

BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

BMJ Open

Medical Expenditure Clustering and Determinants of the Annual Medical Expenditures of Residents: An Empirical Study from Rural China

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2018-022721
Article Type:	Research
Date Submitted by the Author:	02-Mar-2018
Complete List of Authors:	Zhang, Yan; Huazhong University of Science and Technology Tongji Medical College, School of Medicine and Health Management Lu, Shan; Huazhong University of Science and Technology Tongji Medical College, School of Medicine and Health Management Yadong, Niu; Huazhong University of Science and Technology, School of Health and Medicine Management ZHANG, LIANG; Huazhong University of Science and Technology Tongji Medical College
Keywords:	medical expenditure, high-cost, HEALTH ECONOMICS, HEALTH SERVICES ADMINISTRATION & MANAGEMENT, International health services < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

SCHOLARONE™
Manuscripts

Peer Review Only

1 **Medical Expenditure Clustering and Determinants of the Annual**
2 **Medical Expenditures of Residents: An Empirical Study from Rural**
3 **China**

4
5 Yan Zhang^{1,2}, Shan Lu^{1,2}, Yadong Niu^{1,2}, Liang Zhang^{1,2} §

6
7 ¹ School of Medicine and Health Management, Tongji Medical College, Huazhong
8 University of Science and Technology, Wuhan, Hubei, 430030, China

9 ² Research Centre for Rural Health Service, Key Research Institute of Humanities &
10 Social Sciences of Hubei Provincial Department of Education, Wuhan, Hubei,
11 430030, China

12
13 **Email addresses:**

14 Y.Z. Yan Zhang Ph.D. yanzhang@hust.edu.cn

15 S.L. Shan Lu M.D. shanlu@hust.edu.cn

16 Y-D. N. Yadong Niu M.D. nyadong@126.com

17 L.Z. Liang Zhang M.D. zhangliang@mails.tjmu.edu.cn

18
19 **§Corresponding author:**

20 Liang Zhang, zhangliang@mails.tjmu.edu.cn

21 **Word Count:** 3276

22

Abstract

Objective: To identify the characteristics of high-cost (HC) patients and the determinants of the annual medical expenditures of Chinese rural residents.

Methods: Medical expenditure clustering was performed by using the Lorenz curve and Gini index. T test and χ^2 test were used to identify the characteristics of the respondents and a multi-level regression model was run on MLwiN 2.30 to examine the determinants of their annual medical expenditures.

Design: A cluster sampling study was performed to identify those residents who availed healthcare services and to assign them to high- (top 5%), moderate- (top 30%) and low-cost (others) groups based on their annual medical expenditures.

Setting: The population-based database of the 2014 New Rural Cooperative Medical System was processed to obtain the annual healthcare utilisation of all residents in Macheng, China.

Participants: Those residents who availed healthcare services were recruited for the study and their annual medical expenditures were used as the research object.

Results: The medical expenditure of Macheng City residents in 2014 has a Gini index of 0.81 and the HC patients account for 68.01% of the medical expenditures of all residents. Those residents who are female (51.5%), aged over 60 years (34.48%) or burdened with diseases that are difficult to assess are highly likely to generate high medical costs. The annual medical expenditures of people living in the same village and town tends to be approximated. Age, disease category, inpatient status, healthcare utilisation and utilisation level were eventually identified as the determinants of annual

1
2
3
4 45 medical expenditures.

5
6 46 **Conclusions:** The medical expenditures of rural residents are clustering at a

7
8 47 remarkably high level and HC patients are suffering from high economic burden.

9
10 48 Therefore, policymakers must prioritise guiding HC patients to seek healthcare and

11
12 49 strengthening healthcare quality management to reduce unnecessary healthcare

13
14 50 utilisation.

15
16 51

17
18 52 **Keywords:** medical expenditure; high-cost; clustering; patient flow; rural China

19
20 53

21 54 **Strengths and limitations of this study**

22
23 55 • This study is the first to introduce medical expenditure clustering, which can

24
25 56 supplement the findings of other studies on high-cost (HC) patients.

26
27 57 • Population-level data on the annual medical expenditure of residents are seldom

28
29 58 reported. This study was conducted in a city with 889,160 residents according to a

30
31 59 processed New Rural Cooperative Medical System (NRCMS) database, which

32
33 60 contained the inpatient and outpatient records of these residents.

34
35 61 • The Lorenz curve and Gini index were prescribed for medical expenditure

36
37 62 clustering and a three-level linear regression model was used for the aggregation at

38
39 63 the residential and town levels.

40
41 64 • Age, gender, hospitalisation information, geographic factors and diseases were

42
43 65 drawn into the regression model, but some individual factors were absent.

44
45 66

67 **Background**

68 Studies on the distribution of diseases within a population have defined clustering as
69 the uneven morbidity of diseases in time or space.^[1, 2] Medical expenditure clustering
70 is a technique that reflects the uneven distribution on the medical expenditure of a
71 given population. In recent years, researchers have begun to show great interest in
72 medical expenditure clustering and have focused on High - Cost patients (HC patient),
73 who are known for their high annual medical expenditures.^[3] HC patients are
74 identified the top 5% biggest spenders in terms of annual medical expenditures.^[4] The
75 majority of the previous studies have revealed that the medical expenditures of HC
76 patients exceed those of the entire population by 50%.^[5] For instance, in 2014, 52.3%
77 of medical expense in the US have been consumed by HC.^[3] HC patients and their
78 management have attracted research attention because these patients have a high
79 healthcare utilisation rate, along with inappropriate healthcare utilization. ^[6] Caring
80 for these patients can discourage such inappropriate utilisation, reduce their health
81 expenditures avoiding catastrophic health expenditure, conserve social health
82 insurance fund and promote horizontal equity.

83 Medical expenditure clustering has also become a major concern in developing
84 countries. In rural China, the rapid development of NRCMS significantly promoted
85 the healthcare utilisation rates of residents. For instance, the annual hospitalisation
86 rate in rural China increased from 8.7% in 2008 to 14.9% in 2017.^[7] An empirical
87 analysis of seven counties in China in 2015 revealed that 78.6% of inpatient services
88 were distributed in one-third of all inpatients in the area. In addition, one patient in

1
2
3
4 89 Qianjiang District utilised inpatient services 27 times in a single year. Another study
5
6 90 involving 12,600 families in Jiangsu revealed that the HC families accounted for 44.9%
7
8 91 of the total medical expenditures of the entire population.^[8] Moreover, clustering at
9
10 92 the patient level is much higher than that at the family level. We also suspect that the
11
12 93 degree of medical expenditure clustering in rural China is very high and that HC
13
14 94 patients incur unnecessary medical expenditures because of the fragmented healthcare
15
16 95 delivery system in rural China. Residents in rural China also seek healthcare services
17
18 96 from a three-tier (village–town–county) healthcare delivery system and a higher level
19
20 97 of the institution corresponds to a better service capability and higher medical costs.^[9]
21
22 98 Given that patient choice neither follows a specified order nor has limitations and that
23
24 99 general practitioners or consultants are unavailable in most parts of rural China, the
25
26 100 residents make uninformed decisions when choosing amongst hospitals and various
27
28 101 types of healthcare services, thereby resulting in the inappropriate utilisation of
29
30 102 healthcare services. For instance, some of these residents, especially HC patients, may
31
32 103 be given inpatient services when they actually require outpatient services and may be
33
34 104 unnecessarily admitted to higher-level hospitals, thereby increasing their medical
35
36 105 expenditures.^[10] Yingchun revealed that the inappropriate admission rate in five
37
38 106 counties reached as high as 27.6% in 2014.^[11]
39
40
41
42
43
44
45
46
47 107 To reduce the economic burden of rural residents in China, several policies and
48
49 108 strategies, such as the Tiered Healthcare System and Serious Illness Medical
50
51
52 109 Insurance (2016), have been proposed. However, the feedback on the effectivity of
53
54
55 110 these programmes is mixed and the object of focus of these policies remains

1
2
3
4 111 unsubstantiated. Such lack of evidence may be attributed to the difficulty in
5
6 112 identifying high-demand and HC patients.^[12] Waxmonsky J A described HC patients
7
8 113 with multiple or complex conditions, often combined with behavioural health
9
10 114 problems or socioeconomic challenges, they are difficult to monitor.^[13] Moreover,
11
12 115 only few studies have investigated the characteristics and determinants of HC patients
13
14 116 based on population-level data because of the lack of necessary data for exploring
15
16 117 annual medical expenditures. So the critical values, average expenditure and inpatient
17
18 118 service utilisation of HC patients are all unclear in rural China.
19
20 119 Previous studies reveal that HC patients often maintain a high level of medical
21
22 120 expenditure for the following year. For instance, Robst pointed out that 49.2% of HC
23
24 121 patients continuously were classified HC patient, under Florida Medical Assistance
25
26 122 Program (Florida Medicaid) from 2005 to 2010.^[14] Meanwhile, Wodchis et al. found
27
28 123 the ratio is 1/3 based 2009 to 2011 of public health insurance in Ontario, Canada.^[15]
29
30 124 Therefore, HC patients are assumed to possess certain characteristics even though
31
32 125 their distribution is unclear.
33
34
35
36
37
38
39
40
41 126 Based on the above findings, identifying HC patients is a necessary procedure.
42
43 127 Clarifying the medical expenditure clustering of a given population must be taken as
44
45 128 the first step towards such goal. This research focuses on medical expenditure
46
47 129 clustering in particular and distinguishes the distribution and characteristics of HC
48
49 130 patients and the determinants of the annual medical expenditures of residents in
50
51 131 general. This study also aims to guide policymakers and health planners in predicting
52
53 132 and planning the future needs of these patients.
54
55
56
57
58
59
60

133 **Methods**

134 **Study setting**

135 We calculated the annual medical expenditure of the residents based on the outpatient
136 and inpatient services that they have availed within a calendar year. A
137 population-based retrospective cohort was performed in Macheng, a typical rural area
138 in Hubei and a county-level city in central China (Fig. 1). Macheng has a total
139 population of 889,160 and a GDP per capita of 22,758¥.

