Research Paper

Comorbidity Analysis According to Sex and Age in Hypertension Patients in China

Jiaqi Liu[1†], James Ma[3†], Jiaojiao Wang[1†], Daniel Dajun Zeng[1], Hongbin Song[4], Ligui Wang[4], Zhidong Cao[1,2✉]

1. The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China;
2. Cloud Computing Center, Chinese Academy of Sciences, Dongguan, China.
3. College of Business, University of Colorado, Colorado Springs, CO, USA.
4. Institute of Disease Control and Prevention, Academy of Military Medical Sciences, Beijing, China.

†These authors contributed equally to this work.

✉ Corresponding author: Zhidong Cao, Institute of Automation, Chinese Academy of Sciences, No 95 Zhongguancun East Road, 100190, Beijing, China. E-mail: zhidong.cao@ia.ac.cn

Abstract

Background: Hypertension, an important risk factor for the health of human being, is often accompanied by various comorbidities. However, the incidence patterns of those comorbidities have not been widely studied.

Aim: Applying big-data techniques on a large collection of electronic medical records, we investigated sex-specific and age-specific detection rates of some important comorbidities of hypertension, and sketched their relationships to reveal the risk for hypertension patients.

Methods: We collected a total of 6,371,963 hypertension-related medical records from 106 hospitals in 72 cities throughout China. Those records were reported to a National Center for Disease Control in China between 2011 and 2013. Based on the comprehensive and geographically distributed data set, we identified the top 20 comorbidities of hypertension, and disclosed the sex-specific and age-specific patterns of those comorbidities. A comorbidities network was constructed based on the frequency of co-occurrence relationships among those comorbidities.

Results: The top four comorbidities of hypertension were coronary heart disease, diabetes, hyperlipemia, and arteriosclerosis, whose detection rates were 21.71% (21.49% for men vs 21.95% for women), 16.00% (16.24% vs 15.74%), 13.81% (13.86% vs 13.76%), and 12.66% (12.25% vs 13.08%), respectively. The age-specific detection rates of comorbidities showed five unique patterns and also indicated that nephropathy, uremia, and anemia were significant risks for patients under 39 years of age. On the other hand, coronary heart disease, diabetes, arteriosclerosis, hyperlipemia, and cerebral infarction were more likely to occur in older patients. The comorbidity network that we constructed indicated that the top 20 comorbidities of hypertension had strong co-occurrence correlations.

Conclusions: Hypertension patients can be aware of their risks of comorbidities based on our sex-specific results, age-specific patterns, and the comorbidity network. Our findings provide useful insights into the comorbidity prevention, risk assessment, and early warning for hypertension patients.

Key words: Hypertension, Comorbidity, Electronic Medical Records, Detection Rate, Network Analysis.

Background

Hypertension, or high blood pressure, is one of the most important risk factors that can lead to cardiovascular diseases, and is thus regarded as a serious public health problem. The prevalence of hypertension has been increasing in most areas worldwide [1, 2]. In China, hypertension is the leading preventable risk factor for death among Chinese adults aged 40 years and older [3, 4]. Moreover, hypertension has a large number of comorbidities, which greatly affect hypertension patients' quality of life [5-7]. In previous

years, researchers and medical practitioners have made a tremendous effort to study the comorbidities of hypertension [8-10]. Specifically, heart disease [2, 11], diabetes [12, 13], and obesity [14, 15] are the most widely studied comorbidities of hypertension. Some other diseases, such as allergic respiratory disease [9], sleep-disordered breathing [16], and chronic kidney disease [17], have also been studied as potential comorbidities of hypertension. Hypertension and some of its comorbidities have shown high correlations in terms of their prevalence. An example of this type of correlations is that the prevalence of hypertension in patients with diabetes is as high as 92.7% [18].

Moreover, the sex-specific and age-specific analyses of comorbidities of hypertension have resulted in various important findings [1, 19-21]. Specifically, the incident rates of comorbidities in hypertension patients with a different sex and age can significantly differ. An example of this difference is that the incidence of hypertension and hypercholesterolemia combined is 20% for women versus 16% for men, and ranges from 1.9% for those aged 20–29 to 56% for those aged 80 years and older [22]. Additionally, patient's age and sex need to be considered for treatment of these comorbidities [23, 24]. An example of the situation is that treatment for hypertension patients who are 80 years or older with indapamide has been proved to be effective and can also reduce the patient's risk of stroke [23]. Research has shown that untreated male hypertension patients are more likely to suffer from cognitive impairment than untreated female hypertension patients do [25]. Thus, hypertension should be treated and controlled as early as possible for male patients before they encounter dementia.

There has been increasing interest in analyzing disease relationships using network theory [26, 27]. The disease network is particularly useful when analyzing the co-occurrence of different diseases. Specifically, the disease network denotes an individual disease with a vertex, and the co-occurrence of two diseases with an edge connecting those two diseases. The disease network summarizes the connections among diseases and shows progress of disease preferentially along the edges or links [28]. The frequency of co-occurrence relationships among important comorbidities could provide useful insight into describing the disease development process, and thus result in doctor's and patient's awareness of diseases at the early stage of development. Studying the comorbidity co-occurrence of hypertension using the disease network may be an effective tool for determining meaningful comorbidity relationships that other approaches have not reported.

