

# 基于疗效加权FP-Growth算法的干燥综合征组方用药挖掘分析\*

丁弋美, 侯佳奇, 薛 鸢\*\*

(上海中医药大学附属岳阳中西医结合医院 上海 200437)

**摘要:**目的 以数据挖掘方法探索干燥综合征组方用药规律,并探讨加权FP-Growth算法在中药方剂挖掘中的应用,为中医临床用药分析提供新思路。方法 以中医临床疗效作为加权标准,以集对分析为基础计算疗效基数 $\rho$ ,以 $w = (\rho_{\text{对照组}} - \rho_{\text{实验组}}) / \rho_{\text{对照组}} + \delta$ 计算最终加权值,以加权FP-Growth算法为关联规则分析方法进行数据挖掘,分别计算支持度及提升度,对挖掘结果进行支持度和提升度的分析,综合得出分析结果。结果 最常用中药前5位为:麦冬、生地黄、甘草、白芍、北沙参。综合支持度及提升度分析挖掘结果后提出以麦冬、生地黄、北沙参、甘草为基本底方,丹参、玄参、黄芪为祛瘀,解毒,补气加减方向的代表。并挖掘出了两个值得关注的高支持度药团:补脾肾的茯苓-山茱萸-山药-丹皮药团及补气血的白芍-太子参-五味子-当归药团。结论 干燥综合征临床用药以养阴生津药为主,配伍具有经典方特色。加权FP-Growth算法用于组方用药分析可提炼核心组方思路及加减方向,有助于中医药经验的继承与发展。

**关键词:**干燥综合征 用药分析 关联规则 FP-Growth算法 集对分析

doi: 10.11842/wst.20210922006 中图分类号: R-058 文献标识码: A

干燥综合征是一种以侵犯自身外分泌腺为特征的全身免疫性疾病,其中泪腺及唾液腺累及最为常见,临床表现为以口干、眼干为代表的一系列干燥症状,此外还有全身疲劳、关节痛和多器官累及<sup>[1]</sup>。其发病率约为61/10万人,女性发病率高于男性。患者血清中可检测出自身免疫性抗体:抗Ro/SSA及抗La/SSB。病理显示唇腺内局灶淋巴细胞浸润<sup>[2]</sup>,伴BAFF、IFN- $\alpha/\gamma$ 等促炎性因子异常增多,从而引发持续性炎症反应,造成腺体不可逆破坏。

干燥综合征患者病情持久且反复,但目前尚无能根治该疾病的药物<sup>[3]</sup>,临床常用的糖皮质激素及免疫抑制剂<sup>[4]</sup>的疗效也缺乏循证医学支持。而中医药有着多靶点的独特优势,能同时兼顾口眼干燥,乏力及疼痛等多种临床症状,提高患者生活质量。有Meta分析显示中医药治疗有较好的临床疗效<sup>[5]</sup>,亦有研究发现

中药可以通过降低 $T_H1$ 及 $T_H2$ 的水平来改善干燥症状<sup>[6]</sup>。临床医家的理法方药各有特色,内容详实且富有变化,通过汇总性的数据挖掘可以分析其用药特点,探寻其组方脉络。目前针对临床方剂的数据挖掘多基于Apriori算法,用SPSS实现的关联规则挖掘,但这种方法难以引入加权项,不能满足临床立足于疗效的需要,目前在该方面的研究尚未引起学者的重视,存在较大发展空间。本研究以知网数据库及PubMed平台收录文献为基础,以疗效加权进行基于FP-Growth算法的干燥综合征中医用药规律的挖掘,有助于提取临床经验及归纳总结证治规律。

## 1 思路与方法

### 1.1 文献检索思路

(1)中文检索式:(SU='干燥综合征'+ 'SS') AND

收稿日期:2021-09-22

修回日期:2022-08-25

\* 上海市卫生和计划生育委员会科研课题项目(20184Y0082):解毒通络生津方对原发性干燥综合征唾液代谢组学的影响,负责人:侯佳奇。

\*\* 通讯作者:薛鸢,主任医师,博士生导师,主要研究方向:中西医结合治疗风湿病研究。

(SU='中医'+ '中西医'+ '中药'+ '方'+ '散'+ '汤'+ '饮'+ '膏'+ '丸'+ '中草药')。

(2) 英文检索式 : (Sjögren syndrome) AND ((traditional Chinese medicine) OR (herb))。

(3) 纳入标准: ①以中药或中药结合其他治疗方法治疗干燥综合征的临床研究; ②论文中有明确方剂组成者; ③设立对照组且有临床疗效数据及结论者; ④发表时间介于2005年1月-2021年11月。

