Study on Grouping Model of Diseases in Traditional Chinese Medicine Hospital Based on Two-Step Clustering Method

Huang Zhengzheng[1], Yang Sirui[1,*]

[1]Hubei University of Chinese Medicine, Wuhan, Hubei, China
*Corresponding author

Abstract: *The purpose of this study is to explore the applicability of the current CHS-DRG grouping and payment scheme in Traditional Chinese Medicine (TCM) hospitals by grouping the discharged cases in TCM hospitals. It also aims to provide a method reference for discussing the related grouping of TCM disease and syndrome diagnosis. The first-page data of inpatient medical records in some TCM hospitals above the county level, from 2015 to 2020, were taken as the research object. The first-page data of inpatient medical records from 2015 to 2017 were grouped according to the CHS-DRG grouping rules, and the grouping results were obtained. Subsequently, two-stage clustering methodology was used to further group the bad groups, and the TCM disease grouping model was preliminarily formed. The model was verified by the first-page data of inpatient medical records from 2018 to 2020. The results showed that under the CHS-DRG grouping model, 494,150 cases were divided into 359 DRG groups. Only 92 cases with CV < 1 and cases > 100 were grouped poorly. After adding TCM diagnostic codes into the grouping model, the grouping effect of the disease group with CV > 1 was significantly increased, and the grouping results met the CHS-DRG grouping criteria. To fully leverage the benefits of Traditional Chinese Medicine and provide simple, convenient, empirical, and inexpensive medical services, TCM hospitals can explore TCM case grouping models that conform to the characteristics of TCM. In case grouping, TCM characteristic diagnosis and treatment elements should be included, and the study of medical insurance payment mode suitable for TCM hospitals should be conducted. Additionally, improving the hospital hardware facilities and providing an intelligent, automatic medical record management system should be prioritized. Further standardizing TCM diagnosis and code filling in the medical record information system and strengthening the study of TCM medical record management and coding will improve the quality of medical record.*

Keywords: *Hospital of traditional Chinese medicine, Diagnosis of disease related groups, Two Step Clustering*

1. Introduction

TCM hospitals are medical institutions that provide diagnosis and treatment services for patients with TCM characteristic technologies to meet people's needs for TCM services, and are the main carriers for the inheritance and development of TCM. According to the National Extracts of Tradition Chinese Medicine Statistics 2019[1], TCM hospitals account for 15.23% of the total number of hospitals, 15.26% of the number of patients, 13.58% of the number of beds, and 16.4% of the total number of medical services. Compared with the previous year, TCM service capacity has been highlighted to some extent, but the proportion of general hospitals and medical services provided is still small. In addition, the development and innovation of TCM diagnosis and treatment technology is relatively slow, the proportion of TCM services is slightly lower, the brain drain of TCM, and TCM hospitals are gradually "westernized", which greatly limits the demand and supply of TCM. The main reasons are that TCM diagnosis and treatment service items are not covered by medical insurance reimbursement, the service cost of TCM and the labor value of TCM medical staff cannot be effectively reflected. Therefore, it is necessary to have a reasonable and effective way of medical insurance compensation to improve the development dilemma of TCM hospitals and encourage medical staff and patients to choose TCM services.

In October 2019, the National Medical Insurance Administration officially released the Technical Specifications for National Medical Insurance DRG Grouping and Payment and the National Medical Insurance DRG (CHS-DRG) Grouping Scheme, which provided the grouping criteria and technical specifications for the pilot cities to carry out the DRG pilot project smoothly [2]. In May 2019, traditional

medicine originating from TCM was officially included in the 11th Revision of the International Classification of Diseases (ICD-11). On October 11, 2021, the revised "Classification and codes of diseases and ZHENG of traditional Chinese medicine " (GB/T 15657-2021) was released, which added 741 names of TCM diseases and 974 names of TCM syndromes on the basis of the 1995 version, improving the coding database of TCM diseases and syndromes, and providing a basis for TCM hospitals to implement related grouping of disease and syndrome diagnosis [3].

