A China Healthcare Security Diagnosis Related Groups Script Development based on R Language

Junmin Zhu (✉ JuinZhu@outlook.com)

Fujian Medical University Affiliated Nanping First Hospital

Yunhai Fan

Fujian Medical University Affiliated Nanping First Hospital

Quanhui He

Fujian Medical University Affiliated Nanping First Hospital

Wanglin Peng

Fujian Medical University Affiliated Nanping First Hospital

Research Article

Keywords: China Healthcare Security Diagnosis Related Groups (CHS-DRG), R language, simulated grouping, accuracy

Posted Date: February 9th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1073001/v2>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Abstract

Background: To develop a set of R scripts that could efficiently and accurately identify the home page information of medical records and perform China Healthcare Security Diagnosis Related Groups (CHS-DRG) simulating grouping.

Methods: Based on the CHS-DRG grouping rules, we abstracted the DRG grouping process into a standard algorithm and compiled the R script Z-DRG. The DRG simulating groupings by Z-DRG were compared with the DRG results from the regional CHS-DRG integrated service platform to evaluate the accuracy.

Results: The Z-DRG includes one function module (zdrfun. R), one operation module (zdrpro. R) and one database form (zdr codes.RData). The function module set 7 algorithm steps and 8 custom functions. The functions were set for multiple diagnoses, multiple operations, joint diagnosis and operation. Only (17.85±0.11) milliseconds were taken for CHS-DRG simulating grouping of one case. Compared with the regional CHS-DRG results, the accuracy rate was 99.10%. The difference in the number of other diagnoses is the main reason that affected the accuracy.

Conclusions: Z-DRG is easy to operate. The CHS-DRG simulating groupings were efficient and accurate. The simulation results could be effectively applied for medical institutions to carry out CHS-DRG grouping prediction and improve the implementation effect of CHS-DRG payment work.

Background

Diagnosis-related groups (DRGs), one type of case mix classification, is a method that groups patients into several diagnostic groups based on factors such as age, sex, disease diagnosis, comorbidities, complications, treatment methods, outcomes, and resource consumption. It is an important tool for measuring medical quality, evaluating cost-effectiveness, and confirming medical insurance payments. To deepen the reform of medical insurance payments, the Chinese National Healthcare Security Administration promoted DRG Payment national pilot work in 30 cities in 2019 and enabled actual DRG payments in 2021.

At present, research on DRG mainly focuses on building statistical models to improve DRG rules^[1–2], using DRG data for medical care quality evaluation^[3–5], performance appraisal^[6], disease cost measurements^[7], medical behavior research^[8–9], and hospital healthcare^[10]. However, there are few studies on compiling the unified DRG scheme into executable programs for establishing DRG simulation grouping. Research has shown that medical institutions should clarify the principle of DRG, establish timely DRG simulation grouping, improve the quality of the home page of medical records, correct ICD coding, and standardize medical behaviors to ensure that medical institutions maintain healthy development in the DRG payment reform process^[11]. Therefore, we have developed a set of CHS-DRG

simulation grouping scripts based on R language, which could efficiently and accurately perform CHS-DRG simulation grouping, to promote CHS-DRG payment pilot work for medical institutions.

1 Methods

1.1 CHS-DRG grouping principle (Figure 1)

1) Major Diagnostic Categories (MDC)

Based on the main diagnosis of the case, considering the anatomical location, etiology and clinical manifestations, the case was classified as MDC. Organ transplantation, ventilator use ≥ 96 hours or ECMO cases were classified into MDCA (preliminarily grouped diseases and related operations). Cases aged less than 29 days were classified into MDCP (newborns and other neonates with conditions originating in the perinatal period). HIV cases were classified into MDCY (human immunodeficiency virus infections and related operations). Two or more severe traumas were classified into MDCZ (multiple significant trauma). MDCA, MDCP, MDCY and MDCZ were pre-major diagnostic categories (pre-MDCs), which were priority grouped in order. According to sex differences, we could classify reproductive system diseases into MDCM (male) or MDCN (female).

2) Adjacent Diagnosis Related Groups (ADRG)

According to the main diagnosis and main operation of the case, compared with the main diagnosis and main operation of ADRG, we could orderly enroll the case into surgical operation ADRG, nonsurgical operation ADRG and medical diagnosis ADRG.

