

# Research and Reviews: Journal of Pharmacy and Pharmaceutical Sciences

## Recognition of Disease Comorbidity Medication Patterns Based on Network Motif Analysis

Di Chen<sup>1\*#</sup>, Jin Tian<sup>2#</sup>, Yuepeng Yao<sup>2</sup>, Songxing Du<sup>2</sup>, Jieyin Gao<sup>2</sup>, Rongjuan Guo<sup>2</sup>, Yun Wei<sup>2</sup> and Peng Lu<sup>1</sup>

<sup>1</sup>Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

<sup>2</sup>Dongfang Hospital, Beijing University of Chinese Medicine, Beijing 100029, China

#These authors contributed equally to this work

### Research Article

Received date: 15/06/2016

Accepted date: 07/07/2016

Published date: 12/07/2016

#### \*For Correspondence

Di Chen, Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China, Tel: +861062551575

**E-mail:** di.chen@ia.ac.cn

**Keywords:** Disease comorbidity, Network motif analysis, Traditional Chinese medicine, Disease-drug network, Blood stasis

#### ABSTRACT

Disease comorbidity is one common and important issue in medical practice. Special medical interventions will be required for patients under such conditions. Here, we proposed a network-based computation model to discover the comorbidity medication patterns by exploiting the electronic health records of a traditional Chinese medicine (TCM) hospital. One key step of this model is the estimation of three types of associations including disease-disease associations, disease-drug associations and drug-drug associations by statistical analyses. Based on these associations, a disease-drug network (DDN) was constructed. Then comorbidity medication patterns were identified from the DDN through random walk and network motif-based analysis. Taking the circulatory system diseases as examples, we applied this model to obtain the comorbidity relations among different diseases and recognize the corresponding medication patterns. As a result, we threw light on the comorbidities among different circulatory system diseases, and discovered that one disorder of the circulatory system may accompany with another lesion on the other location of the circulatory system. In addition, we identified some meaningful medication patterns which are consistent with TCM theory, for example, one blood-regulation agent alone or in accompany with another agent like a tonic can be used to treat comorbidities among different "blood stasis"-related diseases

### INTRODUCTION

Disease comorbidity refers to the condition that different diseases co-emerge in the same patient <sup>[1]</sup>, such as hypertension and diabetes <sup>[2,3]</sup>. Disease comorbidity is one common health issue in medical practice, and it is more likely to produce worse health outcomes than each single disease <sup>[1,4,5]</sup>. Discovering potential disease comorbidities is very important for the diagnosis <sup>[6,7]</sup>, treatments <sup>[8]</sup>, and preventions <sup>[9,10]</sup> for comorbidity conditions and can also provide some new insights into disease etiology and pathology <sup>[11]</sup>.

Generally speaking, investigations on disease comorbidity will need a large amount of clinical resources. The electronic health record (EHR) <sup>[12]</sup> data, which can provide diagnostic information on diseases or medical management for each patient and reflect the clinical practice patterns, is one such vital data resource. The EHRs have been applied to find novel disease comorbidities in some recent studies <sup>[13,14]</sup>. The disease comorbidity relations are mostly evaluated by co-occurrence statistics: disease pairs with co-occurrences larger than expected are regarded as potential comorbidities, and are often represented in the form of a disease network. For example, Hidalgo <sup>[13]</sup> constructed a disease phenotypic network by statistically detecting comorbidities from more than 30 million patients' medical reports. Two statistical measures, the relative risk (RR) and the  $\phi$ -correlation, were utilized in

this study to quantify the co-associations. Likewise, Roque<sup>[14]</sup> investigated on disease comorbidities by identifying significant co-occurring disease pairs from both the structured and free-text fields of 5543 patient records based on a comorbidity score similar to the RR.

In addition to the EHR-based investigations on disease comorbidities, great efforts have also been devoted to discover the biological or medical commonalities between different diseases, and network-based models have been created to identify molecular<sup>[15,16]</sup> or pathway-based associations<sup>[17]</sup> and interactions between phenotypically distinct diseases. A human disease network has been constructed by connecting diseases with shared genes based on the known disease-causing genes from OMIM<sup>[16]</sup>. Similarly, Li<sup>[17]</sup> tried to find disease associations based on shared pathways between different diseases. To obtain a more systematic understanding of the network interactions between complex diseases, Menche<sup>[15]</sup> employed a disease module which is a connected sub graph containing all relevant molecules for a disease within the interactome to represent each disease, and used the overlap between the modules to predict disease-disease relationships. In addition to the disease-associated molecules or pathways, a human symptoms disease network (HSDN) has been constructed to describe the similarity of two diseases considering their respective symptoms<sup>[18]</sup>.

These network-based studies have illustrated the disease associations from distinctive perspectives, covering genes, pathways, molecular interactome and symptoms, all of which can help to understand the mechanism of disease comorbidities. However, less attention has been paid to the medications for disease comorbidity: how should comorbidity conditions be treated, what is the difference between the treatments for patients with different types of comorbidities? Answers to these questions can help improve the clinical management of comorbidities<sup>[19]</sup>. Although disease treatments vary a lot between different patients considering their specific health conditions, there still will be consistent medication patterns for certain types of disease comorbidities<sup>[5,20,21]</sup>. Consequently, this study aims to reveal the underlying comorbidity medication patterns from the EHRs.

The comorbidities among different types of cardiovascular and cerebrovascular diseases, like hypertensive, stroke and coronary heart disease which are all belong to the diseases of circulatory system according to the ICD10 (International Classification of Diseases 10th Revision), are very common in clinical practices<sup>[22-25]</sup>. In this study, we considered the circulatory system as one illustration to identify the medication patterns for comorbidities among circulatory system diseases. First, we collected the standardized medical records for patients with circulation system diseases from the EHRs of a hospital. Next, three different types of associations: disease-disease associations, disease-drug associations, drug-drug associations were mined from the medical records. Taking the diseases and drugs as network nodes, the high-confidential associations as network edges, we constructed a disease-drug heterogeneous network. Finally, representative comorbidity medication patterns were recognized from the heterogeneous network to describe the medication strategies for comorbidities through a network motif-based analysis. Our results discover that different blood-regulating agents or one blood-regulating agent in accompany with other type agents like the tonics can be applied on patients with comorbidities among the hypertensive diseases, cerebrovascular diseases and ischemic heart diseases.