TCM diagnosis and treatment stresses the holistic concept, which is compatible with the "package and pay" method of grouping according to disease diagnosis [4]. The DRG payment system shifts medical risks from healthcare payers to medical service providers. Medical institutions providing medical services will thereby take the initiative to control costs and prioritize choosing doctors who use traditional Chinese medicine diagnostic and treatment techniques that require lower costs and fewer resources. This payment method can effectively stimulate medical institutions to provide more traditional Chinese medicine services and overcome the challenge of low demand. It may also generate more residual income for the hospital. Therefore, in terms of benefits, DRGs is beneficial to the development of TCM hospitals. The purpose of this paper is to explore the suitability of the current CHS-DRG grouping and payment scheme for TCM hospitals, and to provide a methodological reference for further discussion of TCM disease and syndrome diagnosis related grouping model.

2. Data and methods

2.1 Data Source

In this study, the first page data of inpatient medical records with dual diagnosis information of TCM and Western medicine in some TCM hospitals above county level in a province from 2015 to 2020 were taken as the research object. The first page data of medical records from 2015 to 2017 were used as modeling data, and the first page data from 2018 to 2020 were used as validation data. Length of stay, outbound and admission diagnosis and corresponding diagnosis codes, surgical procedures and surgical operation codes, total hospitalization costs and details were selected. Incomplete cases were screened according to data cleaning rules, and 503,224 valid data were included. Data cleaning rules: First, exclude key information of the first page of inpatient medical records, such as gender, age, western medicine code, TCM disease name code, TCM syndrome code, missing total hospitalization cost and data records with logical errors, such as TCM disease name code filling in western medicine code, and then eliminate extreme values according to DRG rules. Patients with hospitalization duration less than 1 day or more than 100 days and hospitalization cost less than 5 yuan [5]. In the medical record system of TCM hospitals from 2015 to 2020, the diagnostic codes of TCM diseases and syndromes still adopt the 1995 version of the National standard "Classification and codes of diseases and ZHENG of traditional Chinese medicine " (GB/T 15657-1995, TCD). The disease diagnosis codes are strictly in accordance with ICD-10 and TCD, and the operation codes are in accordance with ICD-9-CM-3.

2.2 Research method

Firstly, the TCM disease and Syndrome diagnosis related grouping database system developed by the project team was used to sort out and analyze the data related to the first page of medical records. Then, the self-programmed program was used to group the data related to the first page of medical records according to the grouping principles and standards of CHS-DRG. Finally, combined with the actual grouping effect, the TCM disease name coding was again included as the influencing factor of TCM DRG grouping, and the two-stage clustering method was adopted to further subdivide the disease groups that did not meet the grouping criteria, so as to minimize the differences within the subdivided groups and maximize the differences between the subdivided groups [6].

Due to the selection of indicators both categories in our study type variables such as TCM diagnosis codes, the indicators have numeric variables such as total hospitalization expenses, so such as common hierarchical clustering method, statistical method is not applicable to class such as k-means clustering type variable, and a Two Step Clustering method can simultaneously to clustering on the two types of data, automatically determine the final classification number. The Two Step Clustering method is divided into two steps. The first step is pre-clustering and preliminary classification of cases. The second step is formal clustering, re-clustering the preliminary categories obtained in the first step and determining the final clustering scheme. Two Step Clustering method has a distinct characteristics of the variables can be adopted in clustering analysis is a continuous variables and discrete variables, however, for large sample

data computing speed, using statistics as a clustering distance indicators, and also can be according to certain standard to automatically determine the optimal class number, make the results of clustering validity more assured, It meets the data type and statistical requirements of this study. In this study, BIC was used as the clustering rule. The smaller BIC was, the better the clustering effect was.

2.3 Case combination evaluation index

Coefficient of variation (CV) was used to evaluate the homogeneity within the group, and variance reduction (RIV) was used to evaluate the heterogeneity between each disease group [7]. CV is the ratio of standard deviation to mean of a group of data. The smaller CV is, the smaller the difference of each case in the combination is, and the better the grouping effect is. RIV is the ratio of the sum of squares of mean deviations of subsets to the sum of squares of mean deviations of the population in a set of data. The larger the RIV is, the greater the difference between different disease groups is, and the better the grouping effect is [8].