3) Diagnosis Related Groups (DRG)

Matching all other diagnoses of the case and the MCC (Major Complication & Comorbidity) and CC (Complication & Comorbidity) tables, we determined that the properties of all other diagnoses were MCC, CC or no-CC (without complications and comorbidities). Additionally, we checked the main diagnosis exclusion table name and all other exclusion table names. The property of other diagnosis should turn into no-CC when the other exclusion table name was consistent with the main diagnosis exclusion table name. There were 198 sublists of exclusion tables in the CHS-DRG 1.0 revision. Finally, the cases were further divided into DRG with major complications and comorbidities, DRG with complications and comorbidities, and DRG without complications and comorbidities.

1.2 R scripts compilation

Our research team transformed the CHS-DRG grouping principles into executable scripts by using R language. We refined the CHS-DRG grouping principle into multiple executable algorithm steps and then clarified the indicator input and output of the core algorithm. The R functions were used for functioning the main algorithm. The R script Z-DRG was developed based on R 4.1.0 64-bit environment. It was

trouble-free running on the Windows 7 or Windows 10 operating system with 32-bit or 64-bit, and the software platform above version 4.1 of R language.

1.3 CHS-DRG simulation grouping accuracy evaluation

This study was approved by the Medical Research Ethics Committee of Fujian Medical University Affiliated Nanping First Hospital (NO. NPSY2020100006). The home page of medical records in Jan. to Sep. 2021 from a tertiary hospital was extracted. CHS-DRG simulation grouping was carried out. Comparing the simulation results with the actual DRG results from the regional CHS-DRG integrated service platform, the validity was evaluated with the accuracy rate. The grouping differences were analyzed to effectively improve Z-DRG operating effectiveness.

2 Results

2.1 CHS-DRG grouping process analysis

The CHS-DRG 1.0 revision was divided into 26 MDCs, 376 ADRGs and 618 DRGs. MDCA just considered surgical operations. MDCZ should be determined by combining the main trauma site and the others. Based on the MDCs that were located by the main diagnosis or MDCA, there were 7 surgical operation ADRGs, such as AH1 (Trachea incisive ventilator support ≥ 96 hours or ECMO) and JA1 (Radical resection of breast cancer with breast reconstruction). should combine the main operation and other operations to judge. Under the MDCs where the main diagnosis was located, 27 surgical operation ADRGs and 1 nonsurgical operation ADRG needed joint the main diagnosis and the main operation. Additionally, 127 surgical operation ADRGs and 21 nonsurgical operation ADRGs were based on the main operation, while 187 medical diagnosis ADRGs were based on the main diagnosis. Moreover, 6 surgical operation ADRGs, including YC1 (HIV with extensive operating room procedures), SB1 (Operating room procedures for systemic infection), TB1 (Operating room procedures for mental patients), VC1 (Debridement related to injury), WJ1 (Burns with any operating room procedures except skin grafting), and XJ1 (Other diagnoses of contact with health services accompanied by operating room procedures), should be considered operating room procedures or special surgical operations in these MDCs.

In CHS-DRG, according to the CCs that were MCC, CC or no-CC, 121 ADRGs were subdivided into DRG with MCC, DRG with CC, and DRG with no-CC, while 255 ADRGs directly into subdivision groups without any CCs.

2.2 CHS-DRG simulation grouping algorithm

We standardized the grouping process of CHS-DRG and developed a set of R scripts, named Z-DRG. The Z-DRG included a Function Module (zdrfun. Rc) for customizing functions and algorithms, an operation module (zdrpro. R), which was used to operate programs for compiling DRGs, and a database form (zdrpcodes. RData) to store the grouping information of CHS-DRG.

The function module set 8 custom functions (Table 1) and 7 algorithm steps (Figure 2). The operation module mainly loaded the function module and the database form, imported the home page data of medical records, batched DRG grouping, and generated a 17-column table to display the grouping results. The database form covered a diagnosis DRG grouping table, an operation DRG grouping table, and a CHS-DRG codes table.

The Z-DRG and CHS-DRG simulation grouping scripts were performed according to the following algorithm steps.

