Original Article

Medication regularity of pulmonary fibrosis treatment by contemporary traditional Chinese medicine experts based on data mining

Suxian Zhang[1], Hao Wu[1], Jie Liu[2], Huihui Gu[1], Xiujuan Li[1], Tiansong Zhang[1,3]

[1]Department of Traditional Chinese Medicine (TCM), Jing'an District Central Hospital, Fudan University, Shanghai 200040, China; [2]Shandong University of Traditional Chinese Medicine, Jinan 250014, China; [3]Institute of Standardization, Institutes of Integrative Medicine, Fudan University, Shanghai 200040, China

Contributions: (I) Conception and design: T Zhang; (II) Administrative support: T Zhang; (III) Searching the literature: S Zhang, J Liu; (IV) Collection and assembly of data: S Zhang, J Liu, H Wu; (V) Data analysis and interpretation: T Zhang; X Li; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Tiansong Zhang. No. 259 Xikang Road, Shanghai 20040, China. Email: zhangtiansong@fudan.edu.cn.

Background: Treatment of pulmonary fibrosis by traditional Chinese medicine (TCM) has accumulated important experience. Our interest is in exploring the medication regularity of contemporary Chinese medical specialists treating pulmonary fibrosis.

Methods: Through literature search, medical records from TCM experts who treat pulmonary fibrosis, which were published in Chinese and English medical journals, were selected for this study. As the object of study, a database was established after analysing the records. After data cleaning, the rules of medicine in the treatment of pulmonary fibrosis in medical records of TCM were explored by using data mining technologies such as frequency analysis, association rule analysis, and link analysis.

Results: A total of 124 medical records from 60 doctors were selected in this study; 263 types of medicinals were used a total of 5,455 times; the herbs that were used more than 30 times can be grouped into 53 species and were used a total of 3,681 times. Using main medicinals cluster analysis, medicinals were divided into qi-tonifying, yin-tonifying, blood-activating, phlegm-resolving, cough-suppressing, panting-calming, and ten other major medicinal categories. According to the set conditions, a total of 62 drug compatibility rules have been obtained, involving mainly qi-tonifying, yin-tonifying, blood-activating, phlegm-resolving, qi-descending, and panting-calming medicinals, as well as other medicinals used in combination.

Conclusions: The results of data mining are consistent with clinical practice and it is feasible to explore the medical rules applicable to the treatment of pulmonary fibrosis in medical records of TCM by data mining.

Keywords: Pulmonary fibrosis; medical records of traditional Chinese medicine (TCM); medication regularity; data mining

Submitted Oct 31, 2017. Accepted for publication Feb 09, 2018.
doi: 10.21037/jtd.2018.03.11
View this article at: http://dx.doi.org/10.21037/jtd.2018.03.11

Introduction

Pulmonary fibrosis is a progressive and fatal lung disease characterized by excessive accumulation of extracellular matrix (ECM) and remodelling of the lung architecture. It is frequently associated with idiopathic interstitial pneumonia, collagen diseases and so on (1-4). To date, the prognosis of patients with diffuse pulmonary fibrosis is poor because of lack of effective treatments (1,3,5). In the past ten years, traditional Chinese medicine (TCM) respiratory scholars have attempted to use Chinese medicine therapy to treat this disease, achieving certain curative effects by developing Chinese medicine compounds as well as

studying various well-known Chinese medical cases. At present, scholars have studied the compatibility of Chinese traditional herbal formulae for pulmonary fibrosis (6,7), but data mining research on real-world clinical records in the treatment of pulmonary fibrosis has not been reported. This study intends to use data mining to discover and extract knowledge and explore the experience and medication regularity in the treatment of pulmonary fibrosis from clinical records provided by contemporary Chinese medical specialists so as to provide ideas and TCM methods for treatment of pulmonary fibrosis.

Methods

Diagnostic criteria

Referring to the relevant literature, the selected patient medical records meet at least one of the following conditions (1,5):

(I) Have typical clinical features: mainly for exertional dyspnoea, shortness of breath, cough, and other symptoms; Velcro rale, clubbing, cyanosis, and other signs; radiographic/computed tomographic findings such as diffuse wellability shadow, ground-glass opacity, reticular changes, restricted pulmonary ventilation disorder, and hypoxemia;
(II) Have a clear diagnosis via pathological examination.

Inclusion criteria

(I) In line with the diagnostic criteria, and the cause is unknown (idiopathic pulmonary fibrosis) or connective tissue diseases with pulmonary fibrosis;
(II) Prescription physician (deputy) director of Chinese medicine practitioners;
(III) Medical records that have complete descriptions of clinical symptoms, signs, auxiliary examination data, and other data, as well as a clear drug composition and dose, and that provide a clear evaluation of efficacy.

Exclusion criteria

(I) Occupational disease, drug-induced, infection (including tuberculosis), radiation and other specific causes of clear pulmonary fibrosis, chronic obstructive pulmonary disease/bronchiectasis associated with local fibrosis;
(II) Prescription physicians do not have senior titles;
(III) Incomplete data.

Database establishment

Data sources

Without language restrictions, we searched PubMed, EMBASE, SinoMed, Wanfang, CNKI, and VIP from inception to 10 March 2017, using a combination of keywords and MeSH terms for pulmonary fibrosis, interstitial lung fibrosis, interstitial lung disease, interstitial pneumonia, pharmacotherapy, TCM, traditional Chinese herbal, and alternative medicine. Additional searches included manual retrieval from various core Chinese medicine journals. The pulmonary fibrosis medical records in line with the inclusion criteria were selected from the literature.

Data exacting and inputting

Medical records that were in line with the inclusion criteria were extracted and divided into four parts, including Chinese medicine experts, patients with demographic characteristics, symptoms and information, and Chinese medicinal herbs. After indexing in the original text of the medical records, the relevant data were extracted and then input into a Microsoft Excel sheet, setting up a corresponding database. The database forms an association between the different data through data encoding.