140 **[Figure 1]**

141 Macheng has 2 county hospitals, 22 township hospitals and 207 village clinics. Rural
142 residents are enrolled in the NRCMS, which offers reimbursement for all types of
143 healthcare utilisation (e.g. outpatient and inpatient services in various medical
144 institutions, including the tertiary hospital in urban area). The healthcare utilisation of
145 these residents is recorded in the NRCMS database whenever they request for a
146 reimbursement. This study uses the 2014 NRCMS database for Macheng.

147 **Data processing**

148 This retrospective cohort study identified those residents who availed healthcare
149 services. After screening, 478,051 of the 889,160 residents in the NRCMS database
150 were included in the sample. The samples were processed by using MS Excel 2010.
151 Firstly, the outpatient and inpatient cases were merged into one Excel sheet. Secondly,
152 those cases under the same patient identifier (ID number) were sorted in chronological
153 order. Thirdly, the annual medical and inpatient and outpatient expenditures of each
154 patient, the number of outpatient and inpatient cases and other information were

1
2
3
4 155 recorded. Fourthly, the annual healthcare utilisation cases were inputted into a new
5
6 156 database, where each case represents the annual healthcare utilisation of a resident.
7
8 157 Finally, the residents were sorted in a descending order according to their annual
9
10 158 medical expenditures. Those residents that occupied the top 5% of the sample were
11
12 159 included in the HC group, those in the top 6% to 30% were included in the moderate
13
14 160 cost (MC) group and the other patients were included in the low cost (LC) group.
15
16 161 These three groups represent the various degrees of medical expenditure clustering.
17
18 162 The main programming techniques include Excel formulas (e.g. COUNTIF,
19
20 163 SUMPRODUCT, LOOKUP and IF) and case processing technologies (e.g. split
21
22 164 columns and removal of duplicates). We marked the outstanding diseases of each
23
24 165 resident and adjusted the original ICD-10 disease code to a broader one (for example,
25
26 166 the disease code for COPD was adjusted from J44.900 to J44). Township hospitals
27
28 167 were divided into four levels according to their scale and service capacity. The
29
30 168 distance from and arrival time to county hospitals were individually captured by using
31
32 169 Google Maps.
33
34 170 The land form of towns, the healthcare capacity of township hospitals and the
35
36 171 sociological characteristics (e.g. gender and age), arrival time to the county hospital,
37
38 172 disease category, healthcare utilisation (e.g. annual length of stay (LOS)) and annual
39
40 173 medical expenditures of the patients were collected to build the final database.^[16]
41
42 174 The study protocol conformed to the guidelines of the Ethics Committee of the Tongji
43
44 175 Medical College of Huazhong University of Science and Technology. The patient
45
46 176 information was anonymised and deidentified before the analysis.
47
48
49
50
51
52
53
54
55
56
57
58
59
60

177 **Statistical analysis**

178 Medical expenditure clustering was described by using the Gini coefficient and
 179 Lorenz curvet. The characteristics of the residents in the three groups were compared
 180 by conducting T and χ^2 tests in IBM SPSS Statistics 22.0. Given that the obtained data
 181 showed a hierarchical structure, the determinants of annual medical expenditure were
 182 examined by conducting a multilevel linear logistic regression analysis in MLwiN
 183 2.30.^[17] Annual medical expenditure showed a skewed distribution as expected.
 184 Therefore, we transformed annual medical expenditure to follow the normal
 185 distribution. We accepted the Log10 (x) translation and obtained the following
 186 regression model:

$$\begin{aligned}
 \text{Log10(Annual medical expense)} = & \beta_{0jk}\text{cons} + \beta_1\text{Gender}_{ijk} + \beta_2\text{Age}_{ijk} + \beta_3\text{Family}_{ijk} + \beta_4\text{Distance}_{ijk} + \beta_5\text{Time}_{ijk} + \beta_6\text{Capacity}_{ijk} \\
 & + \beta_7\text{Disease}_{ijk} + \beta_8\text{LOS}_{ijk} + \beta_9\text{Coutinp}_{ijk} + \beta_{10}\text{Outinp}_{ijk} + \beta_{11}\text{Coutinp}_{ijk} + \beta_{12}\text{Towinp}_{ijk} \\
 & + \beta_{13}\text{Cououtp}_{ijk} + \beta_{14}\text{Towoutp}_{ijk} + \beta_{15}\text{Clioutp}_{ijk} + \text{More}_{ijk} \\
 \beta_{0jk} = & \beta_0 + u_{0j} + w_{0jk}
 \end{aligned}$$

188 where β_{0jk} denotes the fixed-effects parameter, u_{0j} denotes the random effects at the
 189 village level and w_{0jk} denotes the random effects at the town level.

190 **Results**

191 **Medical expenditure clustering in sample area**

192 Table 1 presents the medical expenditure clustering results of Macheng in 2014. The
 193 top 5% of the population accounted for 68.01% of the total medical expenditure
 194 whilst the top 20% accounted for over 90%. Figure 2 shows the Lorenz curve of
 195 medical expenditure clustering with a Gini coefficient of 0.814.

196 **[Figure 2]**

197

198 **Table 1. Medical expenditure clustering of Macheng in 2014**

Population ranking (%)	Critical value (¥)	Total medical expenditure (10T ¥)	Cumulative percentage of medical expenditure (10T ¥)	Cumulative percentage of medical expenditure (%)
1	19,123.63	22,634.57	22,634.57	38.73
1-5	4,985.8	17,111.41	39,745.99	68.01
5-10	2,021.45	7,696.77	47,442.76	81.18
10-20	610.22	5,179.36	52,622.12	90.04
20-30	347.29	2,205.70	54,827.82	93.82
30-40	221.26	1,329.19	56,157.01	96.09
40-50	147.37	867.84	57,024.85	97.58
50-60	99.54	582.81	57,607.67	98.57
60-70	63.35	384.42	57,992.09	99.23
70-80	39.95	244.01	58,236.10	99.65
80-90	20.19	141.11	58,377.20	99.89
90-100	3.20	64.17	58,441.37	100.00

199 Table 2 shows the expenditure distribution for the HC, MC and LC groups. The HC
 200 group has an average annual expenditure per capita of over 15,000¥ and a minimum
 201 expenditure of 4,985.80¥ whilst the LC group has a maximum expenditure of 347.29¥.
 202 Figure 3 presents the expenditure composition of these groups.

203 **Table 2. Expenditure distribution for the HC, MC, and LC groups (¥)**

Variable	All	Cost Groups		
		HC	MC	LC
Annual expenditure per capita M(SD)	1,222.49 (6,253.01)	16,618.78 (22,817.95)	1,261.36 (1,112.90)	107.99 (89.26)
Minimum expenditure	424,962.10	424,962.10	4,985.60	347.29
Maximum expenditure	3.20	4,985.80	347.30	3.20
Inpatient expenditure per capita M(SD)	976.09 (6,165.55)	15,866.07 (22,745.43)	728.3 (1,182.59)	0.15 (6.39)
Outpatient expenditure per capita M(SD)	246.4 (559.70)	752.71 (2,083.85)	533.06 (417.68)	107.84 (89.17)

204 **[Figure 3]**

205 **Characteristics of HC patients at various clustering levels**

206 Table 3 shows the demographic characteristics of the patients. Significant differences

207 were observed amongst the three groups for nine demographic items (all $P < 0.001$).

208 Specifically, females accounted for 51.5% and 47.42% of the residents in the HC and

209 LC groups, respectively. Average age also varied across these groups. Specifically,

210 those residents aged over 60 years mostly belonged to the HC group. The residents in

211 the HC and LC groups generally had small (4.01) and large (4.23) family sizes,

212 respectively. Distance from and arrival time to county hospital showed the same

213 change trends in terms of family size. Those residents living nearby a high-capacity

214 township hospital mostly belonged to the HC group (15.19%), by contrast, the

215 residents living nearby a low-capacity township hospital accounted for 30.76% of all

216 members of the LC group. These three groups showed similar distributions across

217 varying geographic and traffic conditions. Specifically, in the LC group, those

218 residents living in poor areas (mountains and county roads) had high accident rates

219 (12.54% and 32.36%, respectively). In addition, HC patients were highly likely to

220 develop ailments, such as cancer, circulatory, digestive and urinary diseases as well as

221 haematological disorders, whilst LC patients frequently developed respiratory

222 diseases.