Although the comorbidities of hypertension have been extensively studied, most existing research is based on medical surveys and public census data. Census data sets show aggregated facts of the general public without detailed information regarding individual patients. In contrast, medical surveys while include some individual level information usually involve a limited number of survey participants because of limited resources. Those surveys are often set for a confined geographical area (i.e., a city or a county), and thus cannot claim to be representative of a larger and broader area. Due to the nature of medical surveys, usually only certain types of participants are willing to reveal their private and sensitive medical related situations. People with a stronger sense of privacy are normally reluctant to reveal their medical history or health-related conditions. Therefore, medical surveys on a voluntary basis may have a biased participant population. The data points that are collected in a medical survey also largely depend on the participant's availability during the time of the survey, the participant's mood at the time, and the survey collector's attitude and human interaction skills. Too many human-related factors can affect the quality of medical surveys. Furthermore, to reduce the survey participant's reluctance, a medical survey is usually composed of a limited number of survey questions so that an interview or a questionnaire can be completed within a short time period. This greatly reduces the versatility of the survey when analyzing the survey results. In summary, because of the time-consuming and labor-intensive nature of medical surveys, the limited number of, and possibly biased, survey participants and survey questions can lead to biased analysis results, and possibly overlook important patterns and relationships in the occurrence of diseases.

In the current study, we leveraged a large, reliable, and extensive data set and analyzed the occurrence patterns of hypertension comorbidities. We also investigated the common comorbidities of hypertension with respect to the patient's sex and age. The co-occurrence relationships among comorbidities of hypertension are also discussed using the disease network approach.

Methods

Study population

Our data set was obtained from a Chinese National Surveillance System, which was initially implemented by the Chinese government in 2010. This surveillance system collects electronic medical records from hospitals and aims to oversee the overall health conditions of the Chinese population. Since 2010, this

system has been adopted by 192 hospitals located throughout China. Although we had access to all 192 hospitals' data in the surveillance system, we intentionally excluded some hospitals that did not appear to present a sufficient and continuous data stream. Some obvious errors and incomplete data points were also removed to maintain the data integrity.

Eventually, we decided to use 6,371,963 hypertension-related high-quality data records from the 110,528,991 electronic medical records that we had access to. Those medical records were dated between 2011 and 2013, and were from 106 hospitals located in 72 cities in China (Figure S1). Those cities are geographically distributed in 29 of 31 provinces in China (excluding two underpopulated provinces, Qinghai and Ningxia). Our data set covers 33.90% of the city population in China. The city population data is based on the sixth Chinese population census published by the National Bureau of Statistics of the People's Republic of China (http://www.stats.gov.cn).

This study was approved by the institutional review board of the Institute of Automation, Chinese Academy of Sciences. The data set was collected by the Chinese government for disease control. All patients gave their informed consent. The patient's privacy was strictly preserved in our study. We only used the patient's sex, age, and clinical diagnostic information to perform our analysis. Patients' identity-related information was masked before we started our study.

Data normalization

The clinical diagnosis in the original electronic medical records was not coded using uniformed and standardized text terms. An example was that some doctors had used "upper infection" as an abbreviation for "upper respiratory tract infection" and others had chosen a different abbreviation for the same diagnosis. To standardize the diagnosis, we applied a natural language processing technique [29, 30] and developed several in-house Python scripts for Chinese text processing and mining. Python [31] has been proved an effective tool for handling similar tasks. Specifically for our study, each electronic medical record was automatically segmented into a series of Chinese words, and these words were then combined to form Chinese phrases according to the probability distribution of those words. In addition to automatic normalization of data, many text ambiguities and synonyms were handled manually. Finally, all medical diagnostic records were converted to standardized and coded diagnostic terms that could be easily manipulated and analyzed.

Statistical analysis

The occurrences of comorbidities were counted in hypertension-related electronic medical records. The comorbidity's occurrence was then utilized to derive the detection rate of the comorbidity which better reflected the comorbidity's prevalence in hypertensive patients. The detection rate of a comorbidity was defined as the ratio of the number of the comorbidity's records to the number of hypertension-related records:

$$DR_{com} = \frac{N_{com}}{N_{HTN}} \times 100\%$$

The sex-specific detection rate was determined as the ratio of the number of each comorbidity in males or females to the number of hypertension cases in the corresponding sex group. The odds ratios and their 95% confidence interval (CI) of each sex-specific detection rate were also calculated. For the age-specific analysis, every 10-year age range between 0 and 99 years was considered an age group (e.g., 0–9 years, 10–19 years). Ages greater than or equal to 100 years were considered as one age group. Because the numbers of each comorbidity in the 0–9 years group and above 99 years group were small, the age-specific detection rates were calculated and analyzed only from 10 years to 99 years. Similarly the age-specific detection rate was determined as the ratio of the number of each comorbidity in each age group to the number of hypertension cases in the corresponding age group. Their 95% CIs were calculated. To analyze the age-specific prevalence trends of the top 20 comorbidities, the expectation maximization class in Weka [32] version 3.7.7 was used to cluster those 20 trends. The expectation maximization [33] algorithm assigns a probability distribution to each trend, which indicates the probability of it belonging to each cluster.

Network analysis

When two comorbidities of hypertension appeared in one electronic medical record, we considered that there was a co-occurrence relationship between this comorbidity pair. The number of co-occurrences between a couple comorbidities can be an important factor to reveal the relationship of those two comorbidities. Thus, we constructed a weighted comorbidity network [34, 35] to study the comorbidities of hypertension and the co-occurrence relationships among those comorbidities.