Data cleaning

This study mainly analysed the Chinese medicinals database and cleaned the database, mainly to standardize herb names. For example, the "Shanyurou (Corin Fructus)" is unified as "Shanzhuyu (Corni Fructus)", "Dabei (Fritillariae Thunbergii Bulbus)" and "Zhebei (Fritillariae Thunbergii Bulbus)" are unified as "Zhebeimu (Fritillariae Thunbergii Bulbus)". The concept of splitting characteristics involved various schemes: firstly splitting the drug name "peach almond" into "Taoren (Persicae Semen)" and "Xingren (Armeniacae Semen Amarum)" and so on, secondly the drug taste "sweetness pungency" is split into "sweetness" and "pungency", thirdly "belongs to the lung and spleen " is split into "lung", "spleen", and finally the effect of "clearing and resolving heat-phlegm" is split into "clearing heat" and "resolving phlegm" and so on.

In the new database of Chinese medicinal herbs, according to the "Pharmacopoeia of the People's Republic of China" (8) (2015 edition), the nature and flavor, meridian tropism and efficacy of each drug are input and digitized. Among them, coldness, hotness, warmth and coolness and

other medicinal properties are valued according to Jiang's method (9); for sourness, bitterness, sweetness, pungency, saltiness, and other drug tastes, meridian tropism and efficacy, if a description of a drug is associated with a certain other drug, it is recorded as 1, otherwise as 0.

Two independent authors gathered and classified the data and crosschecked results after the data had been completed. Any disagreement was resolved by discussion and consensus.

Data mining

(I) Descriptive analysis: the frequency method was used to calculate the type and frequency of each Chinese medicinal;
(II) Cluster analysis: for Chinese medicinals used more than 30 times, cluster analysis was carried out by a hierarchical clustering algorithm using Ward's linkage method according to the nature, flavor, meridian tropism and main efficacy of medicinals (6,8);
(III) Association analysis: the compatibility rules of the couplet medicinal/ group medicinal were extracted according to the association rules analysis (6,10). We specified that the minimum level of support indicating that items are associated was 10%, the minimum confidence for rule generation was 50%, and the minimum lift was 1;
(IV) Link analysis: we used link analysis and other complex network analyses to establish a group of core medicinals (11,12), and found a new prescription;
(V) Data mining tools: descriptive analysis and cluster analysis were completed with Stata13.1, and association rules analysis and link analysis were completed by SAS® Enterprise Miner 4.3.

Interpretation of the results

The results of data mining must be interpreted and evaluated under the guidance of experts in the appropriate field of expertise to determine whether the knowledge discovery is valuable. Two Chinese medical experts used the Delphi method to interpret the results in this study.

Results

Search results

According to the method of retrieval, 38,364 articles were initially retrieved. By reviewing the literature titles, abstracts, and keywords, we obtained 346 pertinent articles. By reading their full text, 112 articles were identified according to the inclusion and exclusion criteria, which were all Chinese articles and were finally confirmed to include 124 medical cases involving 60 doctors, with a total of 328 records of treatment times.

Cluster analysis results

In the 124 medical records, 263 types of herbs were used for 5,455 counts of frequency; Chinese medicinals appearing more than 30 times could be grouped into 53 species, with 3,681 counts of frequency in total. According to the nature, flavor, meridian tropism, and main efficacy of medicinals, it was reasonable to group these into ten categories and nineteen sub-categories in line with the principles of TCM, as shown in *Figure 1* and *Table 1*.

Association rule analysis results

According to the set of principles for drug association analysis, a total of 62 rules, of which a total of 29 rules involve couplet medicinals and a total of 33 rules involve group medicinals resulted. By these principles, 17 rules contain Gancao (Glycyrrhizae Radix Et Rhizoma), which is considered a "Harmonizing medicinal" rather than a specific treatment medicinal, so these rules were deemed "uninteresting" and were also ignored. So, the interesting rules in couplet medicinals were 21 and in group medicinals were 24. Thus, *Table 2* shows the 45 association rules that were considered significant.

Link analysis results

The data contains 179 nodes and 2,542 links. We only show the results of links among all the main medicinals that were used in the cluster analysis in order to determine the core medicinals, as shown in *Figure 2*. It is found that a new prescription for the treatment of pulmonary fibrosis is composed of Huangqi (Astragali Radix), Dangshen (Codonopsis Radix), Maidong (Ophiopogonis Radix), Wuweizi (Schisandrae Chinensis Fructus), Dilong (Pheretima), Danshen (Salviae Miltiorrhizae Radix Et Rhizoma), Danggui (Angelicae Sinensis Radix), Zhebeimu (Fritillariae Thunbergii Bulbus), Banxia (Pinelliae Rhizoma), Taoren (Persicae Semen), Kuxingren (Armeniacae Semen Amarum), Ziwan (Asteris Radix Et Rhizoma), Gancao

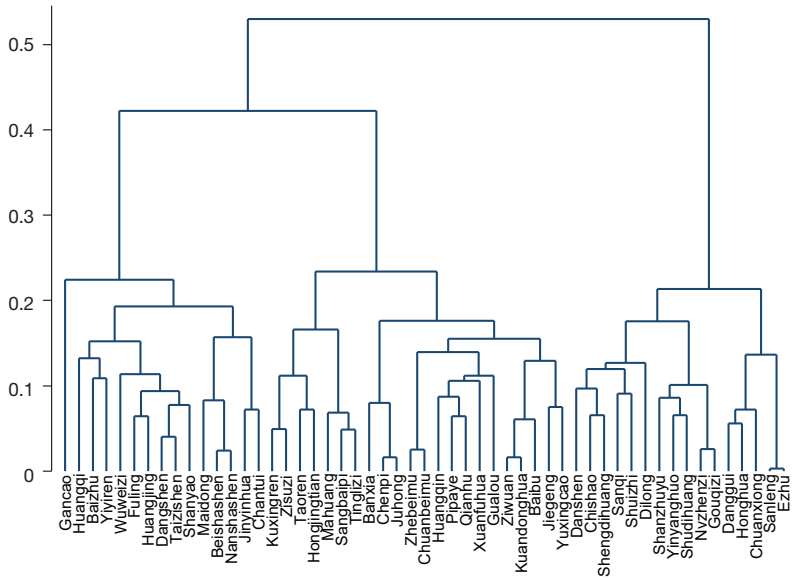

Figure 1 Cluster analysis results.