223 **Table 3. Distribution of the demographic characteristics of residents in the three cost groups**
 224 **(n = 478,051)**

Variable	All n (%)	Cost Groups n (%)			T/F Value
		HC	MC	LC	
All	478,051	23,922 (5.00)	119,492 (25.00)	334,637 (70.00)	
Gender					
Male	242,244 (50.67)	11,602 (48.5)	54,704 (45.78)	175,938 (52.58)	1,673.95*
Female	235,804 (49.33)	12,320 (51.5)	64,786 (54.22)	158,698 (47.42)	
Age (years)					
Mean (SD)	41.97 (20.11)	49.45 (18.47)	45.56 (20.14)	40.14 (19.92)	4,739.48**
Less than 20	77,706 (16.4)	1,374 (5.75)	15,218 (12.79)	61,114 (18.47)	10,792.85*

20–39	114,619 (24.19)	5,142 (21.51)	23,166 (19.47)	86,311 (26.08)	
40–59	181,426 (38.29)	9,147 (38.27)	48,561 (40.81)	123,718 (37.39)	
60–79	93,338 (19.7)	7,753 (32.44)	29,926 (25.15)	55,659 (16.82)	
More than 79	6,706 (1.42)	487 (2.04)	2,118 (1.78)	4,101 (1.24)	
Family size					
Mean (SD)	4.2 (1.54)	4.01 (1.62)	4.15 (1.59)	4.23 (1.51)	350.63**
Distance from CH (km)					
Mean (SD)	30.7 (15.15)	29.34 (15.48)	30.16 (15.44)	31 (15.01)	234.87**
Arrival time to CH (min)					
Mean (SD)	46.5 (18.03)	44.88 (18.6)	45.95 (18.64)	46.81 (17.76)	200.56**
Capacity of TH					
First level (strong)	56,866 (11.9)	3,633 (15.19)	15,868 (13.28)	37,365 (11.17)	690.05*
Second level (general)	276,701 (57.88)	13,304 (55.61)	69,057 (57.79)	194,340 (58.07)	
Third level (weak)	144,484 (30.22)	6,985 (29.2)	34,567 (28.93)	102,932 (30.76)	
Geography					
Plain	175,767 (36.77)	9,753 (40.77)	47,877 (40.07)	118,137 (35.3)	2,701.41*
Hilly	117,197 (24.52)	5,792 (24.21)	31,915 (26.71)	79,490 (23.75)	
Hilly and mountainous	126,341 (26.43)	5,641 (23.58)	25,665 (21.48)	95,035 (28.4)	
Mountainous	58,746 (12.29)	2,736 (11.44)	14,035 (11.75)	41,975 (12.54)	
Traffic condition					
National standard roads	214,571 (44.88)	10,499 (43.89)	53,544 (44.81)	150,528 (44.98)	885.97*
Provincial standard roads	113,392 (23.72)	6,561 (27.43)	31,008 (25.95)	75,823 (22.66)	
County roads	150,088 (31.4)	6,862 (28.68)	34,940 (29.24)	108,286 (32.36)	
Disease category					
Cancer	4,665 (0.98)	2,686 (11.23)	1,452 (1.22)	527 (0.16)	87,301.73*
ENT disease	41,772 (8.74)	2,475 (10.35)	7,114 (5.95)	32,183 (9.62)	
Respiratory disease	181,929 (38.06)	3,312 (13.84)	30,142 (25.23)	148,475 (44.37)	
Circulatory disease	24,721 (5.17)	3,781 (15.81)	9,701 (8.12)	11,239 (3.36)	
Digestive disease	45,723 (9.56)	2,838 (11.86)	9,539 (7.98)	33,346 (9.96)	
Urinary disease	8,048 (1.68)	1,360 (5.69)	3,364 (2.82)	3,324 (0.99)	
Endocrinology disease	5,785 (1.21)	133 (0.56)	792 (0.66)	4,860 (1.45)	
Haematological disorders	2,995 (0.63)	427 (1.78)	739 (0.62)	1,829 (0.55)	
Bones and muscles	41,675 (8.72)	2,988 (12.49)	11,993 (10.04)	26,694 (7.98)	
Obstetrics and gynaecology	15,844 (3.31)	2,223 (9.29)	6226 (5.21)	7,395 (2.21)	
Others	104,894 (21.95)	1,699 (7.1)	38,430 (32.16)	64,765 (19.36)	

225 *Pearson's chi-square test.

226 **ANOVA.

227 Table 4 shows the distribution of the healthcare utilisation of all residents in 2014.

228 The average number of inpatient cases demonstrated a decreasing trend for the HC,

229 MC and LC groups (2.36, 0.45 and 0.02, respectively). Similar results were obtained

230 for the number of inpatient cases in 3 levels and annual LOS (25.69, 3.28 and 0.1),
 231 but the opposite results were observed for those residents without prior hospitalisation
 232 experience (0.07, 62.94 and 99.95). The average number of outpatient cases (8.28,
 233 13.57, 5.05), and that at the town and clinic levels showed an inverted V-shaped
 234 distribution in 3 groups. Specifically, the residents in the MC groups showed a high
 235 mean number of outpatient cases at the town and clinic levels (3.69 and 8.61). By
 236 contrast, the number of outpatient cases at the county level showed a decreasing trend
 237 in all three groups (1.76, 1.27 and 0.21).

238 **Table 4. Distribution of the healthcare utilisation characteristics of the three groups (n =**
 239 **478,051)**

Variable	All n (%)	Cost Groups			F
		HC	MC	LC	
Outpatient number					
Mean (SD)	7.34(8.26)	8.28(9.1)	13.57(11.63)	5.05(4.89)	58,639.94**
Outpatient level (time)					
County	0.55 (1.51)	1.76 (2.85)	1.27 (2.26)	0.21 (0.66)	34,430.07**
Township	1.96 (4.14)	2.16 (4.5)	3.69 (6.48)	1.32 (2.57)	15,398.27**
Clinic	4.83 (7.71)	4.36 (7.72)	8.61 (11.78)	3.52 (4.94)	20,854.66**
Inpatient number					
Mean (SD)	0.23 (1.21)	2.36 (4.33)	0.45 (1)	0 (0.02)	55,409.19**
0	409,682 (85.70)	17 (0.07)	75,207 (62.94)	334,458 (99.95)	274,894.92
1–3	64,955 (13.59)	20,840 (87.12)	43,936 (36.77)	179 (0.05)	*
4–7	2,736 (0.57)	2,447 (10.23)	289 (0.24)	0 (0)	
More than 8	678 (0.14)	618 (2.58)	60 (0.05)	0 (0)	
Inpatient level (time)					
Outside the county	0.04 (0.38)	0.58 (1.53)	0.04 (0.22)	0.01 (0.01)	28,920.75**
County	0.14 (1.07)	1.49 (4.09)	0.25 (0.90)	0.01 (0.01)	24,788.76**
Township	0.06 (0.32)	0.29 (0.96)	0.16 (0.45)	0.01 (0.02)	18,940.21**
Annual LOS					
Mean (SD)	2.09 (10.66)	25.46 (36.69)	3.28 (7.91)	0.01 (0.1)	88,485.86**

240 *Pearson's chi-square test.

241 **ANOVA.

242 **Determinants of the annual medical expenditure of residents**

243 A multilevel linear regression was performed and the patients, villages and towns

244 were identified as levels 1, 2 and 3, respectively. Table 5 displays the results of the
 245 explanatory variables to fit the three variance component models. Age, disease
 246 category, and all terms for healthcare utilisation were identified as the major
 247 determinants of the annual medical expenditures of residents. No significant
 248 relationship was observed amongst gender, family size, distance to county hospital,
 249 arrival time to county hospital, home geography, traffic condition and township
 250 hospital capacity. If the other factors were kept constant, then the medical
 251 expenditures of patients would increase along with age, number of outpatients, LOS
 252 and level of utilisation. Those patients with cancer, circulatory, digestive, urinary,
 253 obstetric and gynaecological diseases or haematological disorders had a higher
 254 probability to incur high annual medical expenditures than those with respiratory,
 255 ENT, endocrinal, skeletal and muscular diseases.

256 **Table 5 Three-level linear regression model analysis of annual medical expenditures (n =**
 257 **478,051)**

	Parameter estimate	Standard error	χ^2	P
Fixed factors				
Constant	1.842	0.063	862.479	<0.001
Gender (baseline: male)				
Female	0.003	0.002	5.051	0.025
Age (years)	0.003	0.001	4690.838	<0.001
Family size (person)	0.001	0.001	1.536	0.215
Distance to CH (km)	-0.001	0.001	0.276	0.599
Time to CH (min)	0.001	0.001	0.006	0.938
Capacity of TH	-0.073	0.030	5.797	0.016
Geography (baseline: plains)	0.013	0.020	0.423	0.515
Traffic condition (baseline: national)	0.004	0.022	0.036	0.849
Disease category (baseline: respiratory)				
Cancer	0.759	0.008	8286.833	<0.001
ENT disease	-0.095	0.003	1084.129	<0.001
Circulatory disease	0.303	0.004	4948.426	<0.001
Digestive disease	0.024	0.004	11.12	0.001
Urinary disease	0.304	0.006	2197.804	<0.001

Endocrinology disease	-0.070	0.007	6.756	0.009
Haematological disorders	0.113	0.010	127.488	<0.001
Bones and muscles	0.030	0.008	10.168	0.001
Obstetrics and gynaecology	0.390	0.005	6056.819	<0.001
Else	0.126	0.003	0.294	0.588
Annual LOS (days)	0.015	0.001	18287.060	<0.001
Inpatient level (times)				
Outside the county	0.275	0.002	15149.683	<0.001
County	0.064	0.001	5024.366	<0.001
Township	0.260	0.003	8241.838	<0.001
Outpatient level (times)				
County	0.134	0.001	62795.795	<0.001
Township	0.049	0.001	56609.642	<0.001
Clinic	0.031	0.001	82166.996	<0.001
Random factors				
Town variance	0.005	0.002	9.114	0.003
Village variance	0.007	0.000	298.486	<0.001
Patient scale parameter	1	0.00	-	-

258 Discussion

259 Clustering of the medical expenditures of residents in rural China

260 The medical expenditures of rural residents showed an extremely uneven distribution.

261 With a Gini coefficient of 0.814 in 2014, medical expenditure was clustered in the

262 minority population. The HC group accounted for 68.01% of the total medical

263 expenditures; this proportion is much higher than that observed in the US, where the

264 HC group accounted for 52.3% of the total medical expenditures in the same year.^[3]

265 Meanwhile, the LC group, 70% of population, accounted for 2.42% of the total

266 medical expenditures. The annual medical expenditure per capita for the entire

267 population was 1,222.49 ¥, which is nearly similar to that for the MC group (1261.36¥)

268 and 7.35% for the HC group. The maximum annual medical expenditure of this

269 population was 424,962.1¥, which is more than four times larger than the annual

270 reimbursement amount provided by NRCMS (100,000¥).

