

基于数据挖掘探究国家专利复方治疗骨结核的用药规律

邓 杨¹, 李学清¹, 乐思微¹, 刘锦涛¹, 王 凯¹, 刘汝专^{2*}

¹广西中医药大学研究生院, 广西 南宁

²广西中医药大学附属瑞康医院脊柱外科, 广西 南宁

收稿日期: 2024年7月27日; 录用日期: 2024年9月18日; 发布日期: 2024年9月27日

摘 要

目的: 基于R语言平台分析国家专利复方治疗骨结核的用药规律, 为骨结核的中医药治疗提供参考依据。方法: 检索“国家知识产权局-中国专利公布公告”网站中治疗骨结核的复方专利数据, 采用Excel软件提取纳入文献信息并进行四气五味归经统计, 利用R语言进行频数统计、关联规则分析、聚类分析。结果: 共纳入70首专利复方, 共包含药物180味, 用药频数大于10的药物分别是当归、甘草、全蝎、蜈蚣、黄芪、鳖甲、没药、乳香、穿山甲、白芨。药物药性以温、寒、平性为主, 药味以甘、辛、苦味为主; 归经以肝、脾、肾经为主。关联规则分析显示获得以当归、甘草、全蝎、蜈蚣、熟地、乳香、没药为核心的药物组合; 层次聚类分析得到4类药物组合。结论: 分析强直性脊柱炎临床用药的常用组合规律, 为临床治疗强直性脊柱炎提供用药思路。

关键词

数据挖掘, 骨结核, 用药规律

To Explore the Medication Rule of National Patent Compound Treatment for Bone Tuberculosis Based on Data Mining

Yang Deng¹, Xueqing Li¹, Siwei Yue¹, Jintao Liu¹, Kai Wang¹, Ruzhuan Liu^{2*}

¹Graduate School, Guangxi University of Traditional Chinese Medicine, Nanning Guangxi

²Department of Spinal Surgery, Ruikang Hospital Affiliated to Guangxi University of Traditional Chinese Medicine, Nanning Guangxi

*通讯作者。

文章引用: 邓杨, 李学清, 乐思微, 刘锦涛, 王凯, 刘汝专. 基于数据挖掘探究国家专利复方治疗骨结核的用药规律[J]. 中医学, 2024, 13(9): 2378-2386. DOI: 10.12677/tcm.2024.139355

Abstract

Objective: To analyze the medication rule of national patent compound treatment based on R language platform and provide a reference for TCM treatment of bone tuberculosis. **Methods:** We searched the compound patent data for the treatment of bone tuberculosis in the website of “State Intellectual Property Office-China Patent Announcement”, extracted the included literature information and collected the four gas and five taste statistics, and used frequency statistics, association rule analysis and cluster analysis in R language. **Results:** A total of 70 patented compounds were included, including 180 drugs with drug frequency greater than 10 were angelica, licorice, scorpion, centipede, Astragalus, turtle nail, myrrh, frankincense, pangolin and *Bletilla striata*. Warm, cold, plain and sweet, spicy and bitter; mainly liver, spleen and kidney. The association rule analysis showed that the drug combinations with angelica, licorice, whole scorpion, centipede, ripening, frankincense and myrrh were obtained; the hierarchical clustering analysis obtained four types of drug combinations. **Conclusion:** The analysis of the clinical combination of ankylosing spondylitis provides medication ideas for the clinical treatment of ankylosing spondylitis.