## MATERIALS AND METHODS

### Source Data and Preprocess

The EHRs were collected from the Dongfang Hospital, which is the second clinical medical college of Beijing University of Chinese Medicine (BUCM) and an A-level tertiary traditional Chinese medicine (TCM) hospital. Each record contains a unique patient ID, the date of visit, diagnosed diseases, and the corresponding prescribed drugs for this visit. All diseases in the EHRs were specified as the ICD10 codes. All drugs were represented by the unique-identifiers designated by this hospital. Only western drugs or traditional Chinese medicine patent prescriptions (TCMPPs) were retained for the study, and each TCMPP was regarded as one drug rather than its herb compositions. To obtain insights into the comorbidity therapeutic strategies for circulatory system diseases (ICD10 codes belonging to chapter IX, i.e., I00-I99), we only collected records in which the diagnostic disease ICD10 codes beginning with the character "I". At last, we obtained 448384 EHRs for 88153 patients, with 356 diseases and 1422 drugs in these records.

### There were 356 Disease-Drug Network

#### *Filtering rare diseases and drugs*

Diseases and 1422 drugs among all the 448384 clinical records. However, a part of diseases and drugs only appeared in very few records. These rare disease or drugs will not lead to prevalent co-occurrences with other diseases or drugs, so they were truncated to save time. Setting the cut-off as appearing in 10 among all the 448384 records, we retained 135 diseases and 1108 drugs for further study (**Tables 1 and 2** for the number of records each disease and drug has appeared in).

#### *Estimating associations*

This study tried to discover three different types of associations: disease-disease associations, drug-drug associations and drug-disease associations. To calculate different types of associations, the medical records were organized into two forms-diagnostic record (DR) and patient record (PR). One DR stores a list of diseases and a list of drugs which are recorded for one visit.

Each PR consists of all the diseases in one specific patient's all historical DRs. Disease-disease associations, which are proposed to mine disease comorbidity conditions, should provide suggestions on whether two diseases co-occur on the same patients; therefore, they were calculated based on the PRs. Otherwise, disease-drug associations and drug-drug associations were applied to discover whether a drug can be used to treat a disease or whether two drugs can be combined considering the doctors' diagnostic experience, so these two types of associations should be determined based on DRs (**Figure 1**).

No matter DRs-based or PRs-based evaluation, we utilized two measures to mine the large medical resource to estimate the associations for entity pairs. One is the chi-square test which can examine whether there is a significant difference between the expected and the observed frequencies [26]. For two entities A and B, a contingency table was constructed to describe the distribution for these two entities as shown in **Table 1**, where *a* is the number of records which don't contain A or B, *b* is the number or records which contain A but not B, *c* is the number of records containing B but not A, and *d* is the number of records containing both A and B. Then the *Chi-square* statistic is calculated as:

$$\chi^2 = \frac{n(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d)}$$

Where *n* represents the total number of records:  $n = a + b + c + d$ .

**Table 1.** A contingency table for entities A and B.

	A=0	A=1	Total
B=0	<i>a</i>	<i>b</i>	<i>a+b</i>
B=1	<i>c</i>	<i>d</i>	<i>c+d</i>
Total	<i>a+c</i>	<i>b+d</i>	<i>a+b+c+d</i>

One problem for Chi-square statistic is that small cell values may make the *Chi-square* test misleading [27]. In this study, we adopted a correction proposed by Yates [27] on the contingency tables with any cell value less than 5:

$$\chi^2_{corr} = \frac{n(|ad - bc| - 0.5n)^2}{(a + b)(c + d)(a + c)(b + d)} \quad \text{if } a < 5 \text{ or } b < 5 \text{ or } c < 5 \text{ or } d < 5$$

Entity pairs with high chi-square statistic values (i.e., low p-values) are often considered as "clinically significant" associations. However, there are two opposite conditions which may lead to large chi-squared values: 1. Two entities co-occur frequently in the medical records (*a*, *b*, *c* is small while *d* is very large); 2. Both entities are absent in most of the medical records (*b*, *c*, *d* is small while *a* is very large). Although the latter one can indicate one entity is dependent on the other, it will not represent meaningful disease comorbidity, drug combination or disease-drug relation if the co-occurrences are very rare. For example, if two diseases A' and B' are with a contingency table in which  $a=9999$ ,  $b=c=0$ ,  $d=1$ , then they will have a very large Chi-square value ( $\chi^2_{corr} = 2499.50$ ,  $p\text{-value}=0.0$ ), however, it is not persuasive enough to say they have comorbidity relations as they only co-occur once among 10000 records. Consequently, the co-occur rate (CR) was applied together with the *Chi-square* test. The CR for entities A and B is the ratio between the number of co-occurrences and the total number of records:

$$CR = \frac{d}{a + b + c + d}$$

There is no clear boundary for isolated entities and associated entities. Typically, we considered an association exists between two entities if they co-occur frequently in the same records substantially more than chance alone. Consequently, Entity pairs with significant Chi-square values ( $\chi^2 > 6.7$  and  $p\text{-value} < 0.01$ ) and with meaningful CR ( $CR > 2.0e-4$ ) were identified as potential associations which are mutual-dependent and not so rare among the records.

### Network construction

According to the association analysis, disease-disease pairs, drug-disease pairs and drug-drug pairs which can meet the requirements for both *Chi-square* test ( $\chi^2 > 6.7$  and  $p\text{-value} < 0.01$ ) and CR ( $CR > 2.0e-4$ ) were taken as the edges of a disease-drug network (DDN), and the corresponding non-redundant nodes for all the edges were taken as the nodes. In addition, a disease comorbidity network (DICN) was constructed by retaining only the disease-disease associations.

### Recognition of Comorbidity Medication Patterns

#### Definition of comorbidity medication motif (CMM)

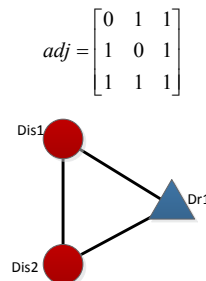
Considering the disease-drug network, which contain two different types of nodes and a great number of associations, there must be various motifs within the network. However, not all recurrent motifs can be applied to explain the medication strategies for disease comorbidities; for example, motifs with only drugs or only one isolated disease.

To make sure there are two diseases which may possess disease comorbidity relations, and each disease is managed by at least one effective drug, a CMM should meet the following requirements:

- a. There should be two connected disease nodes.