(Glycyrrhizae Radix Et Rhizoma) , and other components.

Discussion

As for the model of diagnosis and treatment based on diagnosis individuality and integration of diagnosis, medical records of TCM provide good evidence of Chinese evidence-based practice (13), which reflect the comprehensive use of TCM principles, methods, formulas and medicinal and not only are the real records of medical activities, but also reflect the clinical experience and thought processes of physicians (14). The recent sage Zhang Taiyan said "*the achievement of TCM is noticeable in medical records; in order to get the previous experience, medical records are the best approach to learn; dig into this and do more with less*". However, the schools of TCM are various and influence different physicians to use different medicinals. From the perspective of informatics, medical record data is complicated and intricate experiential data, where valuable information cannot be easily determined individually and should be studied with the assistance of new technology and methods. Data mining (15) is an effective tool to distil knowledge from a mass of incomplete and noisy data, as data mining is the most advanced data processing technology in the era of big data. So, applying data mining technology should bring improvement and further development of TCM academic technology (16). Therefore, we have explored data mining of pulmonary fibrosis medical records of qualified TCM doctors to evaluate the new technology, which plays a positive role in enriching critical thinking systems involving TCM treatments for pulmonary fibrosis.

Cluster analysis

Cluster analysis is the method of grouping a set of objects in a manner in which objects in the same group are more similar to each other than to those in other groups. As a common technique for statistical data analysis and a main unsupervised method of exploratory data mining, cluster analysis can automatically divide a data set into many types to discover the classification regularity implied in TCM clinical data (17).

After the cluster analysis of 53 main medicinals according to nature, flavor, meridian tropism and efficacy, we find that the acquired classification is consistent with clinical practice generally, but there are also some apparent errors. (I) Fuling (Poria) is grouped into the medicinals of boosting qi to engender fluid in the qi-tonifying medicinal category; (II) Yuxingcao (Houttuyniae Herba) is grouped as a cough-suppressing medicinal. By reviewing medical literature from past dynasties and modern pharmacological research, we find that the above classifications are actually reasonable: (I) although in the textbook "*Chinese materia medica*", Fuling (Poria) is grouped into inducing diuresis with bland drugs,

Table 1 Main medicinal classification and frequency

Category	Classification	Sub-categories	Count	Nature and flavor	Meridian tropism	Main efficacy
Category 1	Qi-tonifying medicinal	Huangqi (Astragali Radix)	206	Sweetness; slight warmth	Lung; spleen	Partial to tonify the spleen to promote diuresis
		Baizhu (Atractylodis Macrocephalae Rhizoma)	79	Bitterness; warmth	Spleen; stomach	
		Yiyiren (Coicis Semen)	50	Sweetness; bland; coolness	Lung; stomach	
		Wuweizi (Schisandrae Chinensis Fructus)	90	Sourness; sweetness; warmth	Lung; heart; kidney	Partial to boost qi and engender fluid
		Fuling (Poria)	85	Sweetness; bland; natural	Heart; lung; spleen; kidney	
		Dangshen (Codonopsis Radix)	83	Sweetness; natural	Spleen; lung	
		Taizishen (Pseudostellariae Radix)	67	Sweetness; slight bitterness; natural	Lung; spleen	
		Shanyao (Dioscoreae Rhizoma)	49	Sweetness; natural	Lung; spleen	
		Huangjing (Polygonati Rhizoma)	31	Sweetness; natural	Spleen; lung; kidney	
Category 2	Yin-tonifying and lung heat-clearing medicinal	Maidong (Ophiopogonis Radix)	119	Sweetness; slight bitterness; slight coldness	Heart; lung; stomach	Partial to nourish yin and engender fluid
		Beishashen (Glehniae Radix)	46	Sweetness; slight bitterness; slight coldness	Lung; stomach	
		Nanshashen (Adenophorae Radix)	42	Sweetness; slight coldness	Lung; stomach	
		Jinyinhua (Lonicerae Japonicae Flos)	67	Sweetness; coldness	Lung; heart; stomach	Partial to cleat heat
		Chantui (Cicadae Periostracum)	55	Sweetness; coldness	Lung; liver	
Category 3	Kidney-tonifying medicinal	Shanzhuyu (Corni Fructus)	75	Sourness; astringent; slight warmth	Liver; kidney	Partial to warm and supplement the liver and kidney
		Yinyanghuo (Epimedii Folium)	52	Pungency; sweetness; warmth	Liver; kidney	
		Shudihuang (Rehmanniae Radix Praeparata)	34	Sweetness; slight warmth	Liver; kidney	
		Nvzhenzi (Ligustri Lucidi Fructus)	37	Sweetness; bitterness; coolness	Liver; kidney	Partial to nourish the liver and kidney
		Gouqizi (Lycii Fructus)	31	Sweetness; natural	Liver; kidney	

Table 1 (*continued*)

Table 1 (*continued*)