Step 1. Start up the R work platform. Importing the Database Form (zdrpcodes. RData) by the “load” function, loading the function module (zdrfun. Rc) by loadcmp of the “compiler” package and then checking the required packages, including “reshape2”, “dplyr”, “stringr” and “openxlsx”, we could be ready for loading the CHS-DRG process.

Step 2. Data preparation. The home page information exported from the Medical Records Management System was read into the R platform by the “read” function of the “openxlsx” package. According to the DRG grouping requirement, the home page data, which included medical records number, gender, age, hospitalization days, main diagnosis code main operation code, discharge method, total hospitalization cost, all other diagnosis codes, and all other operation codes, were extracted by the “dsdata” function. Matching the diagnosis DRG grouping table and the operation DRG grouping table from the database form (zdrpcodes. RData), the whole diagnosis code table and the whole operation code table were generated.

Step 3. Pre grouping. First, setting MDC to MDCA, the case was identified to determine whether it was grouped into one ADRG of MDCA through the “uni.s” and “adrg.sj” functions. Second, setting MDC to MDCP, the case was identified to determine whether it was grouped into one ADRG of MDCP by the “adrg.n” function. If the case was grouped into both PU1 (normal newborn) and PV1 (infant diseases diagnosed from neonates (29 days $\leq$ birth age <1 year old)), we could distinguish by whether the case’s age was less than 29 days. Third, setting MDC to MDCY, the case was identified to determine whether it was grouped into one ADRG of MDCY by the “adrg.dj” function. When the patient underwent surgery, we should prioritize determining whether it could group to YC1. Fourth, through the “uni.d” function, the main diagnosis and other diagnoses were combined to determine whether the MDCZ could be grouped. When MDC was MDCZ, the case was classified into the specific ADRG of the MDCZ by the “adrg.n” function. If no surgical ADRG of the MDCZ could be grouped, the case should be grouped into ZZ1 (multiple significant traumas with no operating room procedures).

Step 4. Normal grouping. According to the main diagnosis to set MDC. The “uni.s” function was used to determine whether the case could group into ADRGs that were combined surgical operations. The scripts were in normal grouping through the “adrg.n” function.

Step 5. Special grouping. If the case was grouped into MDCM, MDCN, MDCS, MDCT, MDCV, MDCW, or MDCX, the scripts would start this step. Sex was used to distinguish whether the MDC was MDCM or

MDCN. The operating room procedures were used to determine whether the patients were grouped into SB1, TB1, VC1, WJ1 or XJ1.

Step 6. ADRG discrimination. According to the ADRG results, we could determine the final ADRG by the order of pre grouping, normal grouping, and special grouping.

Step 7. DRG discrimination. If the ADRG of the case could directly correspond to DRG, then this case completed its DRG subdivision group. If it could not, the scripts began to start the CC judgment step. Comparing all other diagnosis exclusion table numbers with the main diagnosis exclusion table number from the whole diagnosis codes table from Step 2 to determine the CCs of all other diagnoses. If the exclusion table number of other diagnoses was the same as that of the main diagnosis, then the CCs of these other diagnoses should be set into no-CC, regardless of the original diagnosis. When the case had MCC, it could be grouped into DRG with MCC. When the case just had CC, it should be grouped into DRG with CC. Otherwise, it should be grouped into DRG without complications and comorbidities.

Finally, we integrated step 2 to step 7 into the “drgs.pro” function. Then, the operation module was carried out to obtain the CHS-DRG simulation grouping results. The result data frame was a 17-column table (table 2).

2.3 Accuracy analysis of CHS-DRG simulation grouping results

Using the Z-DRG, we performed CHS-DRG simulation grouping of 42523 cases in a tertiary hospital from Jan. to Sep. 2021. (17.85 ± 0.11) milliseconds on average was taken for one case. There were 35236 cases of the CHS-DRG grouping results of the healthcare security patients, which were downloaded from the regional CHS-DRG operation management system. The comparative analysis of the DRG grouping results showed that the CHS-DRG simulation grouping results were consistent with those of the regional CHS-DRG operation management system on MDCs and ADRGs. There were 34918 cases with consistent results in the DRG subdivision group, and the overall DRG grouping accuracy rate was 99.10%.