(4) 排除标准: ①纯动物实验研究; ②治疗采用中成药或方剂具体组成未见者; ③个案或无疗效数据者; ④期刊论文与硕博论文内容重合者。

### 1.2 关联规则方法

本研究选择FP-Growth算法作为基础, FP-Growth根据每一个项逐步构建频繁模式树, 通过对频繁模式树的分析挖掘得到项中具体元素间的关联规则。在构建频繁模式树路径时结合加权算法, 使疗效评价纳入挖掘规则中, 更符合研究需要。具体加权FP-Growth算法利用Python 3.9编程实现。

### 1.3 加权规则方法

本研究以疗效为加权基数, 取中医临床疗效为疗效评价指标, 中医临床疗效标准选择“干燥综合征疗效判定标准”, 采用尼莫地平法计算治疗前后的疗效指数: 疗效指数(N) = (治疗前评分 - 治疗后评分) / 治疗前评分 × 100%。以疗效指数划分治愈, 显效, 有效及无效。以集对分析作为权值计算方法进行疗效加权。具体计算方法为:

集对分析4元联系数  $\mu = a + bi + cj + dk$ , 其中a表示“治愈率”, b表示“显效率”, c表示“有效率”, d表

示“无效率”<sup>[7]</sup>。当a=b=c=d=1时, 最佳疗效4元联系数  $\mu = 1 + 1i + 1j + 1k$ , 计算该方疗效与最佳疗效的距离  $\rho$  值,  $\rho$  值越小疗效越好:

$$\rho = \sqrt[2]{(1-a)^2 + i(1-b)^2 + j(1-c)^2 + k(1-d)^2}$$

本次实验设计i=0.666, j=0.333, k=0, 分别计算治疗组的  $\rho_{\text{实验组}}$  及对照组的  $\rho_{\text{对照组}}$ , 取  $w = (\rho_{\text{对照组}} - \rho_{\text{实验组}}) / \rho_{\text{对照组}} + \delta$  作为最终权值,  $\delta$  为调整差值, 本研究  $\delta$  取0.15。

### 1.4 特殊情况处理方法

对于无中医临床疗效的研究, 其权值取平均值纳入研究。

对于研究中方剂用药的模糊项, 作如下处理: ①出现生甘草、甘草的计入甘草, 出现炙甘草的计入炙甘草; ②出现沙参、北沙参的计入北沙参, 出现南沙参的计入南沙参; ③出现地黄、生地黄的计入生地黄, 出现熟地黄的计入熟地黄; ④出现生黄芪、炙黄芪、黄芪的一并计入黄芪。

### 1.5 结果排序及综合分析方法

(1) 支持度排序: 支持度  $s = w_i / w_{\text{总}} = w_i / (w_{\text{均}} * k)$  ( $i \in [1, k]$ , k为纳入项数)

(2) 提升度排序: 在符合一定支持度范围的条件 下(本研究取  $s \geq 0.1$ ) 计算提升度

$$l = \frac{\text{有}x\text{药物时出现}y\text{药物的概率}}{\text{整体出现}y\text{药物的概率}} = \frac{s_{x,y}}{s_y} = \frac{s_{x,y}}{s_x * s_y}$$

(3) 综合分析方法: 首先在高支持度的药物组中 筛选提升度小于1的药物作为第一层次结果, 再分析与第一层次结果中同时有较高支持度及提升度的药 物作为第二层次结果, 最后提取在支持度容许范围内