The nodes of the network represented comorbidities and the diameter of each node was proportional to the detection rate of each comorbidity. An edge in the network indicated the co-occurrence of two comorbidities whom that edge was connecting.

The weight of an edge was the number of co-occurrences of those two comorbidities. When an electronic medical record contained more than two comorbidities of hypertension, the count of every relationship between each possible pair of comorbidities in that record would have an increment of one (e.g., when the record was "hypertension, A, B, C", the count of relationships A-B, A-C, and B-C would all encounter an increment of one). After investigating all hypertension-related electronic medical records, we retained the high-frequency relationships among the top 20 comorbidities. The high-frequency relationships were defined as relationships with a weight of more than 1% of the total number of hypertension-related records.

Several network measures have been adopted to identify the importance of nodes [36]. Three primary methods, namely degree centrality, average degree, and average path length [37], were used to analyze the comorbidity network. Degree centrality is the most readily calculated and understood concept of node centrality. The degree centrality of a comorbidity is the total number of relationships that are directly associated with that comorbidity. A comorbidity with a high degree centrality has more co-occurrence relationships with other comorbidities in the network [38]. The average degree of a network is an overall evaluation about the connections among comorbidities [39]. In addition, path length focuses on the least number of relationships in order to connect two comorbidities. A comorbidity pair with a low path length and high edge weights along the path has a higher risk of co-occurrence in hypertension patients. The path length of any two directly connected comorbidities is one and the number of comorbidities on the shortest path is path length minus one. Similar to the average degree of a network, the average path length of a network is also used to describe the average distance between each comorbidity pair in the network [40]. A frequently used force-directed layout algorithm, the Fruchterman–Reingold algorithm, was used to layout the network.

Results

Detection rates of the top 20 comorbidities

The top 20 comorbidities of hypertension with the highest detection rates were identified (Table 1). Coronary heart disease (CHD), which is one of the most important cardiovascular diseases, had the highest detection rate. Diabetes, hyperlipemia, and arteriosclerosis had a detection rate that was higher than 10%. Cerebral diseases, such as cerebral infarction and cerebral circulation insufficiency, kidney-related diseases, such as nephropathy, renal in-

sufficiency and uremia, and respiratory-related diseases, such as respiratory tract infection, upper respiratory tract infection, and tracheitis had a high detection rate, which indicated that those comorbidities were of a higher risk in hypertension patients than other comorbidities were. Moreover, the detection rates of comorbidities reduced with rank. The detection rate of the last comorbidity, arthritis, was only 1.96%.

Table 1. Detection rates of the top 20 comorbidities of hypertension in China.

No.	Comorbidity	Detection Rate(%)	95% CI
1	Coronary Heart Disease	21.71	21.68-21.74
2	Diabetes	16.00	15.97-16.03
3	Hyperlipemia	13.81	13.78-13.84
4	Arteriosclerosis	12.66	12.63-12.68
5	Cerebral Infarction	7.53	7.51-7.55
6	Move With Difficulty	4.35	4.34-4.37
7	Nephropathy	4.24	4.23-4.26
8	Respiratory Tract Infection	3.95	3.94-3.97
9	Cerebral Circulation Insufficiency	3.87	3.85-3.88
10	Upper Respiratory Tract Infection	3.43	3.42-3.45
11	Renal Insufficiency	3.25	3.23-3.26
12	Tracheitis	3.10	3.09-3.12
13	Osteoporosis	3.04	3.03-3.05
14	Insomnia	2.86	2.85-2.87
15	Uremia	2.73	2.82-2.74
16	Anemia	2.42	2.41-2.44
17	Arrhythmia	2.39	2.38-2.40
18	Gastritis	2.26	2.25-2.27
19	Osteoarthropathy	2.00	1.99-2.01
20	Arthritis	1.96	1.95-1.97

Sex-specific detection rates

The sex-specific detection rates of the top 20 comorbidities of hypertension and their odds ratios were shown in Table 2 and Figure S2. Osteoporosis showed the largest difference between males and females, which suggested that female hypertension patients have a 73.12% higher risk than male hypertension patients in developing osteoporosis. Other bone-related diseases, such as arthritis and osteoarthropathy, also had a higher incidence in female hypertension patients than in male hypertension patients (40.64% vs 36.29%). In addition, insomnia and difficulty with movement threated the health of females more than males (39.78% vs 29.92%). Surprisingly, two cerebral diseases showed different risks in males and females. Cerebral circulation insufficiency was 40.15% more likely to occur in females, while cerebral infarction was 19.05% more likely to occur in males. Moreover, several diseases related to the kidney had a higher morbidity in male hypertension patients than in female hypertension patients. More attention should be paid to renal insufficiency, uremia,

and nephropathy in male hypertension patients (35.39%, 25.56%, and 17.99%) than in female hypertension patients. The sex-specific detection rates of other top comorbidities, including CHD, diabetes, hyperlipemia, and arteriosclerosis, were relatively uniform, with no significant differences between male and female patients.

Age-specific detection rates

The age-specific occurrence distribution of hypertension patients was shown in Figure 1. Based on 6,371,963 electronic medical records, the proportion of hypertension patients who were aged between 50 and 79 years was 71.27% (95% CI: 71.23–71.31%). Only 5.99% of hypertension patients were younger than 40 years. In addition, because there was only a small number of patients who were aged 9 years or older than 100 years, these two age groups were removed from the analysis.