1
2
3
4 271 The three groups showed obvious differences in their expenditure structures.
5
6 272 Hospitalisation expenditures accounted for over 95% of the medical expenditures of
7
8 273 the HC group, thereby supporting the findings of Driessen et al.^[18] Meanwhile,
9
10
11 274 outpatient expenditures accounted for most of the medical expenditures of the LC
12
13 275 group, medical expenditure of the MC group is made up of outpatient and
14
15 276 hospitalization expenditures coequally. The medical expenditures of the HC group
16
17 277 showed a dispersed distribution with a standard deviation that was much higher than
18
19 278 the mean (22,817.95 vs. 16,618.78). By contrast, the medical expenditures of the MC
20
21 279 and LC groups showed a relatively concentrated distribution.
22
23
24
25

26 280 The HC group faces a very high economic burden such that those residents with over
27
28 281 5,000¥ in medical costs are identified as members of this group. The annual medical
29
30 282 expenditure per capita of the HC group is over 16,000¥ whereas the actual
31
32 283 reimbursement ratio of NRCMS is approximately 50%. Therefore, the average Out of
33
34 284 Pocket (OOP) of HC patients is over 8,000¥, which is nearly 80% of the total
35
36 285 consumer spending per capita of rural residents (10,129.8¥).^[7] This ratio can lead to a
37
38 286 catastrophic health spending easily.
39
40
41
42

43 287 **Aggregating annual medical expenditures at the village and town level**

44
45
46 288 The multilevel linear regression revealed that the resident data followed a hierarchical
47
48 289 structure (town–village-residents). The annual medical expenditures of residents were
49
50 290 clustered at the town and village levels, the annual medical expenditures for the same
51
52 291 village and town tend to be approximated. The distribution of annual medical
53
54
55
56
57
58
59
60

1
2
3
4 292 expenditures in Table 3 significantly differs from that in Table 5 in terms of home
5
6 293 geography, traffic condition and capacity of township hospitals. Such differences were
7
8 294 mostly observed at the town level because the residents were living in a town with the
9
10 295 same home geography, traffic condition and township hospital. In addition, 30 towns
11
12 296 in the sample showed differences in their social customs, geographic location and
13
14 297 capacity of their township hospitals whereas the other township hospitals in the
15
16 298 sample had different operating capacities. Taken together, those residents living in
17
18 299 areas with favourable geographic conditions (e.g. plains and national standard roads)
19
20 300 and hospitals with high capacity tend to accumulate high medical expenditures.
21
22 301 Similarly, the effects of distance and arrival time to county hospital were mostly
23
24 302 observed at the village level, shorter distance from and arrival time of these residents
25
26 303 to the county hospital, higher probability of availing county healthcare services, and
27
28 304 higher medical expenditures. The decision of these residents to pursue the most
29
30 305 convenient option can be easily understood.

306 **Determinants of annual medical expenditures**

307 Annual medical expenditures was directly determined by healthcare utilisation, that is,
308 the utilisation of more healthcare services corresponds to a higher medical
309 expenditure as can be seen in Table 5. At the outpatient level, the regression
310 coefficients of village–town–county were 0.031, 0.049 and 0.134. The regression
311 coefficients for inpatients were also higher than those for outpatients. Similar to
312 hospital utilisation, an increase in LOS corresponds to an increase in expenditure.
313 In addition, the higher the age, the higher the expenditure. This case is true especially

1
2
3
4 314 for those residents aged above 60 years (comprising 8.5% of the HC group) because
5
6 315 of their poor physical condition and high tendency to develop one or multiple diseases.
7
8 316 A WHO report revealed that an aging population would increase healthcare spending
9
10 317 but the extent of such increase is less than expected. Although health-related needs
11
12 318 often increase with age, the relationship between healthcare utilisation and health
13
14 319 expenditure is variable.^[19] In fact, the medical expenditures of people from several
15
16 320 high-income countries gradually declined after reaching the age of 75. Other studies
17
18 321 even show that those people aged 80 years and above are major contributors of
19
20 322 medical resources, that is, their share in the consumption of medical resources is
21
22 323 much lower than its share of the population. In terms of gender, the average medical
23
24 324 expenditures of female residents do not differ from those of male residents, there is no
25
26 325 conclusive research on the relationship between gender and medical expenditures.^[20]
27
28 326 Respiratory, urinary, endocrinology, skeletal and muscular diseases have the same
29
30 327 degree of medical expenditure clustering. By contrast, other diseases, such as cancer,
31
32 328 circulatory, digestive, obstetrical and gynaecological diseases or haematological
33
34 329 disorders, have a significantly high annual medical expenditures, with cancer having
35
36 330 the highest regression coefficient of 0.759. Those diseases that are difficult to assess
37
38 331 and can easily lead to death greatly contribute to high medical expenditures.^[21] This
39
40 332 finding can be attributed to the ability of healthcare professionals at the township
41
42 333 level, most of which do not specialise in urinary and cardiovascular diseases or
43
44 334 haematological disorders; therefore, those residents suffering from such diseases tend
45
46 335 to be admitted to county hospitals or hospitals outside their respective counties.^[22]
47
48
49
50
51
52
53
54
55
56
57

336 **Redesigning the health delivery system for the HC group**

337 The intense concentration of medical expenditures reveals an imbalance in healthcare
338 utilisation. Thus, the rural healthcare delivery system must specifically focus on the
339 members of the HC group. Firstly, establishing a monitoring mechanism for HC
340 patients is a primary and key measure whilst the NRCMS database can serve as a
341 source of information for monitoring purposes. Those residents with a notably high
342 healthcare utilisation and those who have been identified as HC patients in the
343 previous year warrant special attention. Moreover, HC patients must be guided in
344 seeking for healthcare services and given priority access to primary healthcare.
345 Secondly, strengthening healthcare quality management plays a vital role in
346 discouraging unnecessary healthcare utilisation cases, such as inappropriate admission
347 and excessive utilisation of inpatient services.^[23] Thirdly, exploring new mechanisms,
348 such as the comprehensive management of patients with chronic diseases, global
349 budget for multilevel institutions, integrated management programmes and integrated
350 delivery of medicine and nursing services to aged residents, can motivate doctors to
351 deliver continued care to HC patients. Continued care can prevent unnecessary service
352 duplications, improve healthcare efficacy and help patients take advantage of primary
353 healthcare services.^[24]

354 **Conclusion**

355 From the city-level population perspective, the rural residents in China demonstrate
356 an intense medical expenditure clustering. Apart from demographic characteristics
357 (e.g. age and disease), healthcare utilisation was identified as a primary determinant

1
2
3 358 of medical expenditure clustering. Policymakers must consider patient prioritisation to
4
5
6 359 guide HC patients in seeking for healthcare services. They must also improve
7
8
9 360 healthcare quality management to discourage the unnecessary utilisation of healthcare
10
11 361 services. Doctors must also be given motivation to deliver continued care to their
12
13 362 patients.

363 **Limitation**

14
15
16
17
18 364 This study has two limitations. Firstly, hospitalisation information, geographic factors,
19
20
21 365 referral status and diseases were included in the regression model whilst other
22
23
24 366 individual factors, such as economic status, education and preference, were ignored.
25
26 367 Secondly, several studies point out that HC residents with a high tendency to develop
27
28
29 368 multiple chronic diseases tend to utilise multidisciplinary and multi-institutional
30
31 369 services.^[25] However, we only considered main disease, the diagnoses from the
32
33
34 370 NRCMS database. These limitations may affect the stability of our findings and
35
36 371 should be addressed in further studies.

372 **Competing interests**

37 373 The authors have declared that no competing interests exist.

374 **Authors' contributions**

38
39
40
41 375 Y.Z. and S.L. participated in conception and design, and the analyses, and wrote the
42
43
44 376 manuscript. Y-D.N. participated in data collection and performed the statistical
45
46
47 377 analysis. L.Z. helped to draft the manuscript, reviewed the manuscript and made final
48
49
50
51 378 changes. All authors have given their final approval of the version to be published.
52
53
54
55
56
57

379 **Funding statement**

380 This research is supported by the National youth Natural Science Foundation of China,
381 grant number (71603088).

382 **Disclosure**

383 No past publication history, no past presentation history

384 **Provenance and peer review**

385 Not commissioned; externally peer reviewed.

386 **Data sharing statement**

387 The anonymized dataset is available through the email of the corresponding author.

388

389 **References:**

- 390 [1] Murakami Y, Okamura T, Nakamura K, Miura K, Ueshima H. The clustering of cardiovascular
391 disease risk factors and their impacts on annual medical expenditure in Japan: community-based cost
392 analysis using Gamma regression models. *BMJ open*. 2013;**3**(3) 'doi': 10.1136/bmjopen-2012-002234['].
- 393 [2] Leech RM, McNaughton SA, Timperio A. The clustering of diet, physical activity and sedentary
394 behavior in children and adolescents: a review. *INTERNATIONAL JOURNAL OF BEHAVIORAL*
395 *NUTRITION AND PHYSICAL ACTIVITY*. 2014;**11** 'doi': 10.1186/1479-5868-11-4['].
- 396 [3] Zulman DM, Chee CP, Wagner TH, Yoon J, et al. Multimorbidity and healthcare utilisation
397 among high-cost patients in the US Veterans Affairs Health Care System. *BMJ OPEN*. 2015;**5**(4) 'doi':
398 10.1136/bmjopen-2015-007771['].
- 399 [4] Coughlin TA, Long SK. Health Care Spending and Service Use among High-Cost Medicaid
400 Beneficiaries, 2002-2004. *INQUIRY-THE JOURNAL OF HEALTH CARE ORGANIZATION*
401 *PROVISION AND FINANCING*. 2009;**46**(4):405-417 'doi': 10.5034/inquiryjrn1_46.4.405['].
- 402 [5] Zulman DM, Ezeji-Okoye SC, Shaw JG, Hummel DL, et al. Partnered Research in Healthcare
403 Delivery Redesign for High-Need, High-Cost Patients: Development and Feasibility of an Intensive
404 Management Patient-Aligned Care Team (ImPACT). *JOURNAL OF GENERAL INTERNAL*
405 *MEDICINE*. 2014;**294**:S861-S869 'doi': 10.1007/s11606-014-3022-7['].
- 406 [6] Hong CS, Siegel AL, Ferris TG. Caring for high-need, high-cost patients: what makes for a
407 successful care management program? Issue brief (Commonwealth Fund). 2014;**19**:1-19.
- 408 [7] NHaFPCoC. *2017 China Statistics Yearbook of Health and Family Planning*. Beijing: China
409 union medical university press, 2017.
- 410 [8] Miao Y, Qian D, Sandeep S, Ye T, et al. Exploring the characteristics of the high-cost population
411 from the family perspective: a cross-sectional study in Jiangsu Province, China. *BMJ open*.