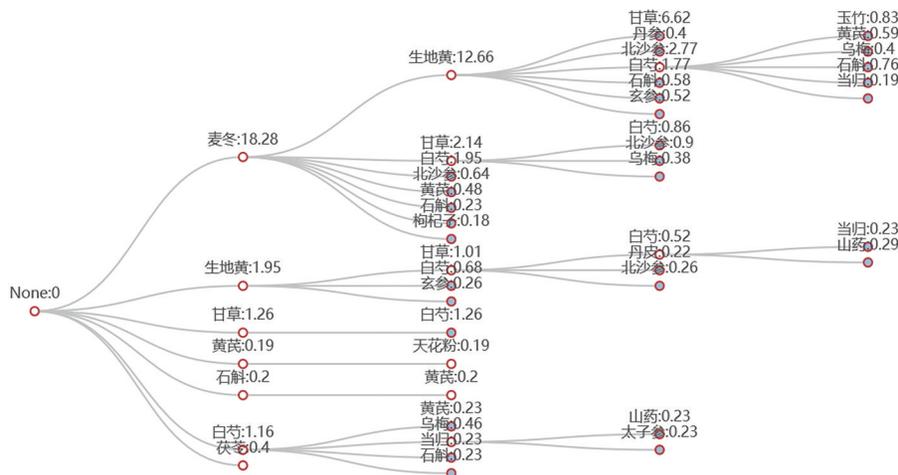


图1 加权频繁模式树构建示意图

( $s \geq 0.1$ )具有高提升度的药组作为第三层次结果。

## 2 数据挖掘结果

### 2.1 检索结果及纳入研究基本信息

根据检索式在知网共找到 1652 条结果,在 PubMed 共找到 121 条结果,根据纳入标准及排除标准,共纳入文献 102 篇,方剂 102 首,其中有中医临床疗效研究 82 项,无中医临床疗效研究 20 项。其中硕博论文 34 篇,期刊文献 68 篇。共来自 65 个研究机构的 93 位不同研究者,研究机构包括 11 所高校及 54 家医院,分布于全国 21 个省级行政区。

表 1 单药挖掘结果(支持度前 10)

序号	药物	加权频数	加权频率
1	麦冬	18.12	0.77
2	生地黄	14.60	0.62
3	甘草	11.03	0.47
4	白芍	10.63	0.45
5	北沙参	9.54	0.40
6	石斛	7.38	0.31
7	玄参	7.30	0.31
8	黄芪	7.25	0.31
9	丹参	6.97	0.29
10	当归	6.81	0.29

### 2.2 加权频繁模式树构建

加权频繁模式树是加权 FP-Growth 算法产生支持度排序的基础,图 1 展示了加权 FP-Growth 算法在 Python 3.9 中构建频繁模式树的过程。

### 2.3 单药加权分析结果

加权 FP-Growth 算法显示治疗干燥综合征最常用的前十味中药是麦冬、生地黄、甘草、白芍、北沙参、石斛、玄参、黄芪、丹参、当归。味以甘(80%)苦(50%)为主,性偏寒凉(70%)。入脾心、肺、胃、肝、肾、经,以心、脾两经为主。功擅养阴(60%),兼能清热(40%),凉血(30%)。具体加权频数及加权频率见表 1。

### 2.4 药对/药组支持度及提升度分析结果

经加权 FP-Growth 算法得出的药对及药组支持度排序结果见表 2。经加权 FP-Growth 算法得出药对支持度后,取  $s \geq 0.1$  的药对进行提升度排序的结果见表 3。药物间以支持度加权的网络关系见图 2。

### 2.5 药物关联分析结果

综合分析加权 FP-Growth 算法下支持度及提升度的结果,药物关联结果显示 3 个层次。

第 1 层是综合高支持度及高提升度药对构成基础方。综合分析发现麦冬、生地黄、北沙参、甘草四药之间支持度高(麦冬-生地黄  $s=0.53$ ;麦冬-北沙参  $s=$

表 2 药对/药组支持度排序结果(支持度前 40)

