脉穴加以针灸,如肾阴虚者取肾俞、照海、三阴交予以施针,肾虚者取中脘、气海、命门予以施针,气血瘀滞者取气海、足三里、三阴交予以施针等^[14-16]。在中药治疗方面常专注“肾为先天之本”“肾生骨髓”“其充在骨”“足少阴气绝,则骨枯……骨肉不相亲,则肉却软”“肾精不足,髓少,则生髓乏源,无以养骨”等理论^[17-18],配合身痛逐瘀汤、六味地黄汤、补肾活血汤、右归丸、金匱肾气丸等经典补肾活血方,从而利筋骨,养髓补血改善骨密度。中医治法方面多从补肾、补肝、健脾、补气血等方面着手^[16]。

中医药不仅能改善患者骨骼代谢,更能预防其本身相关疾病的发展,从而整体改善身体机能。通过我们的数据研究发现,447个处方中杜仲、肉苁蓉和菟丝子最为常见。在古今治疗骨痹的经典名方中代表性方剂如金刚丸、右归丸、补肾活血汤方都是以这些药为基础^[19]。因杜仲、肉苁蓉和菟丝子间配伍疗效甚好,乃至被日本国宝级书籍《医心方》所收录^[20],书上描述苁蓉杜仲茶:肉苁蓉5g,杜仲3g,菟丝子3g,五味子3g,续断3g,红茶5g。用前5味的煎煮液400mL泡茶,冲饮至味淡,多味药合用,可共同补肾益精治病症。正所谓药有个性之特长,方有合群之妙用,希望通过研究中医药配伍规律为中医药治疗骨质疏松症提供一定理论基础。

参考文献

- [1] KHOSLA S, HOFBAUER L C. Osteoporosis treatment: Recent developments and ongoing challenges[J]. *The Lancet Diabetes & Endocrinology*, 2017, 5(11): 898-907.
- [2] ASPRAY T J, HILL T R. Osteoporosis and the ageing skeleton[J]. *Sub-Cellular Biochemistry*, 2019, 91: 453-476.
- [3] NOH J Y, YANG Y, JUNG H. Molecular mechanisms and emerging therapeutics for osteoporosis[J]. *International Journal of Molecular Sciences*, 2020, 21(20): 7623.
- [4] MD K K S. Prevention and rehabilitation of osteoporosis[J]. *Wiener Medizinische Wochenschrift*, 2016, 166(1): 22-27.
- [5] 王文健. 中医药学[M]. 2版. 上海: 复旦大学出版社, 2012.
- [6] 张启明, 王永炎, 张志斌, 等. 中医历代医案数据库的建立与统计方法[J]. *山东中医药大学学报*, 2005, 29(4): 298-299.
- [7] 杜元灏, 李晶, 孙冬伟, 等. 中国现代针灸病谱的研究[J]. *中国针灸*, 2007, 27(5): 373-378.
- [8] 寇龙威, 郭珈宜, 李峰, 等. 针灸治疗膝骨关节炎作用机制研究[J]. *辽宁中医药大学学报*, 2021, 23(1): 196-199.
- [9] 徐东林, 李静. 小儿推拿五经穴与五行学说联系浅探[J]. *中国针灸*, 2022, 42(1): 49-50.
- [10] MCARTHUR C, LEE A, ALROB H A, et al. An update of the prevalence of osteoporosis, fracture risk factors, and medication use among community-dwelling older adults: Results from the Canadian Longitudinal Study on Aging (CLSA)[J]. *Archives of Osteoporosis*, 2022, 17(1): 31.
- [11] FEI H, REN Y F, ZHANG Y, et al. Enriching contextualized language model from knowledge graph for biomedical information extraction[J]. *Briefings in Bioinformatics*, 2020, 22(3): 1-14.
- [12] MARKAZI-MOGHADDAM N, ARAB M, RAVAGHI H, et al. A knowledge map for hospital performance concept: Extraction and analysis: A narrative review article[J]. *Iranian Journal of Public Health*, 2016, 45(7): 843-854.
- [13] 马一帆. 中医医案内容抽取与可视化的研究与实现[D]. 桂林: 广西师范大学, 2021.
- [14] 郭杨, 马勇. 中医药治疗骨质疏松症的常用处方分析[J]. *中国实验方剂学杂志*, 2010, 16(7): 188-191.
- [15] 匡浩铭, 杨张琪, 阮一蔚, 等. 补虚药治疗骨质疏松症的有效性及其安全性 Meta 分析[J]. *中医药导报*, 2022, 28(4): 119-123, 130.
- [16] 邓昶, 周明旺, 付志斌, 等. 骨质疏松症的中医病因病机及其治疗进展[J]. *中国骨质疏松杂志*, 2017, 23(8): 1105-1111.
- [17] 王满宜, 危杰. 股骨颈骨折临床研究的若干问题与新概念[J]. *中华创伤骨科杂志*, 2003(1): 5-9.
- [18] 马远征, 王以朋, 刘强, 等. 中国老年骨质疏松诊疗指南(2018)[J]. *中国老年学杂志*, 2019, 39(11): 2557-2575.
- [19] 白璧辉, 谢兴文, 李鼎鹏, 等. 我国近5年来骨质疏松症流行病学研究现状[J]. *中国骨质疏松杂志*, 2018, 24(2): 253-258.
- [20] 胡冬裴, 顾云之, 顾涣莹, 等. 《医心方》肾藏精理论辨治不孕不育病证特色研究[J]. *实用妇科内分泌电子杂志*, 2019, 6(30): 1-3, 32.