Category	Classification	Sub-categories	Count	Nature and flavor	Meridian tropism	Main efficacy
Category 4	Blood-activating and stasis-dispelling medicinal	Danshen (Salviae Miltiorrhizae Radix Et Rhizoma)	125	Bitterness; slight coldness	Heart; liver; large intestine	Activate blood, dispel stasis
		Chishao (Paeoniae Radix Rubra)	56	Bitterness; slight coldness	Liver; spleen; bladder	
		Shengdihuang (Rehmanniae Radix)	55	Sweetness; coldness	Heart; liver; kidney	
		Sanqi (Notoginseng Radix Et Rhizoma)	40	Sweetness; slight bitterness; warmth	Liver; stomach	
		Shuizhi (Hirudo)	32	Saltiness; bitterness; natural	Liver	
		Dilong (Pheretima)	81	Saltiness; coldness	Liver; spleen; bladder	Clear heat, free the collateral vessels, calm panting
Category 5	Blood-activating and qi-moving medicinal	Danggui (Angelicae Sinensis Radix)	96	Sweetness; pungency; warmth	Liver; heart; spleen	Clear heat, move qi
		Chuanxiong (Chuanxiong Rhizoma)	92	Pungency; warmth	Liver; gallbladder; pericardium	
		Honghua (Carthami Flos)	58	Pungency; warmth	Heart; liver	
		Sanleng (Sparganii Rhizoma)	48	Pungency; bitterness; natural	Liver; spleen	Break blood, move qi
		Ezhu (Curcumae Rhizoma)	48	Pungency; bitterness; warmth	Liver; spleen	
Category 6	Panting-calming medicinal	Kuxingren (Armeniacae Semen Amarum)	152	Bitterness; slight warmth	Lung; large intestine	Partial to direct qi downward to calm panting
		Zisuzi (Perillae Fructus)	93	Pungency; warmth	Lung	
		Taoren (Persicae Semen)	90	Bitterness; sweetness	Heart; liver; large intestine	
		Hongjingtian (Rhodiolae Crenulatae Radix Et Rhizoma)	41	Sweetness; bitterness; natural	Lung; heart	
		Mahuang (Ephedrae Herba)	71	Pungency; slight bitterness; warmth	Lung; bladder	Partial to purge lung of pathogenic fire to calm panting
		Sangbaipi (Mori Cortex)	63	Sweetness; coldness	Lung	
		Tinglizi (Descurainiae Semen Lepidii Semen)	32	Pungency; bitterness; severe coldness	Lung; bladder	
Category 7	Dampness-drying and phlegm-resolving medicinal	Banxia (Pinelliae Rhizoma)	116	Pungency; warmth	Spleen; stomach; lung	Partial to direct qi downward
		Chenpi (Citri Reticulatae Pericarpium)	63	Bitterness; pungency; warmth	Lung; spleen	Partial to regulate qi
		Juhong (Citri Exocarpium Rubrum)	37	Pungency; bitterness; warmth	Lung; spleen	

Table 1 (*continued*)

Table 1 (*continued*)

Category	Classification	Sub-categories	Count	Nature and flavor	Meridian tropism	Main efficacy
Category 8	Heat-clearing and phlegm-resolving medicinal	Zhebeimu (Fritillariae Thunbergii Bulbus)	114	Bitterness; coldness	Lung; heart	Both the role of cough-suppressing
		Chuanbeimu (Fritillariae Cirrhosae Bulbus)	36	Bitterness; sweetness; slight coldness	Lung; heart	
		Huangqin (Scutellariae Radix)	77	Bitterness; coldness	Lung; gallbladder; spleen; large intestine; small intestine	Both the role of qi-descending
		Pipaye (Eriobotryae Folium)	61	Bitterness; slight coldness	Lung; stomach	
		Qianhu (Peucedani Radix)	46	Bitterness; pungency; coldness	Lung	
		Xuanfuhua (Inulae Flos)	32	Bitterness; pungency; saltiness; slight warmth	Lung; spleen; stomach; large intestine	
		Gualou (Trichosanthis Fructus)	31	Sweetness; slightly bitterness; coldness	Lung; stomach; large intestine	
Category 9	Cough-suppressing medicinal	Ziwan (Asteris Radix Et Rhizoma)	92	Pungency; bitterness; warmth	Lung	Partial to moisten the lung to direct qi downward
		Baibu (Stemonae Radix)	50	Sweetness; slight warmth	Lung	
		Kuandonghua (Farfarae Flos)	49	Pungency; slight bitterness; warmth	Lung	
		Jiegeng (Platycodonis Radix)	66	Bitterness; pungency; natural	Lung	Partial to dispel phlegm to drain pus
		Yuxingcao (Houttuyniae Herba)	32	Pungency; slight coldness	Lung	
Category 10	Harmonizing medicinal	Gancao (Glycyrrhizae Radix Et Rhizoma)	208	Sweetness; natural	Heart; lung; spleen; stomach	Tonify the spleen, boost qi, harmonize the actions of various ingredients in a prescription