The top five detection rates of comorbidities in each age group were different (Table 3). Nephropathy, uremia, and anemia were the three biggest risks for hypertension patients who were younger than 39 years, while renal insufficiency was a potential risk to hypertension patients who were younger than 29 years. Hyperlipemia was always in the top five comorbidities through all age groups and was the top comorbidity in the 40–49-year age group. Additionally, CHD, diabetes, and arteriosclerosis became a major risk when hypertension patients were older than 40 years. Another significant risk for older hy-

pertension patients was cerebral infarction being ranked in the top five comorbidities between 50 and 89 years of age and the fourth at 90–99 years of age.

The age-specific detection rates of the top 20 comorbidities of hypertension (Figure 2) were clustered into five classes. First, the age-specific detection rates of CHD, arteriosclerosis, cerebral infarction, insomnia, arrhythmia, gastritis, osteoarthropathy, and arthritis gradually increased as patients got older. The detection rates of these comorbidities at 90–99 years were several times (relative ratio: CHD, 25.91; arteriosclerosis, 22.65; cerebral infarction, 18.55; insomnia, 12.58; arrhythmia, 8.17; gastritis, 3.03; osteoarthropathy, 62.16; and arthritis, 8.02) higher than those at the age of 10–20 years.
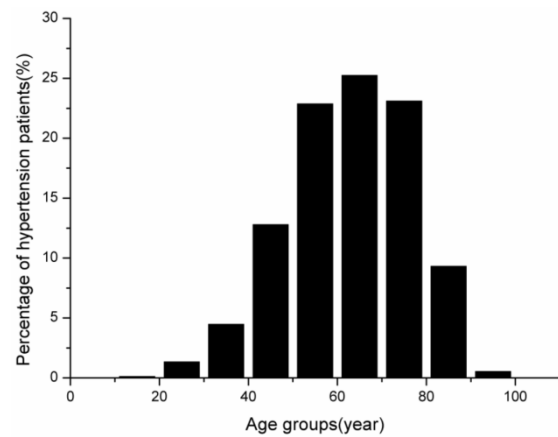

Figure 1. Age-specific distribution of hypertension patients in China.

Table 2. Sex-specific distribution of the top 20 comorbidities of hypertension in China.

No.	Comorbidity	Male Detection Rate(%)	95% CI	Female Detection Rate(%)	95% CI	Odds ratios	95% CI	p-value
1	Coronary Heart Disease	21.49	21.44-21.53	21.95	21.90-22.00	0.973	0.970-0.977	<.00001
2	Diabetes	16.24	16.20-16.28	15.74	15.70-15.78	1.038	1.034-1.042	<.00001
3	Hyperlipemia	13.86	13.82-13.89	13.76	13.72-13.80	1.008	1.003-1.012	0.00057
4	Arteriosclerosis	12.25	12.22-12.29	13.08	13.05-13.12	0.928	0.923-0.932	<.00001
5	Cerebral Infarction	8.17	8.14- 8.19	6.86	6.83- 6.89	1.207	1.200-1.215	<.00001
6	Move With Difficulty	3.82	3.80- 3.84	4.92	4.90- 4.94	0.767	0.761-0.773	<.00001
7	Nephropathy	4.58	4.56- 4.60	3.88	3.86- 3.90	1.189	1.179-1.198	<.00001
8	Respiratory Tract Infection	3.76	3.74- 3.78	4.16	4.14- 4.18	0.899	0.892-0.907	<.00001
9	Cerebral Circulation Insufficiency	3.24	3.22- 3.26	4.54	4.81- 4.56	0.704	0.698-0.710	<.00001
10	Upper Respiratory Tract Infection	3.36	3.34- 3.37	3.51	3.49- 3.53	0.953	0.945-0.962	<.00001
11	Renal Insufficiency	3.72	3.70- 3.74	2.75	2.73- 2.76	1.368	1.355-1.380	<.00001
12	Tracheitis	3.03	3.02- 3.05	3.17	3.15- 3.19	0.954	0.946-0.963	<.00001
13	Osteoporosis	2.24	2.23- 2.26	3.88	3.86- 3.91	0.568	0.563-0.573	<.00001
14	Insomnia	2.4	2.38- 2.41	3.35	3.33- 3.37	0.708	0.702-0.715	<.00001
15	Uremia	3.03	3.01- 3.05	2.41	2.40- 2.43	1.264	1.252-1.276	<.00001
16	Anemia	2.4	2.38- 2.42	2.45	2.43- 2.47	0.979	0.969-0.988	2.5E-05
17	Arrhythmia	2.29	2.28- 2.31	2.5	2.48- 2.51	0.917	0.908-0.927	<.00001
18	Gastritis	2.12	2.10- 2.14	2.41	2.39- 2.43	0.877	0.868-0.886	<.00001
19	Osteoarthropathy	1.7	1.69- 1.72	2.32	2.30- 2.34	0.729	0.721-0.737	<.00001
20	Arthritis	1.64	1.62- 1.65	2.3	2.29- 2.32	0.706	0.698-0.714	<.00001

Table 3. Top five comorbidities of hypertension in each age group.