序号	药对	支持度	序号	药对	支持度
1	麦冬,生地黄	0.53	21	麦冬,枸杞子	0.21
2	麦冬,甘草	0.37	22	麦冬,甘草,玄参	0.21
3	麦冬,北沙参	0.37	23	生地黄,当归	0.20
4	生地黄,甘草	0.32	24	生地黄,丹参	0.20
5	麦冬,白芍	0.30	25	生地黄,石斛	0.20
6	麦冬,玄参	0.29	26	生地黄,枸杞子	0.20
7	麦冬,生地黄,甘草	0.28	27	麦冬,生地黄,黄芪	0.20
8	生地黄,北沙参	0.27	28	生地黄,甘草,玄参	0.19
9	麦冬,生地黄,北沙参	0.26	29	麦冬,乌梅	0.19
10	麦冬,黄芪	0.26	30	麦冬,生地黄,石斛	0.19
11	麦冬,石斛	0.25	31	麦冬,甘草,北沙参	0.18
12	生地黄,玄参	0.25	32	甘草,丹参	0.18
13	麦冬,丹参	0.24	33	麦冬,生地黄,甘草	0.18
14	麦冬,生地黄,玄参	0.23	34	麦冬,生地黄,白芍	0.18
15	生地黄,白芍	0.23	35	麦冬,生地黄,丹参	0.18
16	甘草,北沙参	0.22	36	麦冬,玉竹	0.18
17	麦冬,当归	0.22	37	麦冬,山药	0.17
18	生地黄,黄芪	0.22	38	白芍,当归	0.17
19	甘草,玄参	0.22	39	北沙参,玄参	0.17
20	甘草,白芍	0.21	40	白芍,丹参	0.17



数。仍存在的问题如下:①存在少数文献仅有实验室指标作为疗效标准而缺少临床改善指标,对于此类文献只能以平均疗效作为加权基数;②中医临床疗效不能替代整体疗效,疗效的评价应纳入更多指标,但不同文献指标选取相异,限制了指标选取的范围;③同种指标存在细节区别,如本研究选择的中医临床疗效,不同文献中治愈率,显效率,有效率和无效率标准有所差距,部分以疗效指数 $\geq 80\%$ 为治愈,而部分以疗效指数 $\geq 95\%$ 为治愈。本研究通过对对照组与实验组的比值作为加权基数,尽量避免不同文献间指标细微差异造成的影响;④除疗效外,研究本身的质量也应纳入加权的因素,予高质量研究高权值,低质量研究低权值以进一步优化评价体系。

### 3.1.3 权值量化方法

量化方法是权值的具体计算方式。疗效指标的量化关键在于在最终权值基数中尽可能多保留疗效信息。本研究选择以集对分析为理论基础,通过计算与最佳疗效的距离对疗效指标进行量化。具体方法见 1.3 加权规则,其中选择  $i=0.666, j=0.333$  能够显示治愈,显效和有效间的差别,从而使同等总有效率的疗效得以区分。

为了降低研究间的差异造成的偏倚,对于有对照组的文献,在加权基数  $\rho$  的基础上,以  $w$  作为最终的权值。取差值比  $w_1 = (\rho_{\text{对照组}} - \rho_{\text{实验组}}) / \rho_{\text{对照组}}$ , 绝对差值  $w_2 = \rho_{\text{对照组}} - \rho_{\text{实验组}}$ , 变化率  $= (95\% \text{ 置信区间上值} - \text{下值}) / \text{均值}$ , 结果见表 4。经检验两组  $w$  值都符合正态分布(两组  $P < 0.05$ ), 差值比组为相对数而绝对差值组为绝对数, 相对数更有助于消除不同文献间方法细节差异的偏倚, 所以本研究选择差值比为加权标准。但  $w_1$  的变化率过大, 可能导致部分文献的贡献被忽视, 故进行均值调整, 取调整差值比  $w_3 = (\rho_{\text{对照组}} - \rho_{\text{实验组}}) / \rho_{\text{对照组}} + \delta$ , 使变化率在较为合适的范围内, 取调整差值为 0.15 时, 变化率为 84.1%, 作为最终权值。

### 3.1.4 挖掘排序分析

加权支持度排序得到的结果主要反映药对出现的频率, 即常用药组成的药对出现概率高。提升度排序则反映药物之间的关联性强弱, 提升度高的药对具有配伍意义的可能性更大。但部分小样本会有极高的提升度, 所以在分析时必须综合结果项的支持度与提升度, 本研究取  $s=0.1$  为界限进行提升度分析及排序。