- b. Each disease node should be connected with at least one drug node.
- c. Each drug node should be connected with at least one disease.

Given the motif size as 3 and 4, we generated all candidate CMMs manually and saved them in the same way as a fast network motif detection tool: FANMOD [28]. Each CMM was described by its adjacent matrix, in which the diagonal element stands for the node type: 0 for disease and 1 for drug, other element represents whether two nodes were connected. For example, the adjacent matrix for a three node motif as shown in **Figure 1** is:



**Figure 1.** An example of three-node CMM.

### CMM evaluation

We utilized FANMOD which can discriminate different types of nodes in motif analysis to evaluate all candidate CMMs in the disease-drug network, and retain those which are statistically significant. The significance of each possible CMM was described by Z-score and its corresponding p-value [28]. The frequency for one CMM was calculated as the ratio of the sub graphs belonging to this CMM to all sub graphs with the same size.

### Comorbidity medication pattern

Then, for each CMM, we selected out the sub graphs with the same adjacent matrix of this CMM, and regarded them as the members of this CMM. For each CMM, its sub graph members were further partitioned into different comorbidity medication patterns (CMPs) based on the category information about diseases and drugs. The diseases were classified into 10 categories according to the second level classification in the ICD10 hierarchy, while the drugs were categorized based on the drug organization system of Dongfang hospital. Accordingly, a CMP contains the disease categories, drug categories and the corresponding network structures among them. Each CMP represents a basic comorbidity medication unit describing which types of drugs can be applied on the comorbidities between what kinds of diseases. To measure the prevalence for  $CMP_i$  which is a further-partition of  $CMM_k$ , a score was calculated:

$$score(CMP_i) = \frac{|\{subG \mid subG \in CMP_i\}|}{|\{subGM \mid subGM \in CMM_k, CMP_i \in CMM_k\}| / |\{CMP \mid CMP \in CMM_k\}|}$$

Where the numerator stands for the observed number of sub graph members belonging to  $CMP_i$ , the denominator represents the average number of members belonging to each further-partition of  $CMM_k$ .

## RESULTS AND DISCUSSION

### Framework

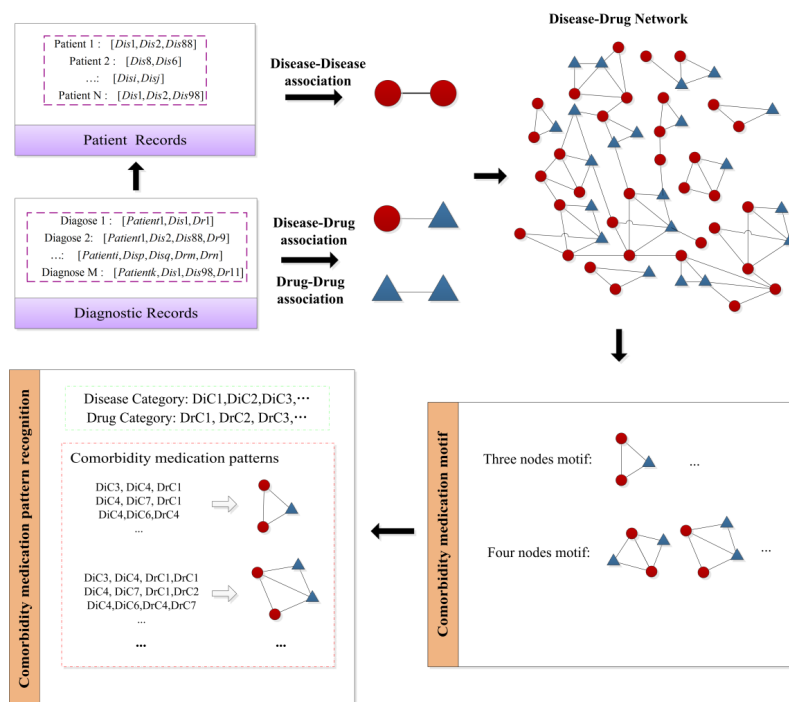
To identify the specific medication patterns for disease comorbidities, our work mainly involves four phases (**Figure 2**):

- a. Collect the circulation system disease-relevant EHRs from the Dongfang Hospital, and process the EHRs to extract standardized patient records and diagnostic records.
- b. Estimate the associations including disease-disease associations, drug-drug associations, and disease-drug associations and construct a disease-drug network based on the significant and meaningful associations.
- c. Generate comorbidity medication motifs which can represent the general comorbidity-drug association structures in the disease-drug network.
- d. Recognize the comorbidity medication patterns which further partition the sub graphs belonging to each comorbidity medication motif into different groups based on the explicit category information of diseases and drugs.

To be noted, the motif and pattern were applied to achieve different purposes: the motifs were utilized to represent the recurrent sub graph structures in the disease-drug network without considering the explicit disease or drug category; while the patterns in this study were more concrete than the motifs, they were applied to classify sub graphs belonging to one motif into more explicit groups considering the specific categories of diseases and drugs, thus to mine which type of comorbidities should be managed by which kinds of drugs by the specific associated formation described by each motif. In addition, since “comorbidity”

here is used to describe the correlations between different diseases, it involves at least two diseases. Consequently, “comorbidity” in this paper will refer to at least two significant co-occur diseases rather than only one of them.

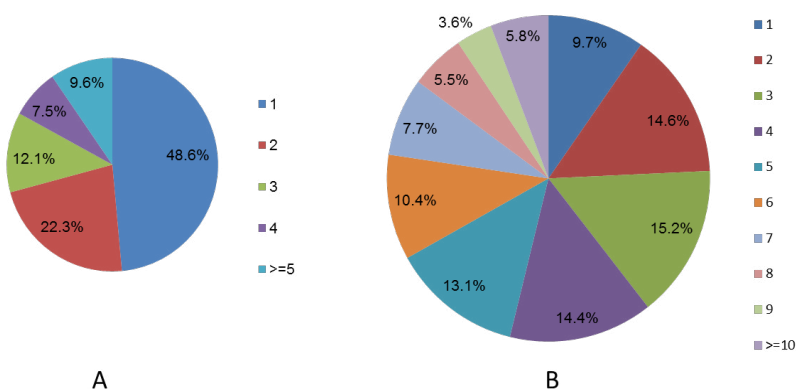
Based on the above work flow, we recognized the comorbidity medication patterns for the circulatory system diseases.