| Age Group	Top 1	Top 2	Top 3	Top 4	Top 5
10-19	Nephropathy	Uremia	Anemia	Hyperlipemia	Renal Insufficiency
20-29	Nephropathy	Uremia	Anemia	Renal Insufficiency	Hyperlipemia
30-39	Nephropathy	Hyperlipemia	Uremia	Anemia	Diabetes
40-49	Hyperlipemia	Diabetes	CHD	Arteriosclerosis	Nephropathy
50-59	CHD	Diabetes	Hyperlipemia	Arteriosclerosis	Cerebral Infarction
60-69	CHD	Diabetes	Hyperlipemia	Arteriosclerosis	Cerebral Infarction
70-79	CHD	Diabetes	Arteriosclerosis	Hyperlipemia	Cerebral Infarction
80-89	CHD	Diabetes	Arteriosclerosis	Hyperlipemia	Cerebral Infarction
90-99	CHD	Arteriosclerosis	Diabetes	Cerebral Infarction	Hyperlipemia
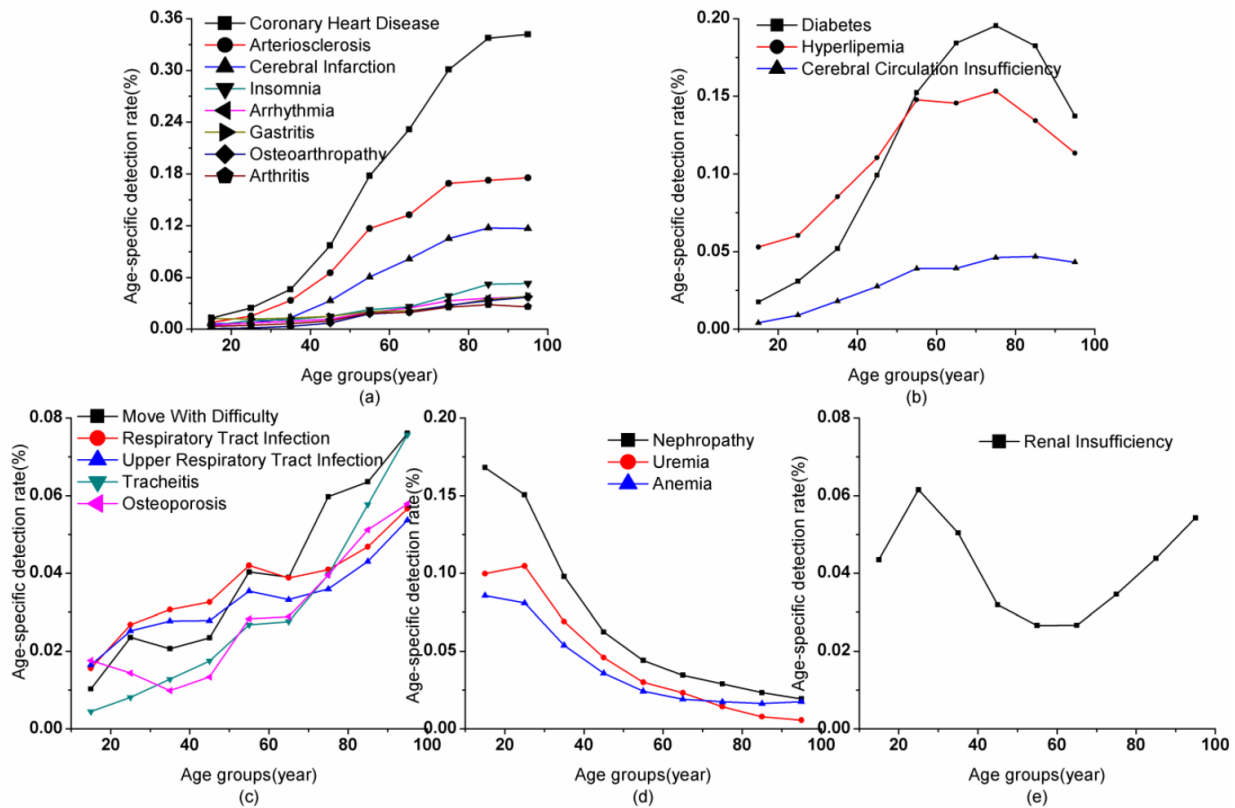
Figure 2. Age-specific patterns of the top 20 comorbidities of hypertension in China.

Second, the age-specific detection rate of diabetes, hyperlipemia, and cerebral circulation insufficiency increased with an increase in age but decreased in older patients. For diabetes and hyperlipemia, the detection rate reached a peak at 70–79 years, with detection rates of 19.55% and 15.33%, respectively. The detection rate of diabetes continued to increase over time, with the highest detection at 40–49 years. However, the detection rate of hyperlipemia flattened off from 50–79 years (14.56–15.33%). Moreover, the detection rate of cerebral circulation insufficiency flattened off from 50–69 years (3.92%) and then peaked at 70–89 years (4.62–4.70% for the two age groups). The detection rates of these three comorbidities greatly decreased in older people after the peak (diabetes: 29.72%; hyperlipemia: 26.01%; and cerebral circulation insufficiency: 8.05%).