Table 2 Medicinal compatibility rules based on association analysis

Relationships	Rules	Lift	Support (%)	Confidence (%)
Couplet medicinals	Dangshen (Codonopsis Radix) ==> Huangqi (Astragali Radix)	1.47	21.83	89.16
	Zisuzi (Perillae Fructus) ==> Kuxingren (Armeniacae Semen Amarum)	1.37	16.81	61.29
	Jinyinhua (Lonicerae Japonicae Flos) ==> Huangqi (Astragali Radix)	1.40	16.81	85.07
	Pipaye (Eriobotryae Folium) ==> Kuxingren (Armeniacae Semen Amarum)	1.86	15.04	83.61
	Danggui (Angelicae Sinensis Radix) ==> Danshen (Salviae Miltiorrhizae Radix Et Rhizoma)	1.44	15.04	53.13
	Wuweizi (Schisandrae Chinensis Fructus) ==> Maidong (Ophiopogonis Radix)	1.46	13.57	51.11
	Chantui (Cicadae Periostracum) ==> Kuxingren (Armeniacae Semen Amarum)	1.87	13.57	83.64
	Taizishen (Pseudostellariae Radix) ==> Maidong (Ophiopogonis Radix)	1.74	12.09	61.19
	Ezhu (Curcumae Rhizoma) ==> Huangqi (Astragali Radix)	1.37	11.80	83.33
	Shanzhuyu (Corni Fructus) ==> Wuweizi (Schisandrae Chinensis Fructus)	2.01	11.80	53.33
	Shengdihuang (Rehmanniae Radix) ==> Huangqi (Astragali Radix)	1.22	11.80	74.07
	Jiegeng (Platycodonis Radix) ==> Kuxingren (Armeniacae Semen Amarum)	1.25	10.91	56.06
	Sanleng (Sparganii Rhizoma) ==> Huangqi (Astragali Radix)	1.23	10.62	75.00
	Hongjingtian (Rhodiolae Crenulatae Radix Et Rhizoma) ==> Huangqi (Astragali Radix)	1.44	10.62	87.80
	Honghua (Carthami Flos) ==> Huangqi (Astragali Radix)	1.02	10.62	62.07
	Kuandonghua (Farfarae Flos) ==> Ziwan (Asteris Radix Et Rhizoma)	2.63	10.32	71.43
	Shanyao (Dioscoreae Rhizoma) ==> Huangqi (Astragali Radix)	1.18	10.32	71.43
	Chishao (Paeoniae Radix Rubra) ==> Danggui (Angelicae Sinensis Radix)	2.21	10.32	62.50
	Jiegeng (Platycodonis Radix) ==> Zhebeimu (Fritillariae Thunbergii Bulbus)	1.53	10.03	51.52
	Taizishen (Pseudostellariae Radix) ==> Danshen (Salviae Miltiorrhizae Radix Et Rhizoma)	1.38	10.03	50.75
	Chenpi (Citri Exocarpium Pericarpium) ==> Banxia (Pinelliae Rhizoma)	1.58	10.03	53.97

Table 2 (*continued*)

Table 2 (*continued*)

Relationships	Rules	Lift	Support (%)	Confidence (%)
Group medicinals	Jinyinhua (Lonicerae Japonicae Flos) ==> Zhebeimu (Fritillariae Thunbergii Bulbus) & Huangqi (Astragali Radix)	2.92	12.09	61.19
	Zhebeimu (Fritillariae Thunbergii Bulbus) & Jinyinhua (Lonicerae Japonicae Flos) ==> Huangqi (Astragali Radix)	1.57	12.09	95.35
	Zhebeimu (Fritillariae Thunbergii Bulbus) & Huangqi (Astragali Radix) ==> Jinyinhua (Lonicerae Japonicae Flos)	2.92	12.09	57.75
	Jinyinhua (Lonicerae Japonicae Flos) & Huangqi (Astragali Radix) ==> Zhebeimu (Fritillariae Thunbergii Bulbus)	2.14	12.09	71.93
	Huangqi (Astragali Radix) & Danggui (Angelicae Sinensis Radix) ==> Danshen (Salviae Miltiorrhizae Radix Et Rhizoma)	1.59	12.09	58.57
	Danggui (Angelicae Sinensis Radix) & Danshen (Salviae Miltiorrhizae Radix Et Rhizoma) ==> Huangqi (Astragali Radix)	1.32	12.09	80.39
	Maidong (Ophiopogonis Radix) & Kuxingren (Armeniacae Semen Amarum) ==> Huangqi (Astragali Radix)	1.03	11.80	62.50
	Maidong (Ophiopogonis Radix) & Huangqi (Astragali Radix) ==> Kuxingren (Armeniacae Semen Amarum)	1.14	11.80	51.28
	Pipaye (Eriobotryae Folium) ==> Kuxingren (Armeniacae Semen Amarum) & Chantui (Cicadae Periostracum)	4.47	10.91	60.66
	Chantui (Cicadae Periostracum) ==> Pipaye (Eriobotryae Folium) & Kuxingren (Armeniacae Semen Amarum)	4.47	10.91	67.27
	Pipaye (Eriobotryae Folium) & Kuxingren (Armeniacae Semen Amarum) ==> Chantui (Cicadae Periostracum)	4.47	10.91	72.55
	Pipaye (Eriobotryae Folium) & Chantui (Cicadae Periostracum) ==> Kuxingren (Armeniacae Semen Amarum)	2.17	10.91	97.37
	Kuxingren (Armeniacae Semen Amarum) & Chantui (Cicadae Periostracum) ==> Pipaye (Eriobotryae Folium)	4.47	10.91	80.43
	Huangqi (Astragali Radix) & Chuanxiong (Chuanxiong Rhizoma) ==> Danshen (Salviae Miltiorrhizae Radix Et Rhizoma)	1.54	10.91	56.92
	Danshen (Salviae Miltiorrhizae Radix Et Rhizoma) & Chuanxiong (Chuanxiong Rhizoma) ==> Huangqi (Astragali Radix)	1.30	10.91	78.72
	Kuxingren (Armeniacae Semen Amarum) & Danshen (Salviae Miltiorrhizae Radix Et Rhizoma) ==> Huangqi (Astragali Radix)	1.02	10.62	62.07
	Kuxingren (Armeniacae Semen Amarum) & Chuanxiong (Chuanxiong Rhizoma) ==> Huangqi (Astragali Radix)	1.44	10.32	87.50
	Huangqi (Astragali Radix) & Chuanxiong (Chuanxiong Rhizoma) ==> Kuxingren (Armeniacae Semen Amarum)	1.20	10.32	53.85
	Huangqi (Astragali Radix) & Chuanxiong (Chuanxiong Rhizoma) ==> Dangshen (Codonopsis Radix)	2.20	10.32	53.85
	Dangshen (Codonopsis Radix) & Chuanxiong (Chuanxiong Rhizoma) ==> Huangqi (Astragali Radix)	1.65	10.32	100.00
	Taizishen (Pseudostellariae Radix) ==> Maidong (Ophiopogonis Radix) & Kuxingren (Armeniacae Semen Amarum)	2.69	10.03	50.75
	Taizishen (Pseudostellariae Radix) & Maidong (Ophiopogonis Radix) ==> Kuxingren (Armeniacae Semen Amarum)	1.85	10.03	82.93
	Taizishen (Pseudostellariae Radix) & Kuxingren (Armeniacae Semen Amarum) ==> Maidong (Ophiopogonis Radix)	2.02	10.03	70.83
	Maidong (Ophiopogonis Radix) & Kuxingren (Armeniacae Semen Amarum) ==> Taizishen (Pseudostellariae Radix)	2.69	10.03	53.13