Keywords

Data Mining, Bone Tuberculosis, Drug Use Law

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

骨结核(tuberculosis of bone and joint), 是由结核分枝杆菌感染骨、关节、滑膜和脊柱所引起的疾病, 常见特点是起病缓慢, 伤口化脓, 溃后不易收口等, 严重影响患者的骨骼发育, 甚至导致残疾的一类疾病, 严重影响患者的生活和工作[1][2]。骨结核占肺外结核的 35%~50%, 占结核总发病率的 2%~5% [3]。目前西医临床上治疗主要以药物为主, 但对于中后期治疗效果欠佳, 不良反应较大, 甚者需要手术治疗[4]。而祖国医学中医药治疗时注重辨证论治[5], 且中药多靶点、多环节的作用机制有利于对疾病的整体调理, 安全性高, 在骨结核的治疗中具有明显优势[6][7]。现基于 R 语言对国家专利复方治疗骨结核的用药规律进行分析, 为临床治疗骨结核提供用药思路。

2. 数据与方法

2.1. 数据的来源及整理

检索中国专利公布公告网站(<http://epub.sipo.gov.cn>)中药专利复方治疗骨结核的相关信息。检索词为“骨结核”“中药”“方剂”等。时间设立为建库至 2024 年 5 月。

2.2. 专利复方的纳排标准

纳入标准: 治疗骨结核的中药专利复方, 汤剂、颗粒剂、散剂、膏方等口服剂型都可。排除标准: 药酒、搽剂、膏药等复方; 单味或中药提取物的复方; 非主治骨结核的复方。

2.3. 资料的提取与录入

由 2 名研究者按照纳入、排除标准对国家专利发布的复方通过阅读全文进行筛选，产生的分歧由第 3 名研究者进行决断。将所获资料录入 Excel 软件，由两人分别录入数据，并进行交互查验所录入的数据。

2.4. 数据的规范

根据《中国药典(2020 年版)》[8]第一部及《中华本草》[9]将药名及药物的性味进行规范化处理，如：“微温”统一为“温”，“微凉”统一为“凉”；“川断”统一命名为“续断”，“川芎”命名为“川芎”，“元胡”命名为“延胡索”等。

2.5. 数据统计与挖掘

运用 Excel 表格对所纳入药物的性味归经进行统计分析，并绘制雷达图。使用 Excel 表格对涉及中药的用药频数进行统计。应用 R 语言函数进行数据关联规则分析，进行可视化展示，并将关联规则排名前 10 的 2 阶、3 阶数据导出 Excel 表格。利用 R 语言中的函数对高频次药物进行层次聚类，绘制相应谱系图及聚类热图。

3. 结果

研究纳入符合纳排标准的专利复方共 70 首专利复方，共包含药物 180 味。

3.1. 药物性味归经统计

对纳入 70 首专利复方的药物进行性味归经统计，并绘制出雷达图，结果显示治疗骨结核的药物药性以药物药性以温、寒、平性为主；药味以甘、辛、苦味为主；归经以肝、脾、肾经为主，见图 1。

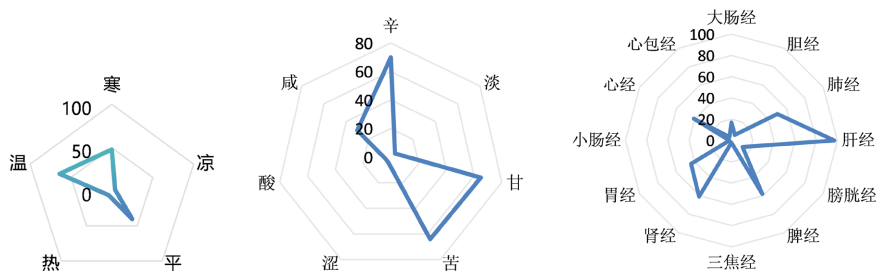


Figure 1. Include the radar chart of the four qi, five flavors, and meridian return of drugs
图 1. 纳入药物四气、五味、归经雷达图