Third, moving with difficulty, respiratory tract infection, upper respiratory tract infection, tracheitis, and osteoporosis had different rates of detection rate increase, and occasionally showed a slight decline depending on age. The detection rate of moving with difficulty greatly increased at 20–29 years (228.86%), 50–59 years (172.18%), and 70–79 years (153.28%) compared with the previous age group. Respiratory tract infection and upper respiratory tract infection had the same patterns in detection rate of greatly increasing at 20–29 years (171.34% and 153.13%, respectively) and 50–59 years (128.87% and 127.20%, respectively). The detection rate of tracheitis greatly increased at almost all age ranges (131.13–181.66%) and declined at 60–69 years (103.18%) compared with previous age groups. The detection rate of osteoporo-

sis was also unique in that it decreased below 40 years and quickly increased by 50–59 years (212.36%).

Fourth, the trend for detection rate did not always show an upward trend. The detection rate of kidney-related diseases continued to fall at most age ranges. The age-specific detection rates of nephropathy, uremia, and anemia fell from 16.81% (10–19 years) to 1.94% (90–99 years), 10.47% (20–29 years) to 0.56% (90–99 years), and 8.57% (10–19 years) to 1.64% (80–89 years), respectively. The decline in detection rate of these diseases in each age group compared with the previous age group was similar.

The last class only contained renal insufficiency whose detection rate reached a peak at 20–29 years (6.15%) and showed a U-shaped curve with increasing age. At 50–69 years old, hypertension patients had the lowest risk in developing renal insufficiency with a detection rate of only 2.66%.

Comorbidity network of hypertension

The comorbidity network comprising high co-occurrence frequency relationships among comorbidities of hypertension was presented in Figure 3 and Table S1. The core of the network included CHD, hyperlipemia, arteriosclerosis, and diabetes whose degree centrality was 13, 7, 7, and 6, respectively. Those comorbidities were directly connected to 75% of all comorbidities. Therefore, hypertension patients who had one of those four comorbidities had a greater health risk. Uremia and anemia were connected to the core network through nephropathy, which indicated that nephropathy was an important indicative variable between those two comorbidities and the core network. Hypertension patients with CHD, hyperlipemia, arteriosclerosis, and diabetes had a relatively low risk of developing uremia and anemia. Gastritis and the comorbidity pair of arthritis and osteoarthropathy were isolated in the core network. The morbidity risk of these three comorbidities was relatively independent to comorbidities in the core network. Moreover, because the average degree was 3.3 and the average path length was 2.09 in the comorbidities network, the comorbidity network showed that the top 20 comorbidities had a strong correlation with each other. Each top 20 comorbidity was directly connected to an average of 3.3 other comorbidities and the number of comorbidities between any two comorbidities was only approximately one.

Discussion

In our study, we obtained a large collection of electronic medical records from 106 prestigious hospitals located in 72 cities in China. The data collection process was mostly automatic and involved little human intervention. The automation of data collection ensured individual patient's medical records to be reliable, extensive, and timely. Compared with other similar research [1, 19, 20, 22], our study was based on a much larger patient base. Our data records were collected while patients were hospitalized, thus the records contained detailed and extensive coverage on patient's medical-related information. Because the medical-related information was for medical diagnostic purposes, the information was highly reliable and objective.
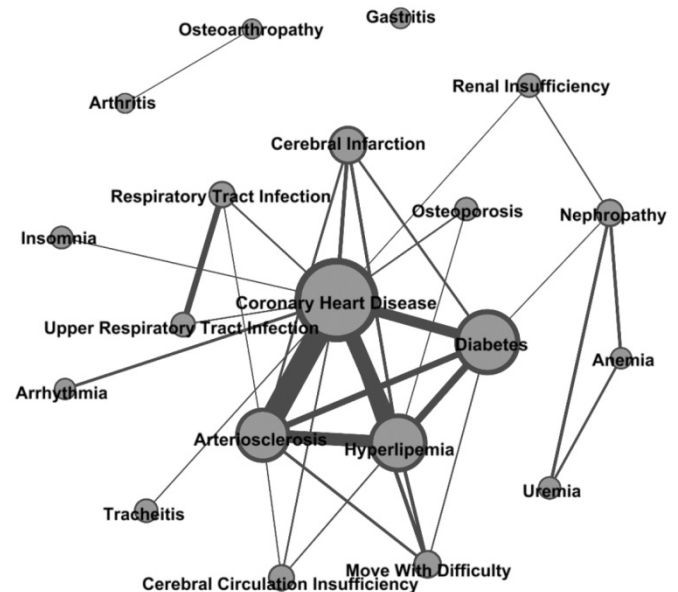

Figure 3. Comorbidity network of hypertension.

To the best of our knowledge, this study was the first to investigate the prevalence of comorbidities of hypertension through a large amount of electronic medical record data rather than using just medical survey or census data. Our findings on the detection rates of comorbidities were sufficiently representative for Chinese population and were insightful for doctors and hypertension patients. The top 20 comorbidities in terms of the detection rate and their co-occurrence relationships implied important health risks to hypertension patients. From our study, more targeted measures can be taken into consideration in order to prevent the deterioration of health of hypertension patients.

The sex-specific and age-specific detection rates of comorbidities described the different risks of comorbidities in hypertension patients with a different sex and age range. We found that female hypertension patients were more likely suffer from osteoporosis, while male hypertension patients were more likely to develop renal insufficiency. Nephropathy, uremia, and anemia were important risk factors in hypertension patients younger than 39 years, while

CHD, diabetes, hyperlipemia, arteriosclerosis, and cerebral infarction were high risk factors in older hypertension patients. Those findings can provide guidelines for the prevention of comorbidities of hypertension. An example of a preventative measure is with diabetes, one of the most frequently observed comorbidities of hypertension, where reducing sugar intake is a common proposal for hypertension patients. In addition, kidney disease is the most important risk factor in young hypertension patients. Therefore, patient's sex and age should be accounted in when proposing prevention measures for hypertension patients.