According to the CHS-DRG grouping principle, the root reason for ADRG grouping consistently with DRG grouping inconsistently lies in the difference in CCs (table 3). The difference analysis of these 318 cases with inconsistent DRG subdivision groups showed that 85.22% of cases underestimated CCs by Z-DRG, which might be because the CC (MCC and CC) tables from the regional CHS-DRG integrated service platform deviated from the CHS-DRG 1.0 revision. In addition, 14.78% were overestimated, which might be because the data used for Z-DRG contained many more other diagnoses.

3 Discussion

At present, we found few reports about using R language to establish a CHS-DRG grouping simulation program. Most software vendors chose to develop a CHS-DRG simulation grouping program for medical institution managers at high cost. Additionally, some vendors created a mobile version of a single case grouping tool to help coders of medical records by taking the main disease diagnosis code and the main

operation code to achieve the basic grouping. During the DRG payment period, the CHS-DRG grouping strategy was needed to help medical institution managers better adapt to the detection regional CHS-DRG grouping results to achieve the combination of DRG and hospital fine management. According to the purpose, we established a set of practical CHS-DRG simulation grouping scripts (Z-DRG) based on R language.

The Z-DRG had five advantages. **First, the Z-DRG was easy to run.** The medical record coders only need to import the home page data of medical records into the Z-DRG. The Z-DRG could automatically complete the code conversion and CHS-DRG simulation grouping. The Z-DRG did not require entry of the main diagnosis and main operation information manually one by one and avoided errors of manual CHS-DRG grouping. **Second, the results of Z-DRG were timely.** Medical record coders could perform the CHS-DRG simulation grouping by Z-DRG at any time without waiting for lagging results from the regional CHS-DRG operation service platform. Timely grouping results could improve the ICD coding level of medical record coders^[11] and provide timely data references for medical institutions to carry out DRG management. **Third, the Z-DRG was efficient and accurate.** The grouping test showed that the CHS-DRG simulation grouping of a single case only took 16.6 milliseconds by Z-DRG. Performing CHS-DRG simulation grouping on a large number of home pages of medical records could obtain the corresponding results quickly. Matching to the results of the regional CHS-DRG operation service platform, the accuracy rate was 99.10%, which greatly enhanced the confidence of medical record coders using Z-DRG for CHS-DRG simulation grouping. According to the results of Z-DRG, clinicians should pay more attention to the correct filling of diagnosis and operation information of the home page of medical records to ensure the accuracy of DRG enrollment^[8]. **Fourth, the grouping results of Z-DRG were standardized.** The Z-DRG extracted the fixed key indicators based on the home page data of medical records and then performed CHS-DRG simulation grouping to generate the results, including the MDC code, ADRG code, DRG code and their names. Z-DRG was used for CHS-DRG simulation grouping. In different periods, different hospitals, different departments, or different clinicians, standard and comparable results were generated by using Z-DRG. **Fifth, the Z-DRG was cost-effective.** The Z-DRG was self-developed based on R language. It did not require a large cost for medical institutions on funds.

Of course, the Z-DRG also had shortcomings. The Z-DRG relied on the R language development environment. Those who did not understand the operating rules of R language required more learning costs in the initial stage. The Z-DRG was applicable to the CHS-DRG 1.0 revision. With the continuous iteration of the CHS-DRG version, the scripts should be updated in time to maintain grouping accuracy. Therefore, our team would continue to accumulate the results of regional CHS-DRG grouping and improve the high-precision simulation grouping of CHS-DRG by introducing machine learning.

4 Conclusion

In summary, we developed a set of CHS-DRG simulation grouping programs based on R language, namely, Z-DRG, which was operated easily, timely, and cost-effective. The CHS-DRG simulation grouping results by Z-DRG were accurate, effective, and standard. It was suitable and valuable for different medical

institutions, departments, or clinicians. Subsequent research would integrate machine learning and visualization functions based on the current scripts to make the Z-DRG more convenient, faster and more intelligent.