3.2. 药物频次统计

将纳入治疗骨结核的用药频数大于 10 的药物分别是当归、甘草、全蝎、蜈蚣、黄芪、鳖甲、没药、乳香、穿山甲、白芨，结果见表 1。

3.3. 药物的关联规则

关联规则分析：关联规则的表达式：A 药=>B 药，其中 A 药为前项，B 药为后项，支持度表示 A 药与 B 药的药对出现的频数，置信度表示在 A 药出现的前提下 B 药出现的概率，提升度代表 A 药的出现对 B 药出现的概率提升的程度，提升度 > 1 时代表有提升效果，提升度为 1 表示两者没有关联，提升度 < 1 表示为无效的强关联[10]。设置最小支持度为 0.07，最小置信度为 0.8，按照置信度的大小降序排序，将排名前 10 药物的 2 阶、3 阶导出 Excel 表格，结果如表 2。

Table 1. Table of drug frequency greater than 10**表 1.** 用药频数大于 10 的药物频数表

中药	频数
当归	22
甘草	17
全蝎	17
蜈蚣	17
黄芪	14
鳖甲	13
没药	11
乳香	11
穿山甲	10

Table 2. Table of 2nd and 3rd order association rules for the top 10 drugs**表 2.** 前 10 药物 2 阶、3 阶关联规则表

序号	关联规则	支持度	置信度	提升度
1	{骨碎补} => {当归}	0.07	1.00	3.00
2	{茯苓} => {当归}	0.09	1.00	3.00
3	{熟地} => {当归}	0.12	0.88	2.66
4	{全蝎} => {蜈蚣}	0.22	0.88	3.42
5	{蜈蚣} => {全蝎}	0.22	0.88	3.42
6	{白术} => {当归}	0.09	0.85	2.57
7	{血竭} => {没药}	0.07	0.83	5.00
8	{茯苓} => {熟地}	0.07	0.83	6.11
9	{乳香} => {没药}	0.13	0.81	4.90
10	{没药} => {乳香}	0.13	0.81	4.90
11	{茯苓, 熟地} => {当归}	0.07	1.00	3.00
12	{鳖甲, 全蝎} => {蜈蚣}	0.10	1.00	3.88
13	{鳖甲, 蜈蚣} => {全蝎}	0.10	1.00	3.88
14	{甘草, 蜈蚣} => {全蝎}	0.07	1.00	3.88
15	{甘草, 熟地} => {当归}	0.09	0.85	2.57
16	{当归, 茯苓} => {熟地}	0.07	0.83	6.11
17	{甘草, 全蝎} => {蜈蚣}	0.07	0.83	3.23
18	{当归, 熟地} => {甘草}	0.09	0.75	2.91
19	{当归, 甘草} => {熟地}	0.09	0.75	5.50
20	{当归, 熟地} => {茯苓}	0.07	0.62	6.87

3.4. 关联规则可视化

运用 R 语言中的函数对关联规则分别进行分组矩阵图、网状图可视化[10]; ① 分组矩阵图: 分组矩

阵是将共同点比较相近的规则聚成类，通过使用聚类方法将规则分组，然后展示关联规则的大体分布，并进行可视化展示；圆圈越大表示支持度越高，颜色越深表示提升度越高，结果见图 2。② 网状图：圆圈越大表示支持度越高，颜色越深表示提升度越高，箭头指向次数多的中药提示为中心药物，结果见图 3。根据药物分组矩阵图、网状图综合得到药物：当归、甘草、全蝎、蜈蚣、熟地、乳香、没药的核心药物组合。