**Figure 2.** Work flows for the network motif-based recognition of disease comorbidity medication patterns. Red circles represent diseases and blue triangles represent drugs.

### A brief look at disease co-occurrences and drug co-administrations

We collected 448384 EHRs for 88153 patients who have been diagnosed with circulatory system diseases during the period 2010-2015. To have a brief look on the possibilities of disease comorbidity and drug combination, we first made a simple statistic on the number of diseases each individual patient has suffered from and the number of drugs prescribed for each diagnosis. In case of diseases, about half (51.4%) of the patients may have been affected by 2 or more than 2 circulatory system diseases based on the EHRs (**Figure 3A**). In case of drugs, only 9.7% diagnoses prescribed only one drug, two or even more drugs were co-administrated in most of the EHRs (**Figure 3B**). Although the simple statistical analysis cannot precisely explain the disease comorbidity or drug combination problem, it can still show that disease co-occurrences and drug co-administrations are very common for circulatory system diseases. Next, to reveal the concrete disease comorbidities and identify the corresponding medication patterns, a network motif-based study was carried out.



**Figure 3.** Statistics on the disease co-occurrences and drug co-administrations. A. Distribution of patients with respect to the number of circulatory system diseases they suffered from. B. Distribution of diagnoses considering how many drugs are prescribed in one diagnosis.

### Network Constructions

#### Disease comorbidity network

Before the construction of a heterogeneous drug-disease network (DDN), we built a disease comorbidity network (DICN), which is a sub graph of the DDN with only disease nodes and their corresponding associations, to make an intuitive description of

the disease comorbidity relations. In the DICN (**Figure 4**), there are 436 edges connecting 75 nodes, where nodes are diseases specified by unique ICD10 codes, and edges connect two diseases which were estimated as frequent and non-occasional co-occurrences (*Chi-square* test  $p$ -value<0.01, and  $CR>2.0e-4$ , see Materials and methods). According to the ICD10 system, diseases belonging to the circulatory system diseases can be divided into 10 main categories (**Table 2**). In this study, diseases belonging to 8 categories were recognized to possess comorbidity relations with each other (**Figure 4**).

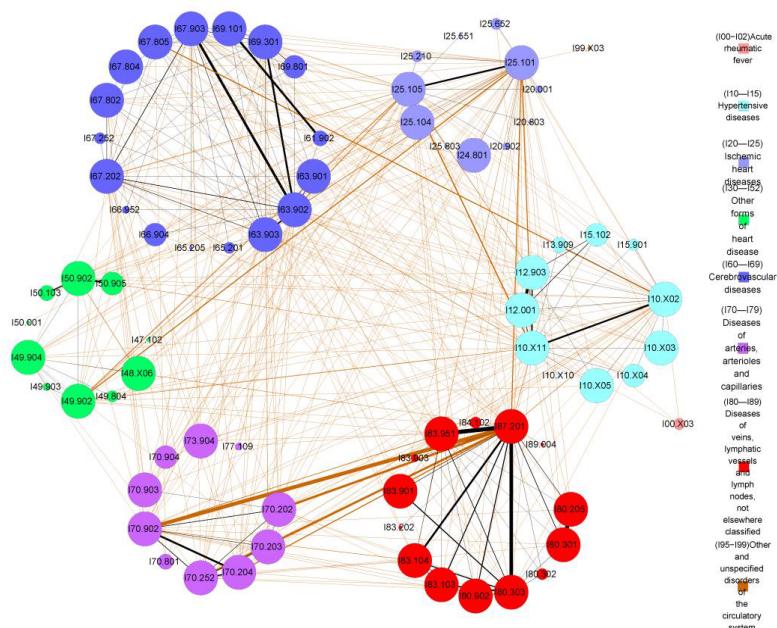
A part of the comorbidity relationships exist in two diseases belonging to same categories. Comparing to all other associations, we can observe that some of these in-category comorbidities like the association between I87.201 (venous insufficiency in lower extremity) and I80.303 (venosclerosis), or the association between I67.903 (cerebrovascular disease, unspecified) and I63.902 (cerebral infarction), are much stronger. This may lie in the anatomical or functional similarity for diseases belonging to same ICD10 category. These highly-associated in-category pairs seem to be trivial and uninteresting as they are just the direct reflections of the taxonomy of the ICD10 system.

However, there are even more comorbidity relations between different categories (**Figures 4 and 5**) for example, the comorbidities between ischemic heart diseases (Category 4) and cerebrovascular diseases (Category 7), as well as the comorbidities between hypertensive diseases (Category 3) and cerebrovascular diseases (Category 7). Among them, the most significant associations are between Category 7 and Category 8, especially the association between I70.902 (arteriosclerosis obliterans) and I87.201 (venous insufficiency in lower extremity). This implies that disorders on cerebrovascular may be comorbid with the problems on veins of lower extremity, and the reason for the comorbidity may lie in that both diseases share the common pathological mechanism that structural changes happen on the blood vessel wall [29,30]. These cross-category comorbidity relations suggest that most of the circulatory system disorders are not isolated: one disorder of the circulatory system may accompany with another lesion on the other location of the circulatory system [31-34].

In addition, according to the node degrees, some high-degree diseases, like I10.X02 (hypertension), I25.101 (coronary heart disease), I63.902 (cerebral infarction), and I87.201 (venous insufficiency in lower extremity) can form comorbidity relations with multiple diseases covering different categories. These diseases may be the primary problem in comorbidity conditions, and patients with such disease should pay more attention to the prevention and detection of their potential comorbidities in time.

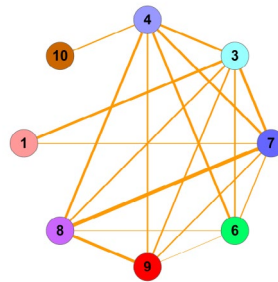
**Table 2.** Main categories in ICD10 Chapter IX-Diseases of circulatory system.