Abbreviations

China Healthcare Security Diagnosis Related Groups

CHS-DRG

Diagnosis Related Groups

DRG

Major Diagnostic Categories

MDC

Adjacent Diagnosis Related Groups

ADRG

Complication & Comorbidity

CC

Major Complication & Comorbidity

MCC

Declarations

Ethics approval and consent to participate

The study was performed in accordance with the guidelines of the Declaration of Helsinki and received exemption from ethical approval from the Medical Ethics Committee of Fujian Medical University Affiliated Nanping First Hospital (NO. NPSY2020100006). According to the committee, informed consent of participants was not required because the patient's key information (e.g., name, ID number, admission number, etc.) had been de-identified and anonymized.

Consent for publication

Not applicable

Availability of data and materials

The data that support the findings of this study are available from Fujian Medical University Affiliated Nanping First Hospital, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of Fujian Medical University Affiliated Nanping First Hospital.

Competing interests

The authors declare that they have no conflicting interests.

Funding

This work was supported by the Qihang Fund Project of Fujian Medical University (2019QH1248) and the Health Science and Technology Project of Fujian (2021RKA014).

Authors' contributions

JZ designed the study, created the R scripts (Z-DRG) and was a major contributor in writing the manuscript. YF analyzed and interpreted the inconsistent data on the CHS-DRG simulation grouping and tested the Z-DRG. QH and WP collected the home page of medical records and tested the Z-DRG. All authors read and approved the final manuscript.

Acknowledgements

Not applicable

Authors' information

Not applicable

References

1. Wu SW, Pan Q, Chen T. Research on diagnosis-related group grouping of inpatient medical expenditure in colorectal cancer patients based on a decision tree model. *World J Clin Cases*. 2020 Jun 26;8(12):2484-2493. doi: 10.12998/wjcc.v8.i12.2484. PMID: 32607325; PMCID: PMC7322429.
2. Luo AJ, Chang WF, Xin ZR, Ling H, Li JJ, Dai PP, Deng XT, Zhang L, Li SG. Diagnosis related group grouping study of senile cataract patients based on E-CHAID algorithm. *Int J Ophthalmol*. 2018 Feb 18;11(2):308-313. doi: 10.18240/ijo.2018.02.21. PMID: 29487824; PMCID: PMC5824089.
3. Reutersberg B, Salvermoser M, Trenner M, Geisbüsch S, Zimmermann A, Eckstein HH, Kuehnl A. Hospital Incidence and In-Hospital Mortality of Surgically and Interventionally Treated Aortic Dissections: Secondary Data Analysis of the Nationwide German Diagnosis-Related Group Statistics From 2006 to 2014. *J Am Heart Assoc*. 2019 Apr 16;8(8):e011402. doi: 10.1161/JAHA.118.011402. PMID: 30975011; PMCID: PMC6507201.
4. Jeon MJ, Choo SP, Kwak YH, Kim DW, Kim EH. The effect of diagnosis-related group payment system on the quality of medical care for pelvic organ prolapse in Korean tertiary hospitals. *PLoS One*. 2019 Aug 20;14(8):e0220895. doi: 10.1371/journal.pone.0220895. PMID: 31430319; PMCID: PMC6701833.
5. Jung YW, Pak H, Lee I, Kim EH. The Effect of Diagnosis-Related Group Payment System on Quality of Care in the Field of Obstetrics and Gynecology among Korean Tertiary Hospitals. *Yonsei Med J*. 2018 Jun;59(4):539-545. doi: 10.3349/ymj.2018.59.4.539. PMID: 29749137; PMCID: PMC5949296.

6. Ma Y, Wang W. The impact of diagnosis related group payment on the performance of public hospitals. *Am J Transl Res*. 2021 Jun 15;13(6):6796-6801. PMID: 34306429; PMCID: PMC8290653.
7. Tan JY, Senko C, Hughes B, Lwin Z, Bennett R, Power J, Thomson L. Weighted activity unit effect: evaluating the cost of diagnosis-related group coding. *Intern Med J*. 2020 Apr;50(4):440-444. doi: 10.1111/imj.14373. PMID: 31111636.
8. Hu WY, Yeh CF, Shiao AS, Tu TY. Effects of diagnosis-related group payment on health-care provider behaviors: A consecutive three-period study. *J Chin Med Assoc*. 2015 Nov;78(11):678-85. doi: 10.1016/j.jcma.2015.06.012. Epub 2015 Sep 1. PMID: 26341451.
9. Romeyke T, Noehammer E, Ch Scheuer H, Stummer H. Patient-Centred Multidisciplinary Inpatient Care-Have Diagnosis-Related Groups an Effect on the Doctor-Patient Relationship and Patients' Motivation for Behavioural Change? *Glob J Health Sci*. 2016 Oct 1;8(10):56011. doi: 10.5539/gjhs.v8n10p100. PMID: 27302451.
10. Zou K, Li HY, Zhou D, Liao ZJ. The effects of diagnosis-related groups payment on hospital healthcare in China: a systematic review. *BMC Health Serv Res*. 2020 Feb 12;20(1):112. doi: 10.1186/s12913-020-4957-5. PMID: 32050962; PMCID: PMC7017558.
11. Aiello FA, Roddy SP. Inpatient coding and the diagnosis-related group. *J Vasc Surg*. 2017 Nov;66(5):1621-1623. doi: 10.1016/j.jvs.2017.08.030. PMID: 29061276.