3. Data analysis results

3.1 Basic grouping

According to the technical specification for CHS-DRG grouping and payment, the DRG group with no less than 100 cases in the disease group, coefficient of variation (CV) < 1, and the relative difference of average cost between DRG groups within the same ADRG subdivision no less than 20% was considered to have a good grouping effect [9]. The first page data of inpatient records collected from some TCM hospitals above the county level in a province during 2015-2017 were grouped by using the CHS-DRG grouping scheme, and a total of 359 groups were obtained. Among them, there were 193 disease groups with CV less than 1. 153 patients with CV greater than or equal to 1; as the number of cases in 13 disease groups was less than or equal to 1, there was no CV. There were only 92 effective disease groups with more than 100 cases and CV less than 1.

3.2 Two step cluster analysis

The effect of CHS-DRG on direct grouping of inpatients in TCM hospitals was not good, and only 25.63% of the patients were effective. Therefore, a TCM case grouping model was proposed to further adjust the grouping scheme and optimize the grouping results. TCM pays attention to giving treatment on the basis of syndrome differentiation. The first page of TCM medical records not only requires the diagnosis of TCM diseases, but also requires the diagnosis of TCM syndrome with characteristics, and determines the diagnosis and treatment path and medication of diseases according to the syndromes. Therefore, in this study, the case combinations with the number of cases greater than 5000 and CV greater than 1 were selected from 153 poorly grouped disease groups, as shown in Table 1. Based on the original ADRG grouping, further subgroups were added into TCM disease name coding, and cases were grouped by two-step clustering method using SPSS.25.0 software.

Table 1: Disease groups with CV greater than 1 and case number greater than 6000

| Grouping | Group name | CV | Case load | Cost | | | | |
				Average value	Median	Standard deviation	Minimum	Maximum
BR29	Ischemic disorders of the brain	113.60%	43842	6881.706	5063.995	7814.981	6.90	398829.55
CW19	Various types of cataract	102.50%	6475	5836.465	4926.31	5981.518	32.05	140450.31
ES29	Respiratory infection/inflammation	118.50%	22885	3742.021	2795.51	4433.028	15	182774.88
EZ19	Other respiratory disorders	165.50%	5265	7997.755	4860	13237.564	30.9	191150.4
FR49	Coronary atherosclerosis/thrombosis/ occlusion	116.50%	21825	7140.047	5160.64	8314.94	21.71	203704.49
LX15	Urinary calculi, obstruction and urethral stricture without complication or complication	100.20%	10695	4577.756	2740	4588.167	6	53950.91
IU29	Neck, waist and back disorders	103.40%	66895	5071.061	3992.74	5242.219	10.24	201929.6

Table 2: Two-step cluster grouping of BR29 patients from 2015 to 2017

Grouping	Case load	Average value	Standard deviation	Variance	CV	RIV
1	3915	20031.58	19866.11	394662524.30	99.20%	28%
2	39927	5592.31	3123.28	9754899.99	55.80%	

Table 3: Two-step cluster grouping of BR29 patients from 2015 to 2017

Grouping	Case load	Average value	Standard deviation	CV	RIV
1	2907	22302.508	21476.947	96.30%	39%
2	46005	6645.074	4158.83	62.60%	

Due to limited space, this paper selected BR29 (Ischemic disorders of the brain) as a typical case for subgroup study, and detailed analysis results are shown in Table 2. The BR29 group was subdivided into two groups by two-step clustering. The CV decreased from 113.60% to less than 1 for each subgroup, and the number of cases in each subgroup was more than 100. The number of cases and CV in each subgroup met the grouping criteria. The RIV of 28% was obtained by variance calculation of group deviation and total deviation, and the difference between groups was large. The overall grouping effect was good, and the grouping result met the requirements of CHS-DRG scheme.