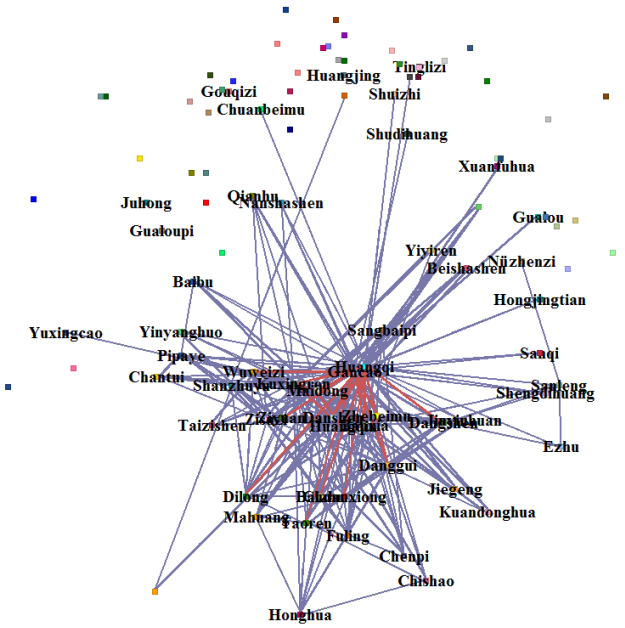

Figure 2 Schematic diagram of the core medicinals.

it is also used as a tonic for its function of tonifying qi and fortifying the spleen in practice. Since the Ming and Qing dynasty, doctors have regarded it as a tonifying spleen-yin drug. As said in *Yaopinhuayi*, "*Sweetness bland belong to earth which are used to tonifying spleen Yin; strong earth can engender metal and boost the lung qi*", and in *New Compilation of Materia Medica*, "*engendering fluid and humor…boosting the lung*", which proves it has the function of boosting qi to engender fluid; (II) although Yuxingcao (Houttuyniae Herba) is grouped as a heat-clearing and detoxicating medicinal, modern research proves that it indeed plays a role in cough-suppression (18).

According to the principle of "syndrome differentiation by effects of medicinals" of TCM, the pathogenesis of pulmonary fibrosis can be roughly inferred. This study found that of 263 Chinese medicinals used in the treatment of pulmonary fibrosis, qi-tonifying, blood-activating, lung fire-clearing, phlegm-resolving, cough-suppressing and panting-calming medicinals are used at a higher frequency, where tonic medicine is mainly qi-tonifying and yin-tonifying while meridian tropism mostly belongs to lung, spleen and kidney, reflecting that the majority of doctors are aware of their effect on the pathogenesis of pulmonary fibrosis (6,19): (I) the location of the disease is in the lungs, and is closely related with the spleen and kidneys; (II) the nature of the disease is a deficiency in origin and excess in superficiality, where insufficiency of the qi(yin) of lung, spleen and kidney belongs to deficiency in origin, and phlegm turbidity, static blood and heat toxin belong to excess in superficiality. The two influence each other and have reciprocal causation. Therefore, the TCM treatment is mainly to tonify the lung, fortify the spleen and boost the kidney to reinforce a healthy qi, and to activate blood and dispel stasis, resolve phlegm and clear heat in order to cure its superficiality. Regarding the lung governing qi and controlling respiration and other dispersing and descending physiological functions, the use of lung-qi dispersing and descending medicine to regulate qi is in line with clinical practice.

Couplet medicinal/group medicinal rule analysis

Association rules have been widely used to identify interesting relationships between item sets in large databases (20). Uncovered relationships can be represented in the form of association rules that will predict the occurrence of an item based on the occurrences of other items in the transaction.

When exploring couplet medicinals or group medicinals, we should combine the support, confidence, and lift to consider the association rules. Furthermore, the association rules for satisfying the user-defined minimum cut-off values of confidence and support are interesting. The lift reflects the closeness rating of the elements in the association rules, based on which >1 is meaningful. In the obtained medicinals association rules, all are interesting; but because the role of Gancao (Glycyrrhizae Radix Et Rhizoma), which is a "Harmonizing medicinal", has a large possibility of conjunction with other medicinals, we should regard the association rules between Gancao (Glycyrrhizae Radix Et Rhizoma) and other medicinals as uninteresting.

There are some interesting association rules that are commonly used by ancient and modern physicians for combinations of couplet medicinals. For example, qi-tonifying medicinal Dangshen (Codonopsis Radix)-Huangqi (Astragali Radix), yin-tonifying medicinal Taizishen(Pseudostellariae Radix)-Maidong (Ophiopogonis Radix) and blood-activating medicinal Chishao (Paeoniae Radix Rubra)-Danggui (Angelicae Sinensis Radix). Others include phlegm-resolving and cough-suppression medicinal Chenpi (Citri Exocarpium Pericarpium)-Banxia (Pinelliae Rhizoma), Zisuzi (Perillae Fructus)-Kuxingren (Armeniacae Semen Amarum), Jiegeng (Platycodonis Radix)-Kuxingren (Armeniacae

Semen Amarum) and Ziwan (Asteris Radix Et Rhizoma)-Kuandonghua (Farfarae Flos). These medicinals are Shengyu decoction, Shashen and Maidong decoction, Siwu decoction, Ezhu powder, Erchen decoction, Dingchuan decoction, Xingsu powder and Xiaochaihu decoction and other commonly used combination drugs.