Currently, medically-aided diagnostic technologies primarily focus on preliminary statistics and a probabilistic computing system. In the era of large amounts of data, network-based recommendation technology continues to be developed. The characteristics of comorbidities that are calculated from a comorbidity network could provide valuable information about the relationships between each comorbidity pair. Our findings could rapidly promote the development of new diagnostic technologies, not only for hypertension, but also for other diseases.

In the current study, the high-frequency co-occurrence relationships among comorbidities of hypertension were analyzed and presented by the comorbidity network. Relationships with a high edge weight indicated those two comorbidities had a high co-existing correlation. The core network comprising the top four comorbidities verified a strong co-occurring relation and high risk of those four comorbidities to hypertension patients. Anemia and uremia had a relatively lower relevance with comorbidities in the core network than nephropathy did. Moreover, arthritis, osteoarthropathy, and gastritis had a relatively independent morbidity risk with the core network. In overall, the high-frequency co-occurrence relationships among comorbidities could be important for prevention and treatment of hypertension and its comorbidities.

There are limitations of this study. First, the detection rates achieved from different cities had shown different patterns. Further study, such as spatial analysis, could reveal the reasons for those differences. Second, some ambiguous or casually typed records were ignored due to the insufficiency of the natural language processing techniques that we had utilized. Adopting more effective text mining tools might increase the validity of rules that we used and the likelihood of finding new rules. Third, some seemingly unrelated or undetected patient symptoms might not have been completely and thoroughly recorded in the system. More detailed inspections on the medical data collection process need to be executed to ensure a more comprehensive data collection.

Conclusions

In summary, our analysis of comorbidities of hypertension in China between 2011 and 2013 provided an overview of the detection rate of comorbidities among hypertension patients. Variate detection rates of comorbidities regarding age and sex were presented, and the co-occurrence relationships among comorbidities were analyzed. Our findings can support doctors and patients to make more specific diagnoses and treatment plans by considering patient's age, sex and comorbidity conditions. Our results can also increase people's awareness of the comorbidities of hypertension. Further study on hypertension and its comorbidities will likely improve the life quality of hypertension patients, and be helpful for the prevention of hypertension.

Abbreviations

CHD: Coronary Heart Disease; CI: Confidence Interval.

Supplementary Material

Figures S1-S2, Table S1.
http://www.medsci.org/v13p0099s1.pdf

Acknowledgements

Competing interests

The authors declared that they had no competing interests.

References

1. Gu DF, Reynolds K, Wu XG, Chen J, Duan XF, Muntner P, et al. Prevalence, awareness, treatment, and control of hypertension in China. Hypertension. 2002; 40: 920-7.
2. Wolf-Maier K, Cooper RS, Banegas JR, Giampaoli S, Hense HW, Joffres M, et al. Hypertension, prevalence and blood pressure levels in 6 European countries, Canada, and the United States. JAMA. 2003; 289: 2363-9.
3. He J, Gu DF, Wu XG, Reynolds K, Duan XF, Yao CH, et al. Major causes of death among men and women in China. N Engl J Med. 2005; 353: 1124-34.
4. Sheng CS, Liu M, Kang YY, Wei FF, Zhang L, Li GL, et al. Prevalence, awareness, treatment and control of hypertension in elderly Chinese. Hypertens Res. 2013; 36: 824-8.
5. Al-Tuwijri AA, Al-Rukban MO. Hypertension control and co-morbidities in primary health care centers in Riyadh. Ann Saudi Med. 2006; 26: 266-71.
6. Hirani V, Zaninotto P, Primatesta P. Generalised and abdominal obesity and risk of diabetes, hypertension and hypertension-diabetes co-morbidity in England. Public Health Nutr. 2008; 11: 521-7.
7. Wang R, Zhao Y, He X, Ma X, Yan X, Sun Y, et al. Impact of hypertension on health-related quality of life in a population-based study in Shanghai, China. Public Health. 2009; 123: 534-9.
8. in't Veld AJM. Symptomatic BPH and hypertension: Does comorbidity affect quality of life? Eur Urol. 1998; 34: 29-36.