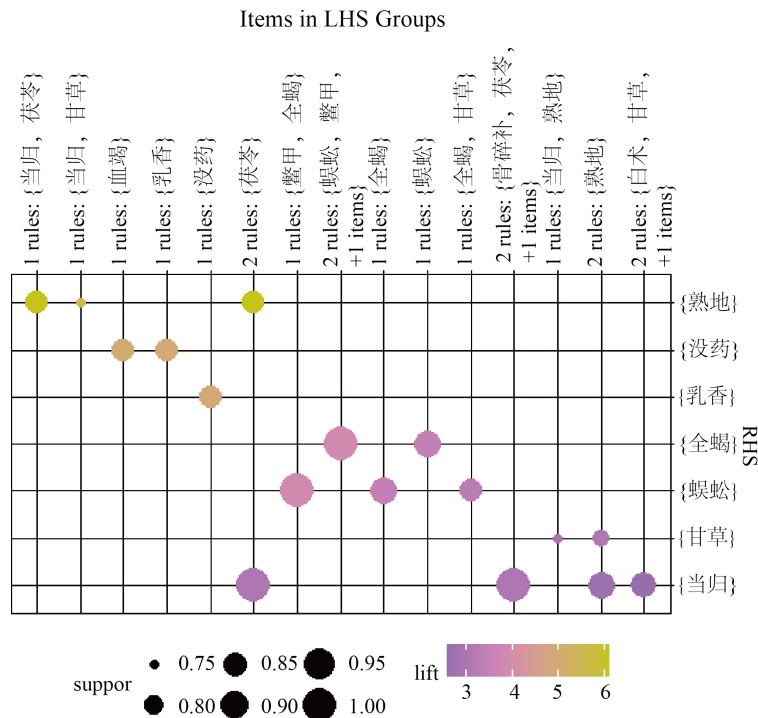


Figure 2. Drug matrix diagram
图 2. 药物矩阵图

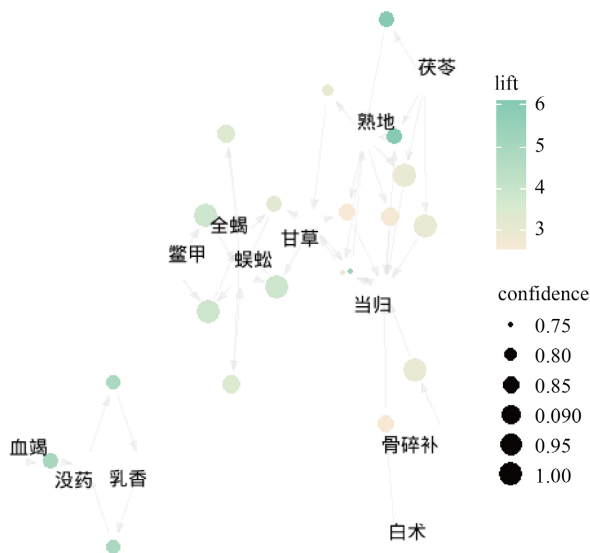


Figure 3. Drug network diagram
图 3. 药物网状图

3.5. 高频药物的聚类分析

聚类分析是将纳入的众多标注数据依据数据库中数据的特征划分不同多个簇集的过程，簇内数据间呈相关关系，不同簇集数据间则不相关，簇集间存在差异性[10]。将纳入药物频数在前 30 味药物进行聚类分析。用投票图可视化展示，可通过最后结果综合得到最佳的聚类数。投票图结果如图 4。投票图综合可得出聚类数为 3 类、4 类、15 类药物组合。做出 4 类谱系图，结果如图 5。最后将高频药物运用 R 语言绘制出聚类热图，可将聚类组合进行观察验证，颜色越深表示药物与药物之间的聚类层次越强，结果如图 6。