On the other hand are the newly discovered drug couplet medicinals and group medicinals, for example, the qi-tonifying medicinal Huangqi (Astragali Radix)-Hongjingtian (Rhodiolae Crenulatae Radix Et Rhizoma) and Huangqi (Astragali Radix)-Shanyao (Dioscoreae Rhizoma), qi-yin-tonifying medicinal Huangqi (Astragali Radix)-Shengdihuang (Rehmanniae Radix), the qi-tonifying and blood-activating medicinals Huangqi (Astragali Radix)-Sanleng (Sparganii Rhizoma), Huangqi (Astragali Radix)-Ezhu (Curcumae Rhizoma), Huangqi (Astragali Radix)-Honghua (Carthami Flos), Huangqi (Astragali Radix)-Chuanxiong (Chuanxiong Rhizoma)-Dangshen (Codonopsis Radix), Huangqi (Astragali Radix)-Chuanxiong (Chuanxiong Rhizoma)-Danshen (Salviae Miltiorrhizae Radix Et Rhizoma) and Huangqi (Astragali Radix)-Danggui (Angelicae Sinensis Radix)-Danshen (Salviae Miltiorrhizae Radix Et Rhizoma), the qi-yin-tonifying and blood-activating medicinal of Taizishen (Pseudostellariae Radix)-Dangshen (Codonopsis Radix), the qi-yin-tonifying and cough-suppressing medicinal Huangqi (Astragali Radix)-Maidong (Ophiopogonis Radix)-Kuxingren (Armeniacae Semen Amarum), the qi-tonifying and blood-activating and cough-suppressing medicinal Huangqi (Astragali Radix)-Chuanxiong (Chuanxiong Rhizoma)-Kuxingren (Armeniacae Semen Amarum), the qi-tonifying and heat-clearing and phlegm-resolving medicinal Huangqi (Astragali Radix)-Zhebeimu (Fritillariae Thunbergii Bulbus)-Jinyinhua (Lonicerae Japonicae Flos) the heat-clearing and phlegm-resolving medicinals Zhebeimu (Fritillariae Thunbergii Bulbus)-Jiegeng (Platycodonis Radix) and Kuxingren (Armeniacae Semen Amarum)-Chantui (Cicadae Periostracum), the blood-activating medicinal Chuanxiong (Chuanxiong Rhizoma)-Danshen (Salviae Miltiorrhizae Radix Et Rhizoma), and so on. Modern pharmacological studies have confirmed that Huangqi (Astragali Radix), Danshen (Salviae Miltiorrhizae Radix Et Rhizoma), Danggui (Angelicae Sinensis Radix), Chuanxiong (Chuanxiong Rhizoma), Huangqin (Scutellariae Radix) and extracts of these medicinals have certain anti-fibrosis functions (21-27). We find that the obtained association rules reflect boosting qi, tonifying yin, activating blood, clearing heat, resolving phlegm and other pulmonary fibrosis basic treatments have an independent use or concomitant uses.

Link analysis

In *Figure 2*, the thickness of a line reflects the correlation between two medicinals. The thicker the line is, the higher the correlation. It can be found that a new prescription for the treatment of pulmonary fibrosis consists of Huangqi (Astragali Radix), Dangshen (Codonopsis Radix), Maidong (Ophiopogonis Radix), Wuweizi (Schisandrae Chinensis Fructus), Dilong (Pheretima), Danshen (Salviae Miltiorrhizae Radix Et Rhizoma), Danggui (Angelicae Sinensis Radix), Zhebeimu (Fritillariae Thunbergii Bulbus), Banxia (Pinelliae Rhizoma), Taoren (Persicae Semen), Kuxingren (Armeniacae Semen Amarum), Ziwan (Asteris Radix Et Rhizoma), Gancao (Glycyrrhizae Radix Et Rhizoma). Huangqi (Astragali Radix) and Dangshen (Codonopsis Radix). These medicinals mutually reinforce each other to invigorate spleen and lung qi. Maidong (Ophiopogonis Radix) and Wuweizi (Schisandrae Chinensis Fructus) reinforce each other to nourish lung and kidney yin. Dilong (Pheretima), Danshen (Salviae Miltiorrhizae Radix Et Rhizoma) and Danggui (Angelicae are matched to activate blood and calm panting. Zhebeimu (Fritillariae Thunbergii Bulbus) and Banxia (Pinelliae Rhizoma) are matched to resolve phlegm. Taoren (Persicae Semen), Kuxingren (Armeniacae Semen Amarum) and Ziwan (Asteris Radix Et Rhizoma) are matched to suppress coughing and to calm panting. Gancao (Glycyrrhizae Radix Et Rhizoma) harmonizes the actions of various ingredients in a prescription. All the drugs in this prescription reflect tonifying of qi and yin, activating blood and resolving phlegm, as well as descending qi to provide panting-calming, which involve a basic method for the treatment of pulmonary fibrosis. This prescription can be regarded as a new finding by data mining and its effectiveness can be further tested by clinical or animal experiments.

Limitations of study

Firstly, the chosen cases are all published and have been successful, but may contain some bias; secondly, although having collected 124 cases, the number chosen is relatively low. This means that we may not have acquired complete findings.

Currently, the summary and inheritance of TCM professors' experiences lies in two directions. One relies on professors' oral or mental instructions and text

summarization; the other is to make the most of information technology and data mining as well as extracting the experience and academic thoughts of qualified doctors. It is believed that these two methods should be combined to reasonably analyse and apply the mined knowledge.