Category	ICD10 codes	Name
1	I00-I02	Acute rheumatic fever
2	I05-I09	Chronic rheumatic heart diseases
3	I10-I15	Hypertensive diseases
4	I20-I25	Ischemic heart diseases
5	I26-I28	Pulmonary heart disease and diseases of pulmonary circulation
6	I30-I52	Other forms of heart disease
7	I60-I69	Cerebrovascular diseases
8	I70-I79	Diseases of arteries, arterioles and capillaries
9	I80-I89	Diseases of veins, lymphatic vessels and lymph nodes, not elsewhere classified
10	I95-I99	Other and unspecified disorders of the circulatory system



**Figure 4.** A disease comorbidity networks.

Diseases represented by ICD10 codes were shown as nodes. Two diseases are connected if they are estimated to have comorbidity relations. Node colors represent the corresponding ICD10 classifications as shown on the right. Nodes with the same colors are connected by black edges, while nodes with different colored are linked by brown edges. Node size is proportional to the node degree in the network. Edge width is proportional to the *Chi-square* value ( $\chi^2$ ) calculated for the disease-disease associations.

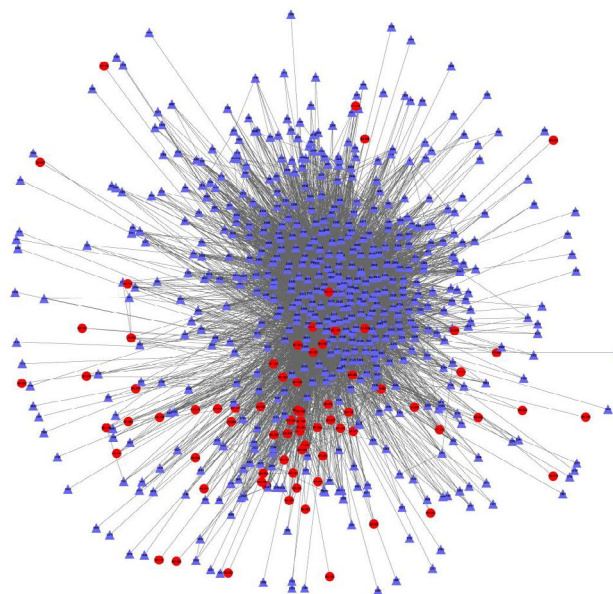


**Figure 5.** A category association networks for circulatory system diseases.

Labels on the nodes represent the categories as identified in **Table 2**. Different categories are distinguished by different colors as **Figure 4**. Edge width is proportional to the category-association score:  $score(C_i, C_j) = 2N_{ij} / N_i / N_j$ , where  $C_i, C_j$  are two different categories,  $N_i$  and  $N_j$  are the number of members belonging to  $C_i$  and  $C_j$  respectively,  $N_{ij}$  is the number of comorbidity associations between the members from  $C_i$  and the members from  $C_j$  in the DICN.

### Disease-drug network

To reflect the complex associations among diseases and drugs comprehensively, a disease-drug network (DDN) was constructed based on three types of associations including disease-disease associations, disease-drug associations and drug-drug associations. In the DDN (**Figure 6**), there were 436 edges among 75 diseases, 7326 edges among 536 drugs, and 2186 edges between 40 diseases and 563 drugs (**Tables 3-5**). There are so many drugs being applied in treating circulatory system diseases: some drugs are highly specific, and can only treat one type of diseases, while the others can be utilized on various diseases. What's more, some drugs can be combined to treat the same diseases. The miscellaneous associations in the DDN make it difficult to obtain knowledge about the general treatment strategies for disease comorbidities, yet, they provide a valuable basis for the further study of disease comorbidity medication patterns.



**Figure 6.** A disease-drug networks for circulatory system diseases.

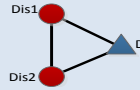
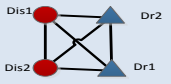
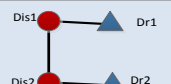


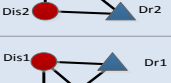

Blue triangle nodes represent drugs, and red circle nodes represent diseases. There are complex associations among different diseases and drugs. A further analysis is essential to obtain meaningful knowledge from this network.

### Motif Analysis

Networks offer a very intuitive representation of different kinds of interactions underlying complex systems. However, the great number of nodes in the network makes it difficult to understand some functional features hidden in the complex network directly. Since the DDN (**Figure 6**) consists of a relatively large number of nodes, it is hard to understand the comorbidity treatment strategies. Network motif (NM), which represents the significantly recurrent sub graph structures, can provide an effective way

to review on the local properties in a big network [35]. Consequently, network motifs were applied to discover meaningful patterns underlying in the DDN. However, some NMs in the DDN, like those with only one disease or no disease cannot explain the medication patterns for disease comorbidities. Further, a comorbidity medication motif (CMM) was defined (see Materials and methods) to reflect the general medication strategies for comorbidities. The CMMs with 3 or 4 nodes were generated and their frequencies were evaluated based on the DDN (Table 3). The three-node CMM suggests that some comorbidity can be cured by the same drug. The four-node CMM listed out 6 candidate forms of 2 diseases-2 drugs medication strategies.

Table 3. Candidate CMM scores.

Size	ID	Graph	Frequency	Z-score	p value	Frequency
3	3_1*		0.045	43.6	0	0.045
4	4_1*		0.025	73.6	0	0.025
	4_2*		0.17	17.7	0	0.17
	4_3*		0.05	10.89	0	0.05
	4_4		0.06	0.09	0.468	0.06
	4_5		0.001	-1.28	0.92	0.001
	4_6		0.010	-6.4	1	0.010

\*: Significant CMMs

The CMM was designed to mine the possible medication forms for disease comorbidity. In the CMMs, different kinds of connections describe the basic relations among diseases and drugs. Connections between two disease nodes represent the comorbidity relations. Connections between one drug and one disease suggest that this disease can be cured by this drug. Two connected drugs mean that these drugs can be prescribed simultaneously. On the contrary, two drugs without connections should not be co-administered, but can be applied in accordance with the sequence of disease progress. These motifs were evaluated by Z-scores and the corresponding p-values which can reflect whether certain motif is with more recurrent subnetworks than random conditions [28], only those with p-values less than 0.01 were considered as significant ones. According to the results, high frequencies are not necessary for certain motifs; most motifs (like CMM 4\_1 and CMM 4\_3) are only with lower frequencies. According to the Z-score of each CMM, among 6 candidate CMMs, only three are significant, and the most significant (z-score=73.6) 4-node CMM is CMM 4\_1. This CMM (CMM 4\_1) suggests certain disease comorbidity conditions can be handled by two drugs in combination and each individual drug has therapeutic effects for both diseases. This strategy accords with the TCM theory of “treating different diseases by the same drugs” and “compatibility of medicines”. To be more concrete, when two comorbid diseases “Dis1” and “Dis2” co-occur on certain patient, doctors can first use a “Dis1”-specific drug “Dr1” on this patient, and then substitute “Dr1” by another drug “Dr2” which can treat both diseases, or in an adverse order considering the disease progress. Next to the most significant CMM, the second one, covering 17% of all four node sub graphs, is with larger frequency than other candidate CMMs, the difference between it and the most significant CMM lies in that these two drugs are not co-administrated, and each drug is with specific effect for only one disease. This strategy may pay more attention to the specificity of each disease. In addition, comparing CMM 4\_1 and CMM 4\_2 with CMM 4\_3, we can discover that there are more medications strategies including one drug which can treat both diseases than applying only disease-specific drugs. These significant motifs, although not necessarily, are more likely to reflect the practical medication strategy.