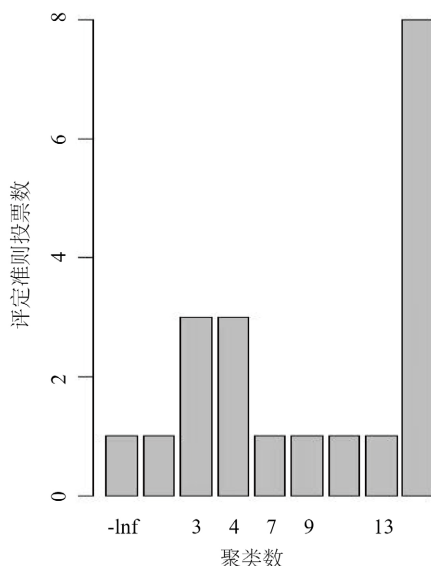


Figure 4. Voting chart

图 4. 投票图

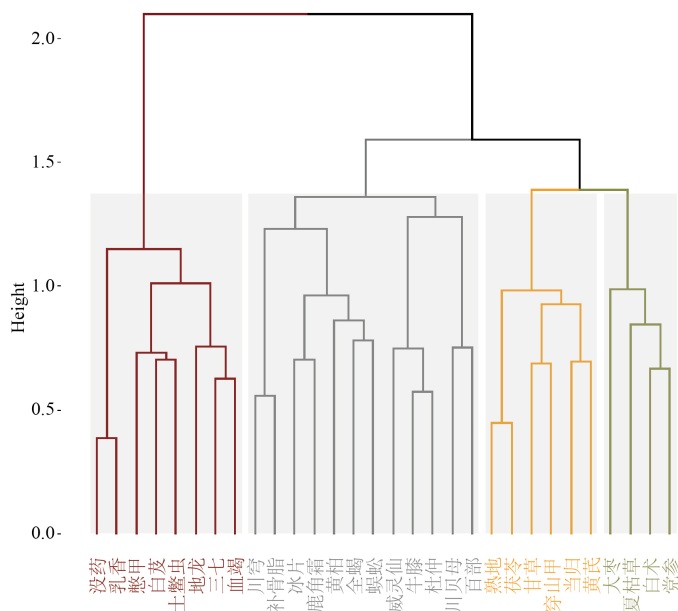


Figure 5. Spectrum diagram of four types of high-frequency drugs

图 5. 高频药物 4 类药物谱系图

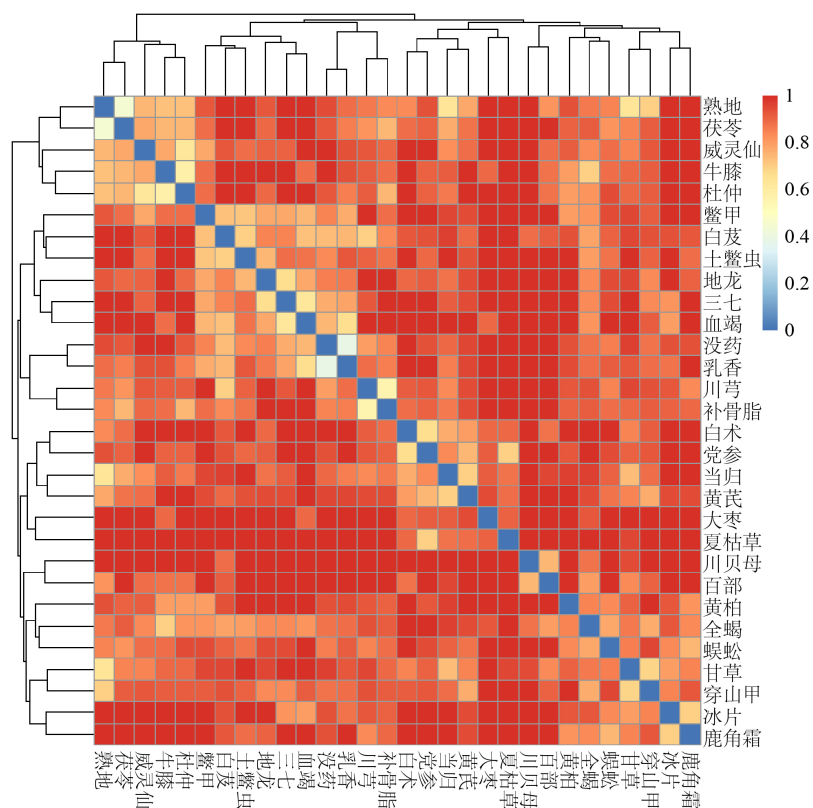


Figure 6. High frequency drug clustering heatmap
图 6. 高频药物聚类热图

4. 讨论

中医学认为骨结核属于“骨痹”“流痰”范畴，统属“阴疽”，现代医家认为本病早期多与寒痰、血瘀及癆虫等因素有关，而中期主要是化热成脓等实证，后期主要为气血与脾肾两亏之虚证。治宜祛痰通络、温阳补血、散寒通滞为法[11][12]。现代临床上应用西药保守治疗骨结核容易产生耐药[13]，而对于中后期的患者常需要配合手术治疗，且有术后再复发的可能[14]。而中医学在治疗骨结核方面具有独特的优势，在临床上广泛运用。现运用数据挖掘技术对未知的大量的中药数据探析其中隐藏的价值与特定联系的过程，能够为发现临床用药规律、指导用药等实际问题提供思路[15]。R 语言相较于传统数据挖掘软件操作灵活，且具有强大的数据可视化技术及多种个性化分析功能[16]-[18]。因此运用 R 语言初步探析现代治疗 AS 的用药规律，为中医辨证施治提供参考。

纳入研究的复方共 70 首专利复方，共包含药物 180 味，对纳入 70 首专利复方的药物进行性味归经统计，结果显示治疗骨结核的药物药性以温、寒、平性为主，药味以甘、辛、苦味为主；甘味能补能和能缓，能补益气血、强壮筋骨；辛味能散能行，能散里寒以除寒湿，能行气通经、活血化痰；苦能燥、能坚，能燥湿止痹痛。归经以肝、脾、肾经为主。肝主筋，肝血充足，则筋膜得以濡养，肾主骨，脾为后天之本，治疗上选择肝肾脾经药物为主。对于纳入药物的频数统计，用药频数大于 10 的药物分别是当归、甘草、全蝎、蜈蚣、黄芪、鳖甲、没药、乳香、穿山甲、白芨。