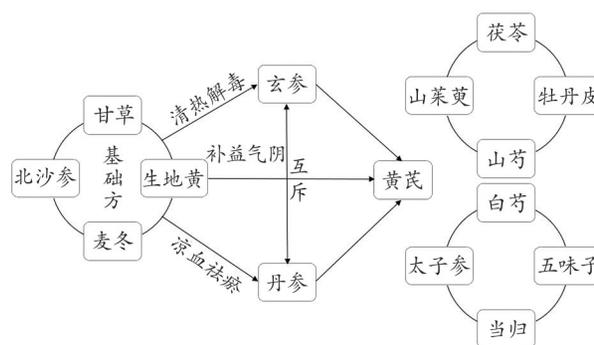


图3 组方思路示意图

表4 加权基数  $w$  对比

项目	差值比 $w_1$	绝对差值 $w_2$	调整差值比 $w_3$
95% 区间	0.010-0.205	0.012-0.211	0.160-0.355
均值	0.082	0.087	0.232
变化率	238%	229%	84.1%

## 3.2 挖掘结果的讨论

### 3.2.1 单药结果讨论

麦冬有养阴生津, 润肺止咳之功, 《神农本草经》<sup>[10]</sup> 中记载: “味甘, 平。主治心腹气结, 伤中, 伤饱, 胃络脉绝, 羸瘦, 短气。” 《景岳全书》<sup>[11]</sup> 言之其可“补上焦之津液, 清胸膈之渴烦。” 现代药理学研究发现麦冬中有效成分主要包括甾体皂苷类、高异黄酮类、多糖类等, 沙明慧<sup>[12]</sup> 研究发现麦冬多糖可通过 TLR4/MyD88/NF- $\kappa$ B 信号通路调控 AQP5 而改善干燥综合征大鼠颌下腺炎症。地黄具有清热凉血, 养阴生津的功效, 地黄中地黄多糖具有调节免疫, 抗疲劳的作用<sup>[13]</sup>, 可能与缓解干燥综合征症状有关。甘草具有补脾益气、清热解毒的功效, 甘草中的甘草三萜皂苷类和甘草黄酮类证实具有较好的抗炎作用, 可通过调节 NF- $\kappa$ B, TNF- $\alpha$  等因子发挥抗炎作用<sup>[14]</sup>。北沙参能养阴清肺, 益胃生津, 北沙参中的北沙参粗多糖具有免疫调节作用, 北沙参二氯甲烷提取物具有抗炎作用<sup>[15]</sup>。白芍具有敛阴止汗, 养血调经的作用, 其中的单萜及其苷类成分是其主要有效成分<sup>[16]</sup>, 杜旭召等<sup>[17]</sup> 的研究表明白芍总苷可能通过抑制 NF- $\kappa$ B 通路进而降低 MMP-13/TIMP-1 比值以及炎症因子的分泌从而发挥抗炎作用。梅寒颖等<sup>[18]</sup> 实验发现白芍总苷可能通过调节 TLR4/MyD88/NF- $\kappa$ B 信号通路活性而抑制炎症因子。

因此, 中医治疗干燥综合征常用药物以滋阴润燥为主, 兼有清热、凉血、补气之功, 与燥痹阴生津亏, 兼有气虚血瘀的病机契合。从现代医学及药理学角度

看,此类中药多具有抗炎,调节免疫的作用,其具体通路及调节机制有待进一步研究。

### 3.2.2 药物关联研究结果讨论

第1层次的基础方中麦冬、生地黄、北沙参、甘草四味药性味皆甘,功擅养阴生津,其中麦冬及北沙参入肺胃,润泽上中二焦,生地黄则能入肝肾经,以资下焦之水源,再配伍甘草补益中气兼能调和诸药,四药合用上养肺阴,中补胃阴,下滋肾阴,符合干燥综合征最主要的病机“津亏”的治疗。