### Comorbidity Medication Patterns

CMMs have revealed the general comorbidity medication formations. However, different types of disease comorbidities may need different kinds of drugs during the medication management. To further understand the medication rules, we made an additional effort to mine the comorbidity medication patterns (CMPs) which are more explicit than the CMMs by including the category information for diseases and drugs (see Materials and methods). Since the EHRs were from a TCM hospital, most of the recorded drugs belong to TCM categories, like the blood-regulating agents, tonics and meridians-relieving agents (Table 4). The recognized CMPs can help us to comprehend the therapeutic principles for disease comorbidities in the main context of TCM theory. The top-10 scored CMPs for three most frequent motifs were shown in Table 5.



For CMM 3\_1, the top-ranked CMPs (CMP 3\_11-CMP 3\_15) show that blood-regulating agents, which are with the main functions of adjusting human body's blood generation and operation, can be used to treat comorbidity conditions among hypertension diseases, ischemic heart diseases and cerebrovascular diseases. This is consistent with the TCM theory of "the same treatment for different disease" [36] which means that the same method can be used to treat different diseases in their development process if there are evidences for the existence of common pathological nature. According to the TCM theory, hypertensive diseases, ischemic heart diseases and cerebrovascular diseases belong to "vertigo", "obstruction", "stroke" respectively. Though the disease locations and clinical manifestations are different, "blood stasis" is the main common syndrome [37]. Therefore, blood-regulating agents can achieve the therapeutic effects for these different diseases through promoting blood circulation and remove blood stasis which is the common pathological factors for these diseases.

For CMM 4\_1, some CMPs like CMP 4\_11, CMP 4\_12, CMP 4\_14 and CMP 4\_17 show that when two drugs are taken into consideration, two different blood-regulating agents can be applied successively on patients with comorbidities among hypertension diseases, ischemic heart diseases or cerebrovascular diseases. Compared to the three-node CMPs, the inclusion of another blood-regulating agent can help to promote blood circulation and remove blood stasis in a different way, and thus to adapt to the comorbidity dynamic developing process. What's more, there can also be another tonic to be in succession with one certain blood-regulating agent for the comorbidity conditions (CMP 4\_13, CMP 4\_15 and CMP 4\_110). According to the basic theory of TCM, Qi is the commander of the blood and Qi could quicken the blood circulation [38,39]. Therefore, Qi-supplementing agents, like the tonics can be used to enhance the regulation effects of the blood-regulating agents on the treatment of the comorbidities among these "blood stasis" diseases.

**Table 4.** Main drug category information.

Drug Category	English Name	Main Functions
24	Meridians-relieving agent	Dredge channels and collaterals, circulate Qi and blood
104	Heat-clearing agent	Clear heat, purge fire, relieve toxicity
106	Dispelling-wind agent	Dissipate external wind, extinguish internal wind, check convulsion
107	Dampness-eliminating agent	Eliminate dampness, promote diuresis, free strangury and discharge turbidity
114	Blood-regulating agent	Stop bleeding, activate blood, resolve stasis, replenish blood
115	Tonic	Replenish Qi, strengthen the body
117	Sedative and tranquilizing	Relieve uneasiness of mind and body
128	Orthopedics and traumatology agent	Promote blood circulation and remove blood-stasis, relieve pain, remove masses, relieves swelling with pain
404	Calcium antagonist	Inhibit inflow of calcium and reduce of calcium overload.
406	Hypo tensor	Control blood pressure
1104	Blood tonics	Replenish blood, treat blood deficiency syndrome
1402	Vitamin B drug	Supplement vitamin B
130302	Oral hypoglycemic agent	Control blood glucose

For CMM 4\_2, the top-scored CMP (CMP 4\_21) suggests that one blood-regulating agent can be combined with a blood tonic to treat comorbidities among hypertensive and ischemic heart diseases (CMM4\_21). Blood stasis will hinder the generation of new blood. Therefore, blood tonics can be combined with blood-regulating agents to promote blood production in the comorbidity of ischemic heart diseases and hypertensive diseases. Considering CMP 4\_22, though cerebrovascular diseases and diseases of arteries, arterioles and capillaries emerge on different locations, these diseases can co-occur [40-43], and the medications for such comorbidities can be composed of blood-regulating agents and meridians-relieving agents.

**Table 5.** Top-10 comorbidity medication patterns for three CMMs.

CMM ID	CMM Information		Comorbidity Medication pattern(CMP)					
	CMM	Rank	CMP ID	Dis1*	Dis2*	Dr1*	Dr2*	Score
3_1		1	3_11	7	7	114	-----	32.24
		2	3_12	7	4	114	-----	25.29
		3	3_13	3	4	114	-----	20.66
		4	3_14	7	3	114	-----	20.13
		5	3_15	4	7	114	-----	13.71
		6	3_16	3	4	406	-----	10.51
		7	3_17	7	3	406	-----	9.26
		8	3_18	3	7	114	-----	8.73
		9	3_19	7	7	404	-----	7.84
		10	3_110	7	3	404	-----	7.12

4_1		1	4_11	3	7	114	114	37.26
		2	4_12	4	7	114	114	34.05
		3	4_13	4	4	115	114	33.09
		4	4_14	4	4	114	114	28.59
		5	4_15	4	7	115	114	26.66
		6	4_16	3	4	106	114	25.05
		7	4_17	3	4	114	114	23.77
		8	4_18	4	7	106	114	22.81
		9	4_19	3	7	114	115	21.52
		10	4_110	3	7	115	114	21.52
4_2		1	4_21	4	3	114	1104	14.44
		2	4_22	8	7	24	114	12.51
		3	4_23	4	4	115	104	11.07
		4	4_24	3	7	107	106	10.11
		5	4_25	4	4	115	107	9.63
		6	4_26	3	7	107	117	9.15
		7	4_27	3	7	128	106	8.66
		8	4_28	9	4	1402	114	8.66
		9	4_29	3	3	130302	104	8.66
		10	4_210	8	3	106	114	8.66

\*: Dis1 and Dis2 represent the disease categories for each CMP; while Dr1 and Dr2 represent the drug categories for each CMP. See Table 2 and Table 4 for the concrete information for disease and drug categories.