Acknowledgements

Funding: This study was supported by the Health and Family Planning Commission Project of Shanghai (No. 20114354), and the Traditional Chinese Medicine Clinical Key Specialist Construction Project of Shanghai Jing'an District (No. JA2016-Z005)

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

References

1. Raghu G, Collard HR, Egan JJ, et al. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. Am J Respir Crit Care Med 2011;183:788-824.
2. Trajano LA, Trajano ET, Lanzetti M, et al. Elastase modifies bleomycin-induced pulmonary fibrosis in mice. Acta Histochem 2016;118:203-12.
3. Todd NW, Luzina IG, Atamas SP. Molecular and cellular mechanisms of pulmonary fibrosis. Fibrogenesis Tissue Repair 2012;5:11.
4. Tabata C, Tabata R, Kadokawa Y. Thalidomide prevents bleomycin-induced pulmonary fibrosis in mice. J Immunol 2007;179:708-14.
5. Raghu G, Rochwerg B, Zhang Y, et al. An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline: Treatment of Idiopathic Pulmonary Fibrosis. An Update of the 2011 Clinical Practice Guideline. Am J Respir Crit Care Med 2015;192:e3-19. Erratum in: Am J Respir Crit Care Med 2015;192:644.
6. Zhang TS, Zhang S, Li XJ, et al. Data mining of the regularity of compound herbal formulae for pulmonary fibrosis. Chin J Inf Tradit Chin Med 2011;18:31-4.
7. Lyu XD, Liu C, Pang LJ, et al. Study on the TCM principle-method-recipe-medicines for treating idiopathic pulmonary fibrosis based on data mining. Chin J Tradit Chin Med Pharm 2016;31:2510-4.
8. National pharmacopoeia editorial board. pharmacopoeia of the People's Republic of China (2015 edition, Part1) . Beijing: China Medical Science and Technology Press 2015:3-385.
9. Jiang YG, Li L, Li RS, et al. Experiment on Data Mining in Compatibity Law of Spleen-stomach Prescriptions in TCM. Modernization of Traditional Chinese Medicine and Materia Medica-World Science and Technology 2003;5:33-7.
10. Li XJ, Zhang TS, Zhang S, et al. Exploring the law of drug treatment of elderly diseases based on data mining. Fujian Journal of Traditional Chinese Medicine 2009;40:38-40.
11. Zhu L, Xue H, Lin S, et al. The association analysis in 445 case of famous practitioner of Chinese medicine about asthma. J Jiangxi Univ Tradit Chin Med 2007;19:83-7.
12. Wang Y, Zhang R, Zhou X, et al. Comparison of General Rules and Individual Differences in Clinical Experience from the Veteran TCM Doctors. Modernization of Traditional Chinese Medicine and Materia Medica-World Science and Technology 2009;11:793-9.
13. Bian Y, Qiu JX. Reflections on the modernization of traditional Chinese medicine. Chin J Inf Tradit Chin Med 2011;18:1-2.
14. Wang YH, Lu JG, Liu T, et al. Study on knowledge discovery in traditional Chinese medical case records. Zhong Xi Yi Jie He Xue Bao 2007;5:368-72.
15. Helma C. Data mining and knowledge discovery in predictive toxicology. SAR QSAR Environ Res 2004;15:367-83.
16. Dong XY, Qi S, Jiang M, et al. Application of data mining in the research of traditional Chinese medicine. Global Traditional Chinese Medicine 2017;10:364-8.
17. Zhou X. Liu B, Zhang X, et al. Data Mining in real-world traditional Chinese medicine clinical data warehouse. In: Poon J, Poon S. editors. Data Analytics for Traditional Chinese Medicine Research. Springer, Cham 2014:189-213
18. He FY, Deng KW, Tang Y, et al. Pharmacological and immunotoxicological effects of Houttuynia cordata Thunb. and its preparation. Chin J Pharmacol Toxicol 2009;23:325-9.
19. Zhang J, Yan YJ. Research progress of traditional Chinese medicine treatment in treating pulmonary fibrosis. Journal of Shandong University of Traditional Chinese Medicine 2009;33:81-5.
20. Yang DH, Kang JH, Park YB, et al. Association rule mining and network analysis in oriental medicine. PLoS One 2013;8:e59241.
21. Luan ZQ, Cao WF. Research development of

astragalus in treating pulmonary fibrosis. Chinese Journal of Experimental Traditional Medical Formulae 2011;17:275-7.
22. Li JT, Wei SC, Liu YQ, et al. Effect of Astragalus polysaccharides on bleomycin-induced pulmonary fibrosis rats. China Journal of Traditional Chinese Medicine and Pharmacy 2011;26:2360-2.
23. Tian SX, Chen JM, Han YL, et al. Influence and mechanism of tanshinone on Bleomycin-induced pulmonary fibrosis in rats. World Chinese Medicine 2014;9:1647-50.
24. Guo G, Wang H. Effect of angelica sinensis on fibroblast in Bleomycin induced pulmonary fibrosis in rats. Medical Journal of Wuhan University 2007;28:560-3.
25. Yue XF, Wang K, Zhang YN. Progress in Studies on Pharmacological Effect of Ligustrazine. Journal of Yu Lin College 2005;15:17-20.
26. Ma RL, Wang WP, Jiang XG, et al. Effects of total flavonoids of scutellaria baicalensis georgi (TFSB) on bleomycin-induced pulmonary fibrosis in rats. Chinese Pharmacological Bulletin 2011;27:537-42.
27. Liu W, Chen XL, Liu JH, et al. The effect of baicalein on bleomycin-induced fibrosis in lungs of rats. Zhongguo Ying Yong Sheng Li Xue Za Zhi 2009;25:145-9.

Cite this article as: Zhang S, Wu H, Liu J, Gu H, Li X, Zhang T. Medication regularity of pulmonary fibrosis treatment by contemporary traditional Chinese medicine experts based on data mining. J Thorac Dis 2018;10(3):1775-1787. doi: 10.21037/jtd.2018.03.11