第2层次以玄参、丹参、黄芪为3个加味代表。玄参体现了清热解毒的思路,它归脾、胃、肾经,能清热凉血、解毒散结。玄参、麦冬、生地黄即传统养阴增液名方增液汤。增液汤本为治疗阳明热结,阴津大伤,燥屎不下而立的“增水行舟”之法,其清热生津力强。而玄参-甘草药对又共能清热解毒。这一加味是在以润为主的底方上增强清解之力,针对燥毒的病机进行治疗。丹参则体现了活血化瘀的思想,其归心肝经,可凉血祛瘀止痛,温病中常用之清营凉血,而与生地黄相配可加强凉血之功,针对阴虚络瘀的病机进行治疗。黄芪则为补气养阴的代表,其味甘,性微温,入肺脾二经,《本草纲目》<sup>[19]</sup>中言黄芪“为补药之长”。不少干燥综合征患者乏力明显,这部分患者常辨证为气阴两虚,方中加黄芪正是针对此病机。黄芪虽为温补之药,但在大队滋阴药中配伍黄芪,有“善补阴者,必于阳中求阴,则阴得阳升而泉源不竭”之妙。此外丹参-玄参药对的提升度小于1,即二药存在一定程度互斥。丹参与玄参皆有凉血之功,但丹参兼能祛瘀,玄参兼能解毒,说明在此二味药的选择上体现组方偏向性的差异。

第3层次为两个高支持度药团。茯苓-山茱萸-山药-丹皮药团源于经典名方六味地黄丸中的两补两泻。其中山茱萸补肝肾,山药补脾肾,茯苓淡渗利湿兼能健脾,丹皮清泻虚热兼能凉血,体现了补泻并用

以滋肾阴的治疗思路。这一药团较之六味地黄丸少熟地黄及泽泻二药,可能是由于熟地黄味厚过于滋腻而泽泻利水作用较强不利于津液留存。另一个白芍-太子参-五味子-当归药团,其核心白芍本身作为支持度排序第四的高频药物,与前五位中麦冬、生地黄、北沙参都有较明显的互斥(提升度均 $<0.85$ ),表明该药团体现的治法与基础方有所差异。其中当归配伍白芍养血滋阴,辛散配酸收,一开一合,动静相宜。太子参加五味子益气生津,与基础方相比清解滋润之力稍弱,而偏向调补气血以生津,体现了不同的用药思路。

## 4 小结

本文在传统关联规则挖掘基础上引入疗效作为加权基数,将疗效这一重要指标纳入分析,更符合中医传承的需要。并且在药物关联结果基础上,进一步分析数据结构并结合理论提出基础方和加味方向,以及发掘体现不同处方思路的药团。这一分析方法既可用于综合不同医家的遣方用药,也可用于提炼个人组方的心得,为临床方剂学的研究提供了新思路,亦是提取临床经验的重要探索。

本文的研究方法尚有多处细节值得结合更多数据深入探索,分为挖掘算法及结果分析两方面。挖掘算法中含有如下内容:①综合权值应纳入的指标及计算方法;②权值变化率适宜范围;③支持度最低基线;④支持度及提升度的权衡标准;⑤多味草药构成的药组如何计算提升度。结果分析中包括如下内容:①具有高支持度但提升度 $<1$ 的药对存在着互斥,这种互斥因素的实际意义;②具有高提升度而在支持度基线边缘的药对,可能代表某些特定配伍,可以进一步结合文献进行探讨;③处于支持度基线之下的药对在计算提升度前已经被舍去,但这些药对可能反映某些医家或地方特色性的用药规律。

## 参考文献

- Ramos-Casals M, Brito-Zeron P, Siso-Almirall A, et al. Primary sjogren syndrome. *BMJ*, 2012, 344:e3821.
- Stefanski A L, Tomiak C, Pleyer U, et al. The diagnosis and treatment of sjogren's syndrome. *Dtsch Arztebl Int*, 2017, 114(20):354-361.
- 干燥综合征诊断及治疗指南. 中华风湿病学杂志, 2010, 14(11): 766-768.
- Ramos-Casals M, Brito-Zerón P, Bombardieri S, et al. EULAR recommendations for the management of Sjögren's syndrome with topical and systemic therapies. *Ann Rheum Dis*, 2020, 79(1):3-18.
- Liu J J, Zhou H M, Li Y P, et al. Meta-analysis of the efficacy in treatment of primary sjögren's syndrome: Traditional Chinese Medicine vs Western Medicine. *J Tradit Chin Med*, 2016, 36(5):596-605.
- Lee G A, et al. Chinese herbal medicine SS-1 inhibits T cell activation and abrogates  $T_H$  responses in Sjögren's syndrome. *J Tradit Chin Med*, 2021, 120(1):651-659.
- 罗月, 蒯伟, 茹意, 等. 基于集对分析加权疗效曲线的银屑病代表方