Furthermore, the CMPs have also indicated that the same kind comorbidity can be managed by different medication strategies. Taking the CMP 4\_11, CMP 4\_19, CMP 4\_110, CMP 4\_26, and CMP 4\_27 as examples, we can find that they all refer to the same pair of disease categories: hypertensive diseases (Category 3) and cerebrovascular diseases (Category 7), but the medication modes are different. This manifests the diversity of medication strategies on comorbidities. However, different patterns should not be regarded as equals: the high-scored CMPs may be more commonly adopted than low-scored ones, i.e., comorbidities between hypertension and cerebrovascular diseases are more likely to be intervened by the medication pattern of CMP 4\_11.

## CONCLUSION

Disease comorbidity is one common and important health issue during medical practice. Special attention need be paid on the medication management of patients with certain comorbidity. The EHRs provide a previous data resource for medical researches, some studies have discovered potential comorbidity associations based on the EHRs. However, less attention has been paid on the corresponding medication modes for diseases comorbidities. Here we take a step forward to recognize the comorbidity medication patterns after we obtained the comorbidity associations. Understanding the comorbidity medication patterns will facilitate treatment on patients with comorbidities, and improve the proper use of appropriate drugs.

First, we estimated on three types of associations: disease-disease associations, drug-disease associations, and drug-drug associations, and represented the significant associations by the DDN. Next, we recognized the comorbidity medication patterns from the DDN by a motif-based analysis. Taking the diseases of circulatory system as examples, we utilized our method and illustrated both the general medication forms and the explicit medication pattern in the context of specific disease and drug categories. Generally speaking, drugs which are with therapeutic effects for multiple diseases are widely adopted for comorbidities. Moreover, the high-score comorbidity medication patterns help us comprehend the medical principles for comorbidities, and provide some references on the medication rules for comorbidities in the context of circulatory system. For example, different blood-regulating agents can be applied successively on comorbidities between different “blood stasis”-centered diseases. Since these patterns are based on motifs with more recurrent subgraphs than random network, they are likely to be generated by some latent law in treating comorbidities. However, they can only describe the medication principle to some degree. It still needs further evaluation on the practical function of each motif, and we will make great efforts on this issue in future study. Even so, these findings may assist the medical staffs to make better medical decisions and managements for patients with comorbidities.

In our study, we only utilized motifs with 3 and 4 nodes to describe the comorbidity medication pattern. The 3-node motif helps us to understand the co-treatment for different diseases by the same drug, while the 4-node motifs consider the pair-wise drugs as a basic unit to see the combination rules. Although in clinical practice there are often more drugs in one prescription for patients with comorbidities, these drugs can be divided into a group of pair-wise drugs and single drugs. Consequently, the 3-node and 4-node motifs are capable to describe the most basic comorbidity medication patterns. To make a specific and explicit evaluation on the comorbidity situations and comorbidity treatment modes for one major type of disease, only circulatory system diseases were taken into account in this study. However, our model is not limited to these diseases; it can also be applied to make estimations on the comorbidity associations among different types of diseases, like endocrine, nutritional and metabolic diseases (ICD10 Chapter IV), mental and behavioral disorders (ICD10 Chapter V), and diseases of the nervous system (ICD10 Chapter VI), if there are enough EHRs for these diseases.

One limitation of this work is that we cannot discriminate the primary disease and the comorbid disease for two associated disease. At present, it can be hypothesized that disease with larger degree on the DRCN may be a primary disease. However, it is not persuasive enough. Therefore, a future study will be conducted to make clear the causal relations and recognize the corresponding medication patterns by taking more information in the EHRs into consideration. Besides, given that all the EHRs were from the Dongfang hospital which is a hospital of TCM, the recognized comorbidity medication patterns are more suitable to reflect the medical experience of TCM practitioners and can be served as references, but they are inadequate to represent the golden standards for comorbidity interventions.

One limitation of this work is that we cannot discriminate the primary disease and the comorbid disease for two associated disease. At present, it can be hypothesized that disease with larger degree on the DRCN may be a primary disease. However, it is not persuasive enough. Therefore, a future study will be conducted to make clear the causal relations and recognize the corresponding medication patterns by taking more information in the EHRs into consideration. Besides, given that all the EHRs were from the Dongfang hospital which is a hospital of TCM, the recognized comorbidity medication patterns are more suitable to reflect the medical experience of TCM practitioners and can be served as references, but they are inadequate to represent the golden standards for comorbidity interventions.

## **ACKNOWLEDGMENTS**

This work was supported by the National Natural Science Foundation of China (Grant No. 81303152), and the State Key Program of National Natural Science Foundation of China (Grant No. 81330086)

## **CONFLICT OF INTERESTS**

The authors declare that there is no conflict of interests regarding the publication of this paper.