关联规则是反映药物之间的相互关联性，运用 R 语言对所纳中药进行关联规则分析，结合分组矩阵图与网状图表明当归、甘草、全蝎、蜈蚣、熟地、乳香、没药这几种药物常伴出现，综合得出治疗骨结核的核心药物为当归、甘草、全蝎、蜈蚣、熟地、乳香、没药。当归补血活血、调经止痛；甘草调和诸药；

全蝎、蜈蚣为虫药，善于攻毒散结，通络止痛；熟地黄补血滋阴，益精填髓；乳香能调气活血、定痛、消肿、生肌；没药能散瘀，消肿，生肌，祛风，止痛。诸药相合，共奏攻毒散结、补血活血、通络止痛的功效。反映出核心药物以补气血、调阴阳、祛邪止痛为主。对所纳入的高频中药进行聚类分析，综合分析后可将药物聚为 4 类并用谱系图展示，得到 4 组新方，组 1 方：乳香、没药、鳖甲、土鳖虫、地龙、三七、血竭；组 2 方：川芎、补骨脂、冰片、鹿角霜、黄柏、全蝎、蜈蚣、威灵仙、牛膝、杜仲、川贝母、百部；组 3 方：熟地、茯苓、甘草、穿山甲、当归、黄芪；组 4 方：大枣、夏枯草、白术、党参。以组 3 方为例，熟地补血滋阴、益精填髓，茯苓健脾，甘草调和诸药，穿山甲活血通络、消肿排脓，当归补血活血，黄芪补气活血。合方行补血活血，祛瘀通络，消肿排脓。聚类 15 类是将药物的组合更加细致，更能凸显药物之间的联系，并能通过聚类热图进行验证。

5. 结语

运用 R 语言数据挖掘中药治疗骨结核的组方用药规律，药物药性以温、寒、平性为主；药味以甘、辛、苦味为主；归经以肝、脾、肾经为主。虽然临床的证型常见的多为复杂的复合证型，但专利复方在辨证论治的条件下发挥着各自的优势。治疗骨结核的核心药物以当归、甘草、全蝎、蜈蚣、熟地、乳香、没药为主。治则上可以补血活血，祛瘀止痛，消肿排脓功效为主。可在核心药物上随证加减活血化瘀、补益肝肾类药物，在临床上可取得较好的疗效。本研究初步探究的用药规律对骨结核临床用药的选药提供了参考，为临床治疗骨结核提供用药的思路。