- 剂用药规律研究. 辽宁中医杂志, 2019, 46(9):1825-1829.
- 8 毕建欣, 张岐山. 关联规则挖掘算法综述. 中国工程科学, 2005, 7(4):88-94.
- 9 俞立平. 客观赋权法本质及在科技评价中的应用研究——以学术期刊为例. 情报理论与实践, 2021, 44(2):50-56.
- 10 顾观光重. 神农本草经. 北京: 人民卫生出版社, 1956:31.
- 11 明·张介宾. 景岳全书. 影印本. 北京: 人民卫生出版社, 1991:1184.
- 12 沙明慧. 麦冬多糖基于TLR4/MyD88/NF- $\kappa$ B信号通路调控干燥综合征颌下腺炎性微环境的机制. 南京: 南京中医药大学硕士学位论文, 2019:12-16.
- 13 陈金鹏, 张克霞, 刘毅, 等. 地黄化学成分和药理作用的研究进展. 中草药, 2021, 52(6):1772-1784.
- 14 于鲁志. 中药甘草抗炎作用药理和临床研究进展. 光明中医, 2017, 32(19):2895-2898.
- 15 王晓琴, 苏柯萌. 北沙参化学成分与药理活性研究进展. 中国现代中药, 2020, 22(3):466-474.
- 16 张燕丽, 田园, 付起凤, 等. 白芍的化学成分和药理作用研究进展. 中医药学报, 2021, 49(2):104-109.
- 17 杜旭召, 杨豪, 邓素玲, 等. 白芍总苷对骨关节炎软骨细胞增殖及分泌表达的影响. 中国骨质疏松杂志, 2016, 22(11):1375-1379.
- 18 梅寒颖, 刘炬, 汤曾耀. 基于TLR4/MyD88/NF- $\kappa$ B信号通路探讨白芍总苷抑制干燥综合征模型小鼠炎症的作用机制. 中药新药与临床药理, 2021, 32(9):1293-1299.
- 19 明·李时珍著清·本草纲目. 上海: 商务印书馆, 1930, 12:85-88.

## The Medication Regularity of Traditional Chinese Medicine in the Treatment of Sjögren Syndrome: An Exploration Based on Clinical-efficacy-weighted FP-Growth Algorithm

Ding Yimei, Hou Jiaqi, Xue Luan

(Yueyang Hospital of Integrated Traditional Chinese and Western Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai 200437, China)

**Abstract:** Objective To analyze the medication regularity of traditional Chinese medicine (TCM) in the treatment of Sjögren Syndrome through data mining and to explore the application of FP-Growth algorithm in the analysis of traditional Chinese medicine formulae, in order to afford new ideas for clinical practice. Methods Clinical effect was chosen as the basis of the weight and the weight was calculated by  $w = (\rho_{control} - \rho_{experimental}) / \rho_{control} + \delta$ . Then the data was analyzed by FP-Growth algorithm. The lift degree and The support degree was calculated. After synthesizing the lift degree and the support degree, a basic formula was proposed. Results The most frequently used Chinese medicine were *Ophiopogonis Radix*, *Rehmanniae Radix*, *Licorice*, *Paeoniae Radix Alba*, *Glehniae Radix*. The basic formula consisted of *Ophiopogonis Radix*, *Rehmanniae Radix*, *Licorice* and *Glehniae Radix*. *Salviae Miltiorrhizae Radix et Rhizoma*, *Astragal Radix* and *Scrophulariae radix* represented three tendencies in formula composition. Two important herb groups were proposed. One consisted of *Poria cocos*, *Corni Fructus*, *Dioscoreae Rhizoma* and *Moutan Cortex*. Another one consisted of *Paeoniae Radix Alba*, *Pseudostellariae Radix*, *Schisandrae Chinensis Fructus* and *Angelicae Sinensis Radix*. Conclusion Yin fluid medicines were most frequently used in the treatment of Sjögren Syndrome. FP-Growth algorithm could be used in the analysis of the medication regularity of TCM to improve the inheritance and development of TCM.

**Keywords:** Sjögren syndrome, Medication regularity of TCM, Association rules, FP-Growth algorithm, Set pair analysis

(责任编辑: 李青)