- 1
2
3 412 2017;7(11):e17185 'doi': 10.1136/bmjopen-2017-017185['].
- 4 413 [9] Li X, Lu JP, Hu S, Cheng KK, et al. The primary health-care system in China. LANCET.
5 414 2017;**390**(10112):2584-2594ELSEVIER SCIENCE INC.
- 6 415 [10] Yip W, Hsiao W. Harnessing the privatisation of China's fragmented health-care delivery.
7 416 LANCET. 2014;**384**(9945):805-818 'doi': 10.1016/S0140-6736(14)61120-X['].
- 8 417 [11] Zhang Y, Chen Y, Zhang X, Zhang L. Current level and determinants of inappropriate admissions
9 418 to township hospitals under the new rural cooperative medical system in China: a cross-sectional study.
10 419 BMC HEALTH SERVICES RESEARCH. 2014;**14**(649).
- 11 420 [12] Bleich SN. Systematic Review of Programs Treating High-Need and High-Cost People With
12 421 Multiple Chronic Diseases or Disabilities in the United States, 2008-2014 (vol 12, E197, 2015).
13 422 PREVENTING CHRONIC DISEASE. 2016;**13** 'doi': 10.5888/pcd13.150275e['].
- 14 423 [13] Waxmonsky JA, Giese AA, McGinnis GF, Reynolds RT, et al. Colorado access' enhanced care
15 424 management for high-cost, high-need Medicaid members: preliminary outcomes and lessons learned.
16 425 The Journal of ambulatory care management. 2011;**34**(2):183-191 'doi':
17 426 10.1097/JAC.0b013e31820f64be['].
- 18 427 [14] Robst J. Developing Models to Predict Persistent High-Cost Cases in Florida Medicaid.
19 428 POPULATION HEALTH MANAGEMENT. 2015;**18**(6):467-476 'doi': 10.1089/pop.2014.0174['].
- 20 429 [15] Wodchis WP, Austin PC, Henry DA. A 3-year study of high-cost users of health care.
21 430 CANADIAN MEDICAL ASSOCIATION JOURNAL. 2016;**188**(3):182-188 'doi':
22 431 10.1503/cmaj.150064['].
- 23 432 [16] Herrin J, St Andre J, Kenward K, Joshi MS, Audet AJ, Hines SC. Community Factors and
24 433 Hospital Readmission Rates. HEALTH SERVICES RESEARCH. 2015;**50**(1):20-39.
- 25 434 [17] Longford NT. Improved approximations for multilevel models with binary responses. JOURNAL
26 435 OF THE ROYAL STATISTICAL SOCIETY SERIES A-STATISTICS IN SOCIETY.
27 436 1997;**160**(3):593.
- 28 437 [18] Driessen RJB, Bisschops LA, Adang EMM, Evers AW, van de Kerkhof PCM, de Jong EMGJ.
29 438 The economic impact of high-need psoriasis in daily clinical practice before and after the introduction
30 439 of biologics. BRITISH JOURNAL OF DERMATOLOGY. 2010;**162**(6):1324-1329 'doi':
31 440 10.1111/j.1365-2133.2010.09693.x['].
- 32 441 [19] Beard JR, Officer AM, Cassels AK. The World Report on Ageing and Health.
33 442 GERONTOLOGIST. 2016;**56**2:S163-S166 'doi': 10.1093/geront/gnw037['].
- 34 443 [20] Owens GM. Gender differences in health care expenditures, resource utilization, and quality of
35 444 care. JOURNAL OF MANAGED CARE PHARMACY. 2008;**143**(3):S2-S6.
- 36 445 [21] Ross JS, Chen J, Lin Z, Bueno H, et al. Recent National Trends in Readmission Rates After Heart
37 446 Failure Hospitalization. CIRCULATION-HEART FAILURE. 2010;**3**(1):97-103.
- 38 447 [22] Hartl S, Luis Lopez-Campos J, Pozo-Rodriguez F, Castro-Acosta A, et al. Risk of death and
39 448 readmission of hospital-admitted COPD exacerbations: European COPD Audit. EUROPEAN
40 449 RESPIRATORY JOURNAL. 2016;**47**(1):113-121.
- 41 450 [23] Feder JL. Predictive Modeling And Team Care For High-Need Patients At HealthCare Partners.
42 451 HEALTH AFFAIRS. 2011;**30**(3):416-418 'doi': 10.1377/hlthaff.2011.0080['].
- 43 452 [24] Ye T, Sun XW, Tang WX, Miao Y, Zhang YD, Zhang L. Effect of continuity of care on
44 453 health-related quality of life in adult patients with hypertension: a cohort study in China. BMC
45 454 HEALTH SERVICES RESEARCH. 2016;**16**(674)BIOMED CENTRAL LTD.
- 46 455 [25] Zulman D, Yoon J, Cohen DM, Wagner TH, Ritchie C, Asch S. MULTIMORBIDITY AND

1
2
3 456 HEALTH CARE UTILIZATION AMONG HIGH-COST PATIENTS: IMPLICATIONS FOR CARE
4 457 COORDINATION. JOURNAL OF GENERAL INTERNAL MEDICINE. 2013;281:S123-S124.
5

6 458 **Figure Legend**
7

8 459 **Figure 1.** Map of Macheng and geographic distribution of the residents.
9

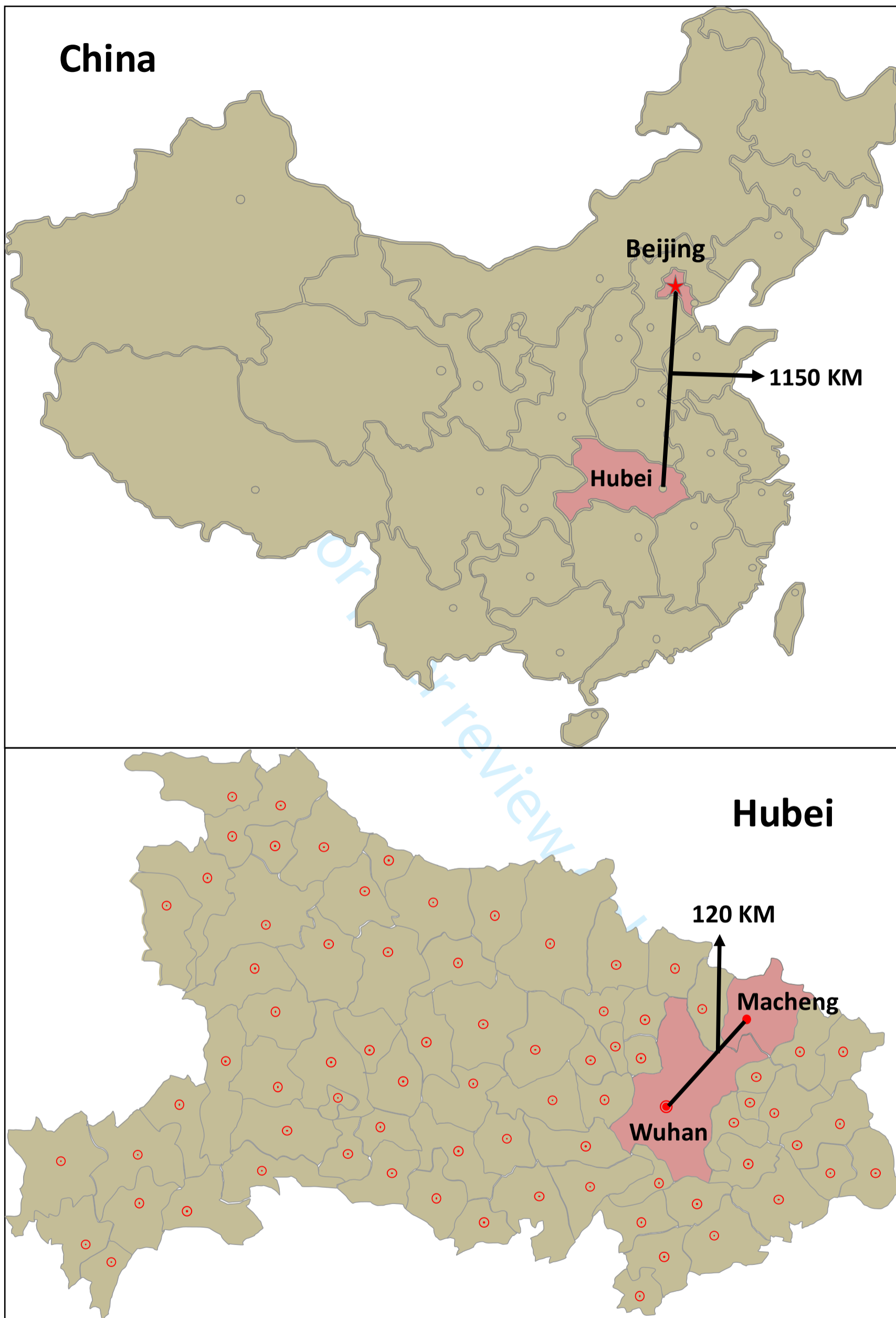
10 460 **Figure 2.** Lorentz curve of the medical expenditure clustering in Macheng in 2014.
11
12