## **REFERENCES**

1. Valderas JM, et al. Defining Comorbidity: Implications for Understanding Health and Health Services, *Annals of Family Medicine*. 2009;7(4):357-363.
2. Froimson LS. Comorbidities of Diabetes and Hypertension in Adult Patients with Bipolar Illness and Schizophrenia: Medical Considerations. Using Motivational Interviewing to Help These Patients, *J Am Psychiatr Nurses Assoc*. 2015;21(1):39-39.
3. Long AN and Dagogo-Jack S. Comorbidities of Diabetes and Hypertension: Mechanisms and Approach to Target Organ Protection, *J Clin Hypertens*. 2011;13(4):244-251.
4. Smith SM and O'Dowd T. Chronic diseases: what happens when they come in multiples? *Br J Gen Pract*. 2007;579:268-270.
5. Mercer SW, et al. Multimorbidity in primary care: developing the research agenda, *Family Practice*. 2009;26:79-80.
6. Canaway R and Merkes M. Barriers to comorbidity service delivery: the complexities of dual diagnosis and the need to agree on terminology and conceptual frameworks, *Australian Health Review*. 2010;34:262-268.
7. Vassallo M, et al. Comorbidity and differential diagnosis, *Medecine Maladies Infectieuses* 2013;43:23-28.
8. Boje CR. Impact of comorbidity on treatment outcome in head and neck squamous cell carcinoma - A systematic review, *Radiotherapy and Oncology*. 2014;110:81-90.
9. Pinhas-Hamiel O and Zeitler P. Who is the wise man? The one who foresees consequences: Childhood obesity, new associated comorbidity and prevention, *Preventive Medicine*. 2000;31:702-705.
10. Swendsen J, et al. Improving targets for the prevention of drug use disorders: Sociodemographic predictors of transitions across drug use stages in the national comorbidity survey replication, *Preventive Medicine*. 2008;47:629-634.
11. Kadam UT, et al. Clinical comorbidity was specific to disease pathology, psychologic distress, and somatic symptom amplification, *J Clin Epidemiol*. 2005;58:909-917.
12. Hayrinen K, et al. Definition, structure, content, use and impacts of electronic health records: A review of the research literature, *Int J Med Inform*. 2008;77:5:291-304.
13. Meyers LA. A Dynamic Network Approach for the Study of Human Phenotypes, *PLoS Computational Biology*. 2009;5:e1000353.
14. Roque FS. Using electronic patient records to discover disease correlations and stratify patient cohorts, *PLoS Comput Biol*. 2011;7:e1002141.
15. Menche J, et al. Disease networks. Uncovering disease-disease relationships through the incomplete interactome, *Science*. 2015;347:1257601,
16. Goh KI, et al. The human disease network, *Proc Natl Acad Sci USA*. 2007;104:21:8685-8690

17. Li Y and Agarwal P. A Pathway-Based View of Human Diseases and Disease Relationships, *Plos One*. 2009;4:2.
18. Zhou XZ, et al. Human symptoms-disease network, *Nature Communications*. 2014;5.
19. Smith SM, et al. Managing patients with multimorbidity: systematic review of interventions in primary care and community settings, *BMJ*. 2012;345.
20. Bower P, et al. Multimorbidity, service organization and clinical decision making in primary care: a qualitative study, *Family Practice*. 2011;28:579-587.
21. Laux G, et al. Co- and multimorbidity patterns in primary care based on episodes of care: results from the German CONTENT project, *Bmc Health Services Research*. 2008;8.
22. Gee RM, et al. Importance of health status and clinical comorbidity in performing exercise among patients with hypertension and coronary artery disease, *Circulation*. 2001;104:642-642.
23. Mancia G, et al. Hypertension and migraine comorbidity: prevalence and risk of cerebrovascular events: evidence from a large, multicenter, cross-sectional survey in Italy (MIRACLES study), *J Hypertens*. 2011;29:309-318.
24. Wirtwein M, et al. Chronopharmacotherapy of hypertension is crucial in mortality in patients with coronary heart disease, *Eur Heart J*. 2014;35:880-880.
25. Polupanov AG, et al. Lipoprotein-Associated Phospholipase A2: Relation to Development of Ischemic Stroke in Patients with Essential Hypertension, *Kardiologiya*. 2014;54:29-34,
26. McHugh ML. The Chi-square test of independence, *Biochemia Medica*. 2013;23:143-149.
27. Yates F. Contingency table involving small numbers and the  $\chi^2$  test, *Supplement to the Journal of the Royal Statistical Society*. 1934;1:217-235.
28. Wernicke S and Rasche F. FANMOD: a tool for fast network motif detection, *Bioinformatics*. 2006;22:1152-1153.
29. Norden c, et al. The Interaction Blood to Vessel Wall in Patients with Arteriosclerosis Obliterans during Vitamin-C Treatment, *International Journal of Microcirculation-Clinical and Experimental*. 1984;3:425.
30. Stenberg B, et al. Effect of venous stasis on vessel wall fibrinolysis, *Thromb Haemost*. 1984;51:240-242.
31. Agle SC, et al. The Association of Periprocedural Hypertension and Adverse Outcomes in Patients Undergoing Catheter-Directed Thrombolysis, *Annals of Vascular Surgery*. 2010;24:609-614.
32. Iijima k, et al. Coronary Artery Calcification and Cerebral Small Vessel Disease Association of Systemic Atherosclerosis, *Circulation*. 2011;75:272-273.
33. Roach REJ. The relationship between venous thrombosis and subsequent arterial cardiovascular disease: causal or mediated by common etiologic factors? *J Thromb Haemost*. 2011;9:442-442.
34. Di Minno MND, et al. Identifying high-risk individuals for cardiovascular disease: similarities between venous and arterial thrombosis in perspective. A 2011 update, *Intern Emerg Med*. 2012;7:1:9-13.
35. Milo R, et al. Network motifs: Simple building blocks of complex networks, *Science*. 2002;298:5594:824-827.
36. Chen D, et al. Molecular mechanism research on simultaneous therapy of brain and heart based on data mining and network analysis. 2013;38:1:91-98.
37. Choi TY, et al. Expert opinions on the concept of blood stasis in China: An interview study, *Chin J Integr Med*. 2014.
38. Xue JM. Discussion on the speciality of TCM term translation and the translating principles viewing from the concept of TCM term "qi". 2008;28:5:471-472.
39. Liu W, et al. Chinese patent medicine for chronic obstructive pulmonary disease based on principles of tonifying qi, promoting blood circulation by removing blood stasis, and resolving phlegm: a systematic review of randomized controlled trials, *J Tradit Chin Med*. 2015;35:1:1-10.
40. Pandya A, et al. Carotid Artery Stenosis: Cost-effectiveness of Assessment of Cerebrovascular Reserve to Guide Treatment of Asymptomatic Patients, *Radiology*. 2015;274:455-463.
41. Reinhard M. Cerebrovascular reactivity predicts stroke in high-grade carotid artery disease: An individual patient data meta-analysis, *Cerebrovascular Diseases*. 2014;37:20.
42. Ninomia T, et al. Brain injury and cerebrovascular fibrin deposition correlate with reduced antithrombotic brain capillary functions in a hypertensive stroke model, *J Cereb Blood Flow Metab*. 20:6:998-1009.
43. Younas F, et al. Multiple intracranial capillary hemangiomas and transient cerebrovascular insufficiency, *Neurological Sciences*. 2011;32:5:963-966.