3.3 Grouping result verification

To ensure the validity of the above research results, the data of BR29 disease group in the inpatient disease group of TCM hospitals from 2018 to 2020 were used for validation. CHS-DRG scheme was used to directly group the inpatients in TCM hospitals from 2018 to 2020, and the CV of the BR29 group was 111.70%. The TCM case grouping model was used, including the factors of TCM disease name coding, and the disease group was subdivided into two disease groups by two-step clustering method, as shown in Table 3. The CV of the subdivision disease group was all less than 1, the number of cases included in each subgroup was more than 100, and the RIV value was more than 20%. The grouping effect was good, and the grouping results were consistent with those from 2015 to 2017, indicating that the grouping results of this subgroup were effective and convincing.

In conclusion, the current CHS-DRG grouping model is not suitable for inpatients in TCM hospitals, but the grouping effect is good when TCM indicators and other influencing factors are added into the grouping model. Therefore, the CHS-DRG grouping model applicable to TCM hospitals can be constructed by adding TCM characteristic indexes to provide decision-making for the optimization of medical insurance payment system in TCM hospitals.

4. Discussion and suggestions

4.1 Improve the TCM medical record information system and standardize the first page of medical records

After sorting out the data of the first page of medical records of TCM hospitals, it was found that there were 1183880 original case data, and 503,224 cases after screening according to the screening rules, indicating a significant decrease in effective case data. Invalid cases, a large number of TCM diagnosis and syndrome coding or fill in the missing errors, "Classification and codes of diseases and ZHENG of traditional Chinese medicine" is not found in the corresponding coding, diagnosis of traditional Chinese medicine hospital medical record home page are filled out coding errors, the information is not complete, the hospital diagnosis and hospitalization expenses key problem of lack of useful information, such as insufficient standard medical record information to fill in, is not conducive to the traditional Chinese medicine hospital medical record management. The data of DRGs grouping and inclusion cost calculation depend on the information related to the first page of medical records, especially the diagnosis and coding of diseases. Non-standard filling data will seriously affect the inclusion of cases, resulting in a lower case inclusion rate. Hospital of traditional Chinese medicine to promote with characteristic of TCM diagnosis related groups pay system, to standardize the clinical path, the specification of the medical record home page to fill in, guarantee the accuracy of the medical record information, still need to improve the medical record information system, make its intelligence, automation, attempts to automatically identify false information, effective medical record quality monitoring data.

4.2 Strengthen the study of TCM medical record management and coding

The first page of medical records not only requires standardized filling, but also requires the coder to

master professional coding knowledge. Among the 680,656 cases screened, the diagnosis codes given by coders were inconsistent with the diagnosis of diseases given by clinicians, and the reasons for filling in the wrong codes indicated that the professional knowledge of medical record coders in TCM hospitals was still lacking. The implementation of DRGs payment system has high requirements on the first page of medical records and disease coding, so the management personnel of TCM hospitals need to strengthen the management of medical records, formulate the filling standard and assessment mechanism of the first page of medical records, so as to ensure the accuracy and standardization of TCM medical record data and improve the quality of medical record data. TCM clinical staff and coders should strengthen the study of coding, master the basic theory of TCM, classification of TCM diseases and syndromes and their codes, improve the coding level, and pay more attention to TCM diagnosis and coding. Hospitals should also carry out coding training sessions regularly to provide opportunities for coding staff to learn and communicate, and to work together to solve coding problems encountered in order to cope with various requirements in clinical work.