13 461 **Figure 3.** Cost composition of the three groups of rural residents in Macheng in 2014.
14
15

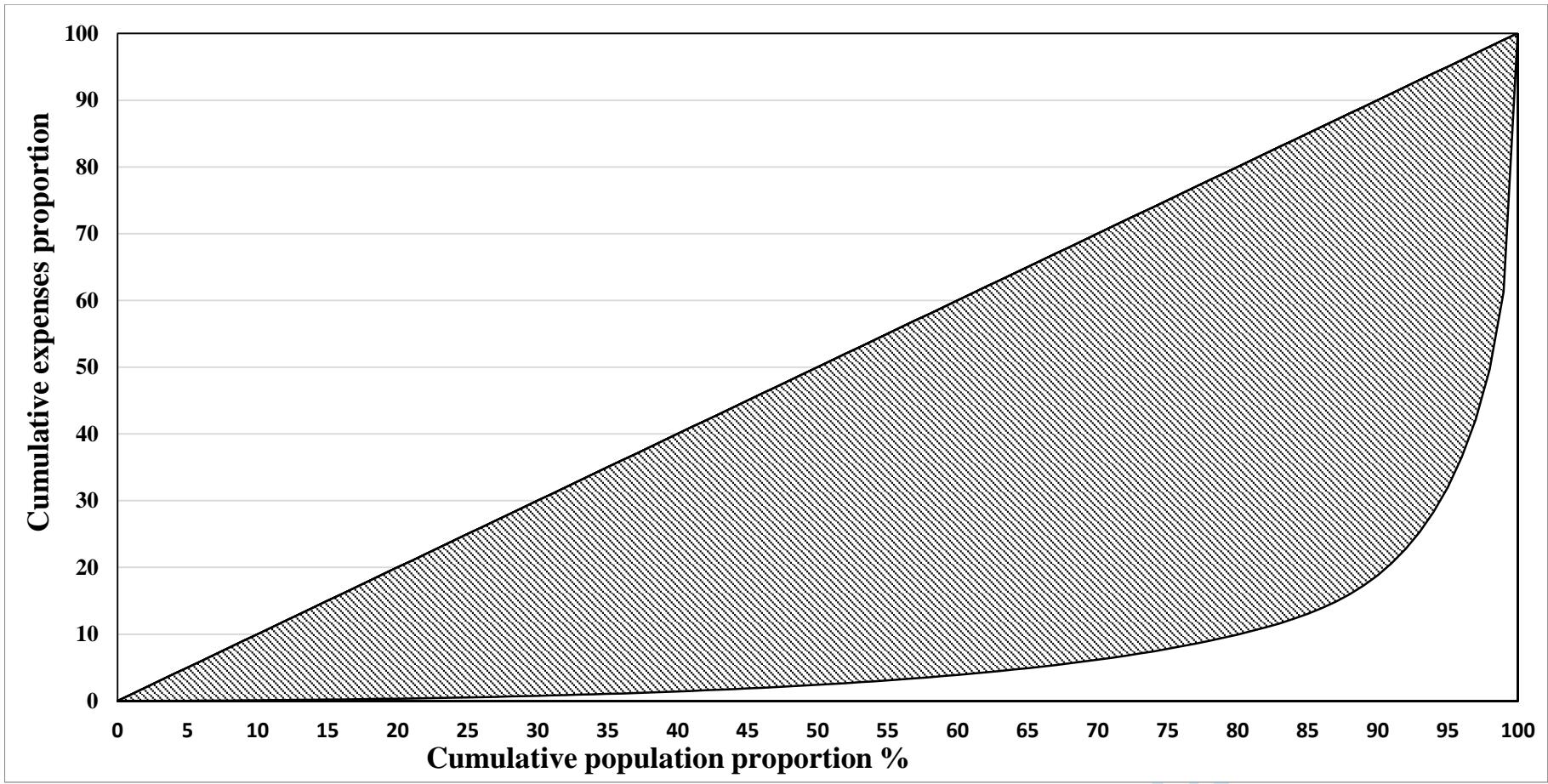
16 462
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

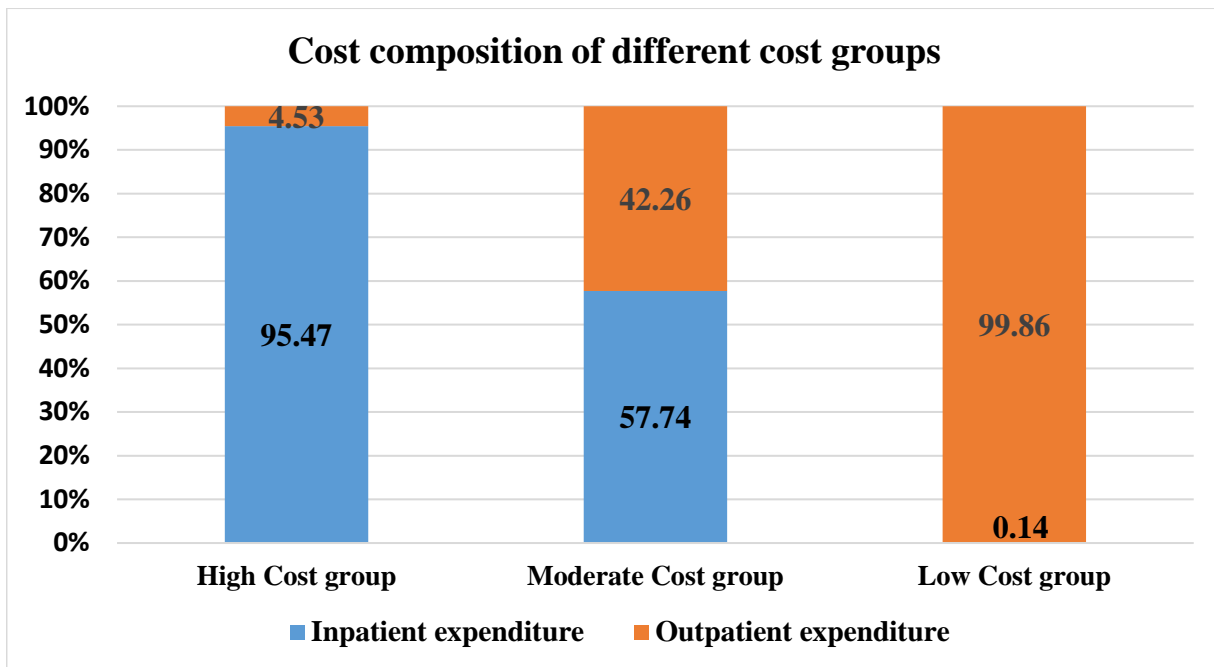
For peer review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46





STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *cross-sectional studies*

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(a) Indicate the study's design with a commonly used term in the title or the abstract	1
		(b) Provide in the abstract an informative and balanced summary of what was done and what was found	2,3
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4,5
Objectives	3	State specific objectives, including any prespecified hypotheses	6
Methods			
Study design	4	Present key elements of study design early in the paper	7
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	7
Participants	6	(a) Give the eligibility criteria, and the sources and methods of selection of participants	8
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	8-9
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	9
Bias	9	Describe any efforts to address potential sources of bias	8
Study size	10	Explain how the study size was arrived at	8
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	inapplicable
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding	9
		(b) Describe any methods used to examine subgroups and interactions	10
		(c) Explain how missing data were addressed	inapplicable
		(d) If applicable, describe analytical methods taking account of sampling strategy	10
		(e) Describe any sensitivity analyses	inapplicable
Results			

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	10
		(b) Give reasons for non-participation at each stage	inapplicable
		(c) Consider use of a flow diagram	inapplicable
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	10
		(b) Indicate number of participants with missing data for each variable of interest	inapplicable
Outcome data	15*	Report numbers of outcome events or summary measures	inapplicable
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	11-15
		(b) Report category boundaries when continuous variables were categorized	11-14
		(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	inapplicable
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	inapplicable
Discussion			
Key results	18	Summarise key results with reference to study objectives	16
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	20
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	16-19
Generalisability	21	Discuss the generalisability (external validity) of the study results	16-19
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	21

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Medical Expenditure Clustering and Determinants of the Annual Medical Expenditures of Residents: A Population-Based Retrospective Study from Rural China

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2018-022721.R1
Article Type:	Research
Date Submitted by the Author:	09-Apr-2018
Complete List of Authors:	Zhang, Yan; Huazhong University of Science and Technology Tongji Medical College, School of Medicine and Health Management Lu, Shan; Huazhong University of Science and Technology Tongji Medical College, School of Medicine and Health Management Yadong, Niu; Huazhong University of Science and Technology, School of Health and Medicine Management ZHANG, LIANG; Huazhong University of Science and Technology Tongji Medical College
Primary Subject Heading:	Health services research
Secondary Subject Heading:	Health policy, Health services research, Medical management
Keywords:	medical expenditure, high-cost, HEALTH ECONOMICS, HEALTH SERVICES ADMINISTRATION & MANAGEMENT, patient flow

SCHOLARONE™
Manuscripts

only

1 **Medical Expenditure Clustering and Determinants of the Annual**
2 **Medical Expenditures of Residents: A Population-Based**
3 **Retrospective Study from Rural China**

4
5
6 Yan Zhang^{1,2}, Shan Lu^{1,2}, Yadong Niu^{1,2}, Liang Zhang^{1,2} §

7
8 ¹ School of Medicine and Health Management, Tongji Medical College, Huazhong
9 University of Science and Technology, Wuhan, Hubei, 430030, China

10 ² Research Centre for Rural Health Service, Key Research Institute of Humanities &
11 Social Sciences of Hubei Provincial Department of Education, Wuhan, Hubei,
12 430030, China

13
14 **Email addresses:**

15 Y.Z. Yan Zhang Ph.D. yanzhang@hust.edu.cn

16 S.L. Shan Lu M.D. shanlu@hust.edu.cn

17 Y-D. N. Yadong Niu M.D. nyadong@126.com

18 L.Z. Liang Zhang M.D. zhangliang@mails.tjmu.edu.cn

19
20 **§Corresponding author:**

21 Liang Zhang, zhangliang@mails.tjmu.edu.cn

22 **Word Count:** 3276

23

Abstract

Objective: To identify the characteristics of high-cost (HC) patients and the determinants of the annual medical expenditures of Chinese rural residents.