参考文献

- [1] 李玮, 严文琪. 骨关节结核病 736 例回顾性分析[J]. 医学综述, 2020, 26(4): 815-818.
- [2] Satchidanandam, V., Kumar, N., Jumani, R.S., Challu, V., Elangovan, S., et al. (2014) The Glycosylated Rv1860 Protein of *Mycobacterium tuberculosis* Inhibits Dendritic Cell Mediated TH1 and TH17 Polarization of T Cells and Abrogates Protective Immunity Conferred by BCG. *PLoS Pathogens*, **10**, e1004176. <https://doi.org/10.1371/journal.ppat.1004176>
- [3] 金阳辉, 石仕元, 赖震. 骨与关节结核的诊断研究进展[J]. 浙江中西医结合杂志, 2016, 26(8): 772-775.
- [4] 李素英, 曹振华, 王冬云, 等. 慢性骨髓炎、骨结核合并营养不良的营养治疗(附 21 例临床报告)[J]. 锦州医学院学报, 1989(4): 235-237.
- [5] 嵇辉, 杨增敏, 芮敏劼, 等. 基于《金匱要略》虚劳理论探析骨与关节结核的辨治思路[J]. 广西中医药, 2020, 43(3): 47-48.
- [6] 杨金录, 杨帆. 中医药治疗骨与关节结核[J]. 中国骨伤, 1995(3): 32-33.
- [7] 刘宇飞, 安甜, 王春晖, 等. 阳和汤的临床应用现状[J]. 中华中医药杂志, 2018, 33(10): 4554-4556.
- [8] 贺凌宇, 曾蕾, 李莉. 基于 R 语言数据挖掘技术探讨国家专利复方治疗多囊卵巢综合征的用药规律[J]. 世界中西医结合杂志, 2023, 18(2): 244-248+255.
- [9] 沈丕安. 中华本草[M]. 上海: 上海科学普及出版社, 2017.
- [10] 樊雪鸣, 王柳丁, 申伟, 等. 基于 R 语言数据挖掘的中药治疗紧张型头痛的用药规律探析[J]. 中草药, 2021, 52(15): 4614-4625.
- [11] 梁健, 冯晶, 乔杰, 等. 阳和汤加味对早期胸腰椎结核患者炎症因子和骨代谢的影响[J]. 湖南中医药大学学报, 2022, 42(10): 1772-1776.
- [12] 蒋志升, 王香阁, 王艳敏, 等. 中西医结合治疗骨结核[J]. 河南医药信息, 2015(7): 51.
- [13] 秦世炳, 董伟杰, 周新华, 等. 正确理解和认识骨与关节结核诊疗的若干问题[J]. 中国防痨杂志, 2013, 35(5): 384-392.
- [14] 李元, 秦世炳. 脊柱结核手术后的复发因素分析及处理[J]. 中国防痨杂志, 2019, 41(4): 369-370.
- [15] 高毅超, 王凡, 郭晶. 数据挖掘技术在中医药领域中的应用概况[J]. 湖南中医杂志, 2019, 35(7): 182-185.

- [16] 刘钰, 余卓芮, 刘岱宁. R 语言可视化的优势及其在空间统计教学中的应用[J]. 高教论坛, 2020(5): 30-33.
- [17] 李杏, 刘涛, 肖建鹏, 等. R 语言在公共卫生领域的应用: Meta 分析[J]. 华南预防医学, 2020, 46(5): 568-570.
- [18] 崔尧, 李建军, 丛芳, 等. 基于 R 语言的数据可视化技术在脊髓损伤康复中的应用[J]. 中国数字医学, 2020, 15(10): 85-87.