9. Aung T, Bisognano JD, Morgan MA. Allergic respiratory disease as a potential co-morbidity for hypertension. Cardiol J. 2010; 17: 443-7.
10. Prudenzano MP, Monetti C, Merico L, Cardinali V, Genco S, Lamberti P, et al. The comorbidity of migraine and hypertension. A study in a tertiary care headache centre. J Headache Pain. 2005; 6.
11. Dzudie A, Kengne AP, Mbahe S, Menanga A, Kenfack M, Kingue S. Chronic heart failure, selected risk factors and co-morbidities among adults treated for hypertension in a cardiac referral hospital in Cameroon. Eur J Heart Fail. 2008; 10: 367-72.
12. Channanath AM, Farran B, Behbehani K, Thanaraj TA. State of Diabetes, Hypertension, and Comorbidity in Kuwait: Showcasing the Trends as Seen in Native Versus Expatriate Populations. Diabetes Care. 2013; 36: E75-E.
13. Weiderpass E, Persson I, Adami HO, Magnusson C, Lindgren A, Baron JA. Body size in different periods of life, diabetes mellitus, hypertension, and risk of postmenopausal endometrial cancer (Sweden). Cancer Causes Control. 2000; 11: 185-92.
14. Lukas A, Kumbein F, Temml C, Mayer B, Oberbauer R. Body mass index is the main risk factor for arterial hypertension in young subjects without major comorbidity. Eur J Clin Invest. 2003; 33: 223-30.
15. Uretsky S, Messerli FH, Bangalore S, Champion A, Cooper-DeHoff RM, Zhou Q, et al. Obesity paradox in patients with hypertension and coronary artery disease. Am J Med. 2007; 120: 863-70.
16. Bixler EO, Vgontzas AN, Lin HM, Ten Have T, Leiby BE, Vela-Bueno A, et al. Association of hypertension and sleep-disordered breathing. Arch Intern Med. 2000; 160: 2289-95.
17. Sarafidis PA, Li S, Chen SC, Collins AJ, Brown WW, Klag MJ, et al. Hypertension awareness, treatment, and control in chronic kidney disease. Am J Med. 2008; 121: 332-40.
18. Abougalambou SSI, Abougalambou AS. A study evaluating prevalence of hypertension and risk factors affecting on blood pressure control among type 2 diabetes patients attending teaching hospital in Malaysia. Diabetes Metab Syndr. 2013; 7.
19. Wu YF, Huxley R, Li LM, Anna V, Xie GQ, Yao CH, et al. Prevalence, Awareness, Treatment, and Control of Hypertension in China Data from the China National Nutrition and Health Survey 2002. Circulation. 2008; 118: 2679-86.
20. Wang J, Ning X, Yang L, Lu H, Tu J, Jin W, et al. Trends of hypertension prevalence, awareness, treatment and control in rural areas of northern China during 1991-2011. J Hum Hypertens. 2014; 28: 25-31.
21. Schillaci G, Pirro M, Vaudo G, Gemelli F, Marchesi S, Porcellati C, et al. Prognostic value of the metabolic syndrome in essential hypertension. J Am Coll Cardiol. 2004; 43: 1817-22.
22. Wong ND, Lopez V, Tang S, Williams GR. Prevalence, treatment, and control of combined hypertension and hypercholesterolemia in the United States. Am J Cardiol. 2006; 98: 204-8.
23. Beckett NS, Peters R, Fletcher AE, Staessen JA, Liu LS, Dumitrascu D, et al. Treatment of hypertension in patients 80 years of age or older. N Engl J Med. 2008; 358: 1887-98.
24. Tanushi H, Dalianis H, Nilsson GH. Calculating prevalence of comorbidity and comorbidity combinations with diabetes in hospital care in sweden using a health care record database. 3rd International Workshop on Health Document Text Mining and Information Analysis 2011, LOUHI 2011, July 6, 2011 - July 6, 2011. Bled, Slovenia: Sun SITE Central Europe CEUR-WS; 2010: 59-65.
25. Kilander L, Nyman H, Boberg M, Hansson L, Lithell H. Hypertension is related to cognitive impairment - A 20-year follow-up of 999 men. Hypertension. 1998; 31: 780-6.
26. Barabasi AL. Network medicine - From obesity to the "Diseasome". N Engl J Med. 2007; 357: 404-7.
27. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL. The human disease network. Proc Natl Acad Sci U S A. 2007; 104: 8685-90.
28. Hidalgo CA, Blumm N, Barabási A-L, Christakis NA. A Dynamic Network Approach for the Study of Human Phenotypes. PLoS Comput Biol. 2009; 5: e1000353.
29. Shi J, Hu M, Shi X, Dai G-Z. Text segmentation based on model LDA. Chin J Comp. 2008; 31: 1865-73.
30. Fu GH, Kit C, Webster JJ. Chinese word segmentation as morpheme-based lexical chunking. Inf Sci. 2008; 178: 2282-96.
31. Sanner MF. Python: A programming language for software integration and development. J Mol Graph Model. 1999; 17: 57-61.
32. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update. ACM SIGKDD explorations newsletter. 2009; 11: 10-8.
33. Moon TK. The expectation-maximization algorithm. ISPM. 1996; 13: 47-60.
34. Roque FS, Jensen PB, Schmock H, Dalgaard M, Andreatta M, Hansen T, et al. Using Electronic Patient Records to Discover Disease Correlations and Stratify Patient Cohorts. PLoS Comput Biol. 2011; 7.
35. Jensen AB, Moseley PL, Oprea TI, Ellesoe SG, Eriksson R, Schmock H, et al. Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients. Nat Commun. 2014; 5: 4022.
36. Bavelas A. Communication patterns in task-oriented groups. J Acoust Soc Am. 1950; 22: 723-30.
37. Newman MEJ. The structure of scientific collaboration networks. Proc Natl Acad Sci U S A. 2001; 98: 404-9.
38. Opsahl T, Agneessens F, Skvoretz J. Node centrality in weighted networks: Generalizing degree and shortest paths. Social Networks. 2010; 32: 245-51.
39. Tang CL, Wang WX, Wu X, Wang BH. Effects of average degree on cooperation in networked evolutionary game. EPJB. 2006; 53: 411-5.
40. Fronczak A, Fronczak P, Holyst JA. Average path length in random networks. PhRvE. 2004; 70.