Tables

Table 1 The connotation and parameter description of Z-DRG custom function

no	function(parameter)	connotation	Parameter description
1	dsdata(data)	Extract the data on the home page of the medical records and return the "drgsdata" for CHS-DRG simulation grouping	data: home page information of medical records
2	uni.s(msadrg, osadrg, mdc)	Combining the main operation and other operations to judge ADRG	msadrg: ADRG vector corresponding to main operation osadrg: ADRG vector corresponding to other operation mdc: default MDC
3	uni.d(mdmdc, odmdc)	Combining the main diagnosis and other diagnoses to judge MDCZ	mdmdc: MDC vector corresponding to the main diagnosis odmdc: MDC vector corresponding to the other diagnosis
4	adrg.sj(msadrg, judge)	Advancing into ADRG according to main operation	msadrg: ADRG vector corresponding to main operation judge: ADRG identification table in CHS-DRG codes table
5	adrg.dj(mdadrg, judge)	Admission to ADRG according to the main diagnosis	mdadrg: ADRG vector corresponding to the main diagnosis judge: ADRG identification table in CHS-DRG codes table
6	adrg.ds(mdadrg, msadrg)	Enrolled into ADRG according to main diagnosis and main operation interaction	mdadrg: ADRG vector corresponding to the main diagnosis msadrg: ADRG vector corresponding to main operation
7	adrg.n(mdadrg, msadrg, mdc, judge)	Set the MDC of the main diagnosis, combine function 4, function 5 and function 6 to judge ADRG	mdadrg: ADRG vector corresponding to the main diagnosis msadrg: ADRG vector corresponding to main operation

mdc: default MDC

judge: ADRG
identification table in
CHS-DRG codes table

8	drgs.pro(drgdata, dcodes, scodes, drgcodes)	integrated step 2 to step 7 to receive the CHS- DRG simulation grouping results	<p>drgdata: extract data from “drgsdata” by case number</p> <p>dcodes: diagnostic DRG group table</p> <p>scodes: operation DRG group table</p> <p>drgcodes: CHS-DRG codes table</p>
---	---	--	---

Table 2 CHS-DRG simulation grouping result form

Variable name	Variable meaning	Variable description
no	Unique identifier	The unique identifier was used for matching the results from the regional CHS-DRG operation management system
gender	Patient gender	1=male, 2=female
age	Patient's age	Text 0 when less than 1 year-old
days	Hospitalization days	
diagnose	Main diagnosis code	
d_name	Main diagnosis name	
surgery	Main operation code	
s_name	Main operation name	
outway	Discharge method	
cost	Total cost	
MDC_code	MDC code	
MDC_name	MDC name	
ADRG_code	ADRG code	
ADRG_name	ADRG name	
DRG_code	DRG code	
DRG_name	DRG name	
DRG_class	DRG type	S=surgical operation N=non-surgical operation M=medical diagnosis

Table3 Cause analysis of inconsistent CHS-DRG simulation grouping results

CHS-DRG simulation grouping results	Regional CHS-DRG grouping results			
	MCC	CC	no-CC	total
MCC	0	35	6	41
CC	121	0	6	127
no-CC	51	99	0	150
total	172	134	12	318

Figures

Figure 1

CHS-DRG grouping principal framework diagram

Figure 2

Working principle of Z-DRG