Methods: Medical expenditure clustering was performed by Lorentz curve and Gini index. T and chi-square tests were performed to identify the characteristics of the respondents and a multi-level regression model was to examine the determinants of their annual medical expenditures.

Design: A cluster sampling study was performed to identify those residents who availed healthcare services and to assign them to high- (top 5%), moderate- (top 30%) and low-cost (others) groups based on their annual medical expenditures.

Setting: The annual healthcare utilisation was calculated by using data from the population-based database of the 2014 New Rural Cooperative Medical System.

Participants: A total of 478,051 residents who availed healthcare services were recruited for the retrospective study in 2014. The annual medical expenditures of these residents were used as the research object.

Results: The total medical expenditures of Macheng City residents for the year 2014 have a Gini index of 0.81 and around 68.01% of these expenditures can be attributed to HC patients. Female residents (51.5%) and persons aged over 60 years (34.48%) who are suffering from diseases that are difficult to diagnose have a high tendency to accumulate high medical costs. The annual medical expenditures of people living in the same village or town tend to be approximated. Age, disease category, inpatient status, healthcare utilisation and utilisation level are identified as the determinants of annual

1
2
3
4 46 medical expenditures.

5
6 47 **Conclusions:** The medical expenditures of rural residents are clustered at a

7
8 48 remarkably high level and HC patients are suffering from high economic burden.

9
10 49 Therefore, policy makers must guide these patients in seeking appropriate healthcare

11
12 50 services and improve their management of healthcare quality to reduce the

13
14 51 unnecessary healthcare utilisation of these patients.

15
16 52 **Keywords:** medical expenditure; high-cost; clustering; patient flow; rural China

17
18
19
20
21 53

22 23 54 **Strengths and limitations of this study**

24
25 55 • This study is the first to introduce the medical expenditure clustering technique,

26
27 56 which findings can supplement the results of previous research on HC patients.

28
29 57 • The annual medical expenditures of residents are seldom reported at the population

30
31 58 level. This study was conducted in Macheng, a city in Hubei Province with 889,160

32
33 59 residents according to the New Rural Cooperative Medical System (NRCMS)

34
35 60 database, which also stores the inpatient and outpatient records of these residents.

36
37 61 • The Lorentz curve and Gini index were used to cluster the annual medical

38
39 62 expenditures data and a three-level linear regression model was used to aggregate

40
41 63 these data at the residential and town levels.

42
43 64 • The age, gender, hospitalisation, geographic and disease data of the sample were

44
45 65 included in the regression model. However, some individual factors were not

46
47 66 included in the model.

67 **Background**

68 The rapid increase in health expenditures greatly impedes the development of the
69 New Rural Cooperative Medical System (NRCMS), the largest basic social health
70 insurance system in rural China that covers 603.46 million rural residents. Specifically,
71 the health expenditures per capita in China have increased from 513.8¥ (83.6\$) in
72 2012 to 1279.2¥ (208.2\$) in 2017 with an annual growth rate of 25.6%, which is
73 much higher than the annual growth in fundraising per capita (16.02%).^[1] Medical
74 expenditure clustering is considered an important factor that motivates such rapid
75 increase in health expenditures.^[2]

76 Studies on the distribution of diseases within a population have defined clustering as
77 the uneven distribution of disease morbidity in time or space.^[3-4] In line with this
78 definition, medical expenditure clustering indicates the uneven distribution of the
79 medical expenditures of a given population. In recent years, researchers have shown
80 great interest in medical expenditure clustering and have specifically focused on
81 high-cost (HC) patients, who are known for accumulating high annual medical
82 expenditures^[5] and comprise the top 5% biggest spenders in healthcare.^[2] Previous
83 studies have revealed that the medical expenditures of HC patients exceed those of the
84 entire population by 50%.^[6] For instance, in 2014, HC patients account for 52.3% of
85 the total medical expenditures in the US.^[5] These patients and their healthcare quality
86 management have attracted much research attention because of their high healthcare
87 utilisation rate and inappropriate utilisation of healthcare services.^[7] Improving the
88 healthcare quality of these patients can discourage such inappropriate utilisation,

1
2
3
4 89 reduce their health expenditures, conserve social health insurance funds and promote
5
6 90 horizontal equity.
7
8 91 Medical expenditure clustering has also become a major concern in developing
9
10 92 countries. In rural China, the rapid development of NRCMS significantly promoted
11
12 93 the healthcare utilisation rates of residents. For instance, the annual hospitalisation
13
14 94 rate in rural China increased from 8.7% in 2008 to 14.9% in 2017.^[1] An empirical
15
16 95 analysis of seven counties in China revealed that 78.6% of inpatient services in 2015
17
18 96 were distributed amongst one-third of all inpatients in the area. In addition, one
19
20 97 patient in Qianjiang District utilised inpatient services 27 times in 2014. Another
21
22 98 study involving 12,600 families in Jiangsu revealed that HC patients accounted for
23
24 99 44.9% of the total medical expenditures of the entire population.^[8] Moreover, the
25
26 100 medical expenditures clustered at the patient level are much higher than those
27
28 101 clustered at the family level. We also suspect that China has a very high degree of
29
30 102 medical expenditure clustering and that HC patients incur unnecessary medical
31
32 103 expenditures because of the fragmented healthcare delivery system in rural China.
33
34 104 Residents in rural China also seek healthcare services from a three-tier (village–town–
35
36 105 county) healthcare delivery system where the higher tiers provide better services and
37
38 106 charge higher medical costs.^[9] Given that patients neither follow a specified order nor
39
40 107 have limitations when seeking healthcare services and that general practitioners or
41
42 108 consultants are unavailable in most parts of rural China, the residents of rural areas
43
44 109 tend to make uninformed decisions when choosing amongst hospitals and various
45
46 110 types of healthcare services, thereby leading to their inappropriate utilisation of
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 111 healthcare services. For instance, some of these residents, especially HC patients, may
5
6 112 be given inpatient services when they actually require outpatient services and may be
7
8 113 unnecessarily admitted to higher-level hospitals, thereby incurring higher medical
9
10 114 costs.^[10] Yingchun revealed that the inappropriate hospital admission rate in five
11
12 115 counties reached as high as 27.6% in 2014.^[11]

13
14
15 116 To reduce the economic burden of rural residents in China, several policies and
16
17 117 strategies, such as the Tiered Healthcare System and Serious Illness Medical
18
19 118 Insurance (2016), have been proposed. However, the effectivity of these programmes
20
21 119 has received mixed feedback and the focus of these policies remains unsubstantiated,
22
23 120 which may be ascribed to the difficulty in identifying high-demand and HC
24
25 121 patients.^[12] Waxmonsky J. A. argued that those HC patients with multiple or complex
26
27 122 conditions, behavioural disorders or socioeconomic problems are particularly difficult
28
29 123 to monitor.^[13] Moreover, only few studies have investigated the characteristics and
30
31 124 determinants of HC patients by using population-level data because of the lack of
32
33 125 necessary data for exploring their annual medical expenditures. Therefore, the critical
34
35 126 values, average expenditures and inpatient service utilisation of HC patients in rural
36
37 127 China remain unclear.

38
39
40 128 Previous studies reveal that HC patients often maintain a high level of medical
41
42 129 expenditure for the following year. For instance, Robst found that 49.2% of patients
43
44 130 under the Florida Medical Assistance Program (Florida Medicaid) were continuously
45
46 131 classified as HC patients from 2005 to 2010.^[14] Meanwhile, Wodchis et al. found that
47
48 132 one-third of residents with public health insurance in Ontario, Canada continuously
49
50
51
52
53
54
55
56

1
2
3
4 133 incurred high medical expenditures from 2009 to 2011.^[15] Therefore, whilst HC
5
6 134 patients are assumed to possess certain characteristics, the distribution of their
7
8 135 medical expenditures remains unclear.

10
11 136 The growth of health expenditures must be controlled and the efficiency of insurance
12
13 137 funds must be enhanced to improve the health insurance system of a specific area.

14
15
16 138 Medical expenditure clustering can be used to monitor the cost efficiency of
17
18 139 healthcare services. Based on the above findings, identifying HC patients is a
19
20 140 necessary procedure and clustering the medical expenditures of a given population
21
22 141 must be taken as the first step towards achieving such goal.

23
24
25
26 142 This research focuses on medical expenditure clustering in particular as well as on the
27
28 143 distribution and characteristics of HC patients and the determinants of the annual
29
30 144 medical expenditures of residents in general. This study also aims to guide policy
31
32 145 makers and health planners in predicting and planning the future needs of these
33
34 146 patients.

35 36 37 38 39 147 **Methods**

40 41 148 **Study setting**

42
43
44 149 We calculated the annual medical expenditures of residents based on the outpatient
45
46 150 and inpatient services that they have availed within a calendar year. A
47
48 151 population-based retrospective study was performed in Macheng, a typical rural area
49
50 152 in Hubei and county-level city in central China (Fig. 1). Macheng has a total
51
52 153 population of 889,160 and a GDP per capita of 22,758¥ (3,704.83\$).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

154 **[Figure 1]**

155 Macheng has 2 county hospitals, 22 township hospitals and 207 village clinics. The
156 residents of this city are enrolled in NRCMS, which reimburses the medical
157 expenditures of these residents for any type of healthcare service (e.g. outpatient and
158 inpatient services in various medical institutions, including tertiary hospitals in urban
159 areas). The healthcare utilisation information of these residents is recorded in the
160 NRCMS database whenever they request for a reimbursement. Therefore, this study
161 uses the 2014 Macheng City data from the NRCMS database.