4.3 Construct disease diagnosis related grouping model with TCM diagnosis and treatment characteristics

In this study, an empirical analysis of the CHS-DRG grouping scheme was conducted based on the first page data of inpatient medical records in TCM hospitals. It was found that the case grouping effect was not good and the case inclusion rate low, indicating that THE CHS-DRG grouping model was not suitable for TCM hospitals. The reason is that the CHS-DRG grouping scheme is designed based on the diagnosis and treatment mode of western medicine, and TCM hospitals have their unique diagnosis and treatment characteristics. TCM disease and syndrome diagnosis are not reflected in the grouping model, while the diagnosis of disease and syndrome is an important factor affecting the medical cost, so the current DRG grouping model is directly adopted. This will lead to the loss of diagnosis information of a large number of TCM diseases and syndromes in cases, and serious deviations in the inclusion of cases, failing to show the characteristics of "simplicity, convenience, experience and simplicity" of TCM, and even causing losses in TCM treatment fees. Based on the theory of TCM diagnosis and treatment, this study proposed the hypothesis that TCM disease name coding would affect the grouping effect and verified it. Combined with the CHS-DRG grouping strategy, the TCM syndrome diagnosis and syndrome coding were included in the subgroup, and the grouping effect was good, indicating that the CHS-DRG grouping idea can be considered in combination with the CHS-DRG grouping idea, and TCM disease and syndrome diagnosis can be included on the basis of this grouping idea, and then the DRGs with TCM characteristics can be subdivided. This grouping idea not only retains the original grouping principle. It also includes the characteristics of TCM diagnosis and treatment to facilitate the grouping of integrated TCM and Western medicine diagnosis and treatment mode. On top of that, there is an attempt to divide the TCM diagnosis and treatment system directly into MDC and operating segment ADRGs based on the diagnosis of traditional Chinese medicine syndromes. This innovative approach aligns better with the characteristics of TCM diagnosis and treatment and can reflect more accurately the advantages of TCM diagnosis and treatment technology. Furthermore, it can encourage doctors and the public to choose traditional Chinese medicine diagnosis and treatment methods and promote the development of TCM. However, putting this practice into action may be relatively difficult. In conclusion, if TCM hospitals want to explore the diagnosis-related grouping model applicable to TCM, it is necessary to consider the influence of TCM characteristic diagnosis and treatment mode on total hospitalization cost, and to construct a diagnosis-related grouping model with TCM characteristics by considering such factors as syndrome differentiation treatment, acupuncture and massage, and prescription.

5. Conclusion

The existing CHS-DRG scheme is not suitable for grouping inpatients in TCM hospitals. In the future, it is necessary to study the grouping scheme of characteristic TCM treatment methods, consider the characteristics of TCM, and provide direction and strategy for medical insurance payment mode of TCM hospitals.

References

[1] Department of Planning and Finance, National Administration of Traditional Chinese Medicine. The national statistics of Chinese medicine extracts [EB/OL]. http://www.satcm.gov.cn/2019tjzb/main.htm, 2020-03-26.

[2] Cui Bin, Zhu Zhaofang. Discussion on key Links of Formulation and Implementation of CHS-DRG [J]. China Medical Insurance, 2021, (05):47-51.

[3] GB/T 15657-1995, 2021, Classification and codes of diseases and ZHENG of traditional Chinese medicine [S].

[4] Liu H Y. Discussion on payment mode of TCM by disease [D]. Huazhong University of Science and Technology, 2013.

[5] Sun Yang, Zhou Da, Luo Bin. DRGs Design Principle and Application Practice Twelve lectures [M]. Hubei: Hubei Science and Technology Press, 2019.

[6] Chen Yin, Liu Shiyang, Sun Jing, et al. Research on inpatient satisfaction based on second-order clustering analysis [J]. Chinese journal of hospital management, 2018, 34(02):104-109.

[7] Sun Fei, Han Junyang, Zhang Wenqian, et al. Case Combination Study of Acute Myocardial lnfarction Based on Decision Tree Model [J]. Chinese Medical Record, 2021, 22(03):75-79.

[8] Zeng Yanbing, Lin Peng, Fang Ya. Study on DRGs grouping of patients with viral hepatitis based on CHAID algorithm [J]. Chin J health statistics, 2015, 32(03):514-517.

[9] National Healthcare Security Administration of the People's Republic of China. About print and distribute disease diagnosis related group (DRG) paid the national pilot specification and grouping scheme [EB/OL]. (2019-10-24). http://www.nhsa.gov.cn/art/2019/10/24/art_37_1878.html.