162 **Data processing**

163 Those residents of Macheng who availed healthcare services in the past were
164 identified through a retrospective study. After screening the data, 478,051 of the
165 889,160 resident records stored in the NRCMS database were included in the sample
166 and processed by using MS Excel 2010. Firstly, the outpatient and inpatient cases
167 were inputted into a single Excel sheet. Secondly, those cases under the same patient
168 identifier (ID number) were sorted chronologically. Thirdly, the annual medical
169 (inpatient and outpatient) expenditures of each patient, the number of outpatient and
170 inpatient cases and other information were recorded. Fourthly, the annual healthcare
171 utilisation cases were inputted into a new database, where each case represents the
172 annual healthcare utilisation of a resident. Finally, the residents were sorted in a
173 descending order according to the annual medical expenditures stated in their records.
174 Those residents who occupied the top 5% and 6%–30% of the sample were included
175 in the HC and moderate cost (MC) groups, respectively, whilst the other patients were

1
2
3
4 176 included in the low cost (LC) group. These three groups represent the various degrees
5
6 177 of medical expenditure clustering.

7
8 178 The main programming techniques employed in this paper include Excel functions
9
10
11 179 (e.g. COUNTIF, SUMPRODUCT, LOOKUP and IF) and case processing
12
13 180 technologies (e.g. split columns and removal of duplicates). The outstanding diseases
14
15
16 181 of each resident were marked and the original ICD-10 disease codes were adjusted to
17
18 182 broader ones (for example, the disease code for COPD was adjusted from J44.900 to
19
20
21 183 J44). Township hospitals were divided into four levels according to their scale and
22
23 184 service capacity. The distance from and arrival time to county hospitals were
24
25
26 185 individually captured by using Google Maps. The exchange rate of the US dollar
27
28 186 against the RMB in 2014 was 6.1428.

29
30 187 The land form of towns, the healthcare capacity of township hospitals and the
31
32
33 188 sociological characteristics (e.g. gender and age), arrival time to county hospitals,
34
35
36 189 disease categories, healthcare utilisation (e.g. annual length of stay or LOS) and
37
38 190 annual medical expenditures of the residents were collected to build the final
39
40 191 database.^[16]

41 42 43 192 **Statistical analysis**

44
45 193 Firstly, the medical expenditures of the residents were clustered by using the Gini
46
47
48 194 coefficient and Lorenz curve. The Gini coefficient is a digitised representation of
49
50
51 195 medical expenditure clustering. A larger Gini coefficient corresponds to a higher
52
53 196 degree of medical expenditure clustering. Then, the characteristics of the residents in
54
55 197 the HC, MC and LC groups were compared by conducting T test and chi-square test

198 in the IBM SPSS Statistics 22.0 software. At last, the determinants of annual medical
 199 expenditure were then examined by conducting a linear logistic regression analysis.

200 Two key observations were obtained at this stage. Firstly, the obtained data showed a
 201 hierarchical structure. Therefore, the determinants of annual medical expenditure were
 202 examined by conducting a multilevel linear logistic regression analysis using MLwiN
 203 2.30.^[17] Secondly, the annual medical expenditure showed a skewed distribution as
 204 expected. Therefore, this variable was transformed to follow a normal distribution.

205 The patient, village and town were assigned to levels 1, 2 and 3, respectively. After
 206 accepting the Log10 (x) translation, the following regression model was obtained:

$$\begin{aligned} \text{Log10(Annual medical expense)} = & \beta_{0jk}\text{cons} + \beta_1\text{Gender}_{ijk} + \beta_2\text{Age}_{ijk} + \beta_3\text{Family}_{ijk} + \beta_4\text{Distance}_{ijk} + \beta_5\text{Time}_{ijk} + \beta_6\text{Capacity}_{ijk} \\ & + \beta_7\text{Disease}_{ijk} + \beta_8\text{LOS}_{ijk} + \beta_9\text{Coutinp}_{ijk} + \beta_{10}\text{Outinp}_{ijk} + \beta_{11}\text{Coutinp}_{ijk} + \beta_{12}\text{Towinp}_{ijk} \\ & + \beta_{13}\text{Cououtp}_{ijk} + \beta_{14}\text{Towoutp}_{ijk} + \beta_{15}\text{Clioutp}_{ijk} + \text{More}_{ijk} \\ \beta_{0jk} = & \beta_0 + u_{0j} + w_{0jk} \end{aligned}$$

208 where β_{0jk} denotes the fixed-effects parameter whilst u_{0j} and w_{0jk} denote the random
 209 effects at the village and town levels, respectively.

210 Patient and public involvement

211 No patients or members of the public were involved in this research.

212 Ethical approval

213 The study protocol conformed to the guidelines of the Ethics Committee of the Tongji
 214 Medical College of Huazhong University of Science and Technology and was
 215 registered in the Chinese Clinical Trial Registry (ChiCTR-OOR-14005563). The
 216 patient information was anonymised and de-identified before the analysis.

217 Results

218 Medical expenditure clustering in the sample area

219 Table 1 presents the clustering results for the medical expenditures of Macheng
 220 residents in 2014. Amongst these residents, the top 5% and 20% accounted for 68.01%
 221 and 90% of the total medical expenditures of the city, respectively. Figure 2 shows the
 222 Lorentz curve of the clustering results, which have a Gini coefficient of 0.814.

223 [Figure 2]

224 **Table 1. Medical expenditure clustering results for Macheng residents in 2014**

Population ranking (%)	Critical value (¥)	Total medical expenditures (10T ¥)	Cumulative percentage of medical expenditures (10T ¥)	Cumulative percentage of medical expenditures (%)
1	19,123.63	22,634.57	22,634.57	38.73
1-5	4,985.8	17,111.41	39,745.99	68.01
5-10	2,021.45	7,696.77	47,442.76	81.18
10-20	610.22	5,179.36	52,622.12	90.04
20-30	347.29	2,205.70	54,827.82	93.82
30-40	221.26	1,329.19	56,157.01	96.09
40-50	147.37	867.84	57,024.85	97.58
50-60	99.54	582.81	57,607.67	98.57
60-70	63.35	384.42	57,992.09	99.23
70-80	39.95	244.01	58,236.10	99.65
80-90	20.19	141.11	58,377.20	99.89
90-100	3.20	64.17	58,441.37	100.00

225 Table 2 shows the medical expenditure distribution of the HC, MC and LC groups.

226 The HC group has an average annual expenditure per capita of over
 227 15,000¥ (2,441.6\$) and a minimum expenditure of 4,985.80¥ (811.61\$) whilst the LC
 228 group has a maximum expenditure of 347.29¥ (56.54\$). Figure 3 presents the
 229 expenditure composition of these groups.

230 **Table 2. Expenditure distribution of the HC, MC and LC groups (¥)**

Variable	All	Cost Groups		
		HC	MC	LC

Annual expenditure per capita M(SD)	1,222.49 (6,253.01)	16,618.78 (22,817.95)	1,261.36 (1,112.90)	107.99 (89.26)
Minimum expenditure	424,962.10	424,962.10	4,985.60	347.29
Maximum expenditure	3.20	4,985.80	347.30	3.20
Inpatient expenditure per capita M(SD)	976.09 (6,165.55)	15,866.07 (22,745.43)	728.3 (1,182.59)	0.15 (6.39)
Outpatient expenditure per capita M(SD)	246.4 (559.70)	752.71 (2,083.85)	533.06 (417.68)	107.84 (89.17)

231

[Figure 3]**232 Characteristics of HC patients at various clustering levels**

233 Table 3 shows the demographic characteristics of the patients. The three groups
 234 showed significant differences ($P < 0.001$) for nine demographic items. Specifically,
 235 females accounted for 51.5% and 47.42% of the HC and LC groups, respectively.
 236 Most of the residents aged over 60 years were assigned to the HC group. The
 237 residents in the HC and LC groups generally had small (4.01) and large (4.23) family
 238 sizes, respectively. Family size, distance from county hospitals and arrival time to
 239 county hospitals all showed the same change trends. Most of the residents living near
 240 a high-capacity township hospital were assigned to the HC group (15.19%). By
 241 contrast, 30.76% of the members in the LC group were living near a low-capacity
 242 township hospital. These three groups also showed similar distributions across
 243 varying geographic and traffic conditions. Specifically, those members of the LC
 244 group who were living in poor areas (mountains and county roads) showed high
 245 accident rates (12.54% and 32.36%, respectively). In addition, the members of the HC
 246 group were highly likely to develop cancer, circulatory, digestive and urinary diseases
 247 as well as haematological disorders whilst the members of the LC group often

248 developed respiratory diseases.

249 **Table 3. Distribution of the demographic characteristics of residents in the three cost groups**
 250 **(n = 478,051)**

Variable	All n (%)	Cost Groups n (%)			T/F Value
		HC	MC	LC	
All	478,051	23,922 (5.00)	119,492 (25.00)	334,637 (70.00)	
Gender					
Male	242,244 (50.67)	11,602 (48.5)	54,704 (45.78)	175,938 (52.58)	1,673.95*
Female	235,804 (49.33)	12,320 (51.5)	64,786 (54.22)	158,698 (47.42)	
Age (years)					
Mean (SD)	41.97 (20.11)	49.45 (18.47)	45.56 (20.14)	40.14 (19.92)	4,739.48**
Less than 20	77,706 (16.4)	1,374 (5.75)	15,218 (12.79)	61,114 (18.47)	10,792.85*
20–39	114,619 (24.19)	5,142 (21.51)	23,166 (19.47)	86,311 (26.08)	
40–59	181,426 (38.29)	9,147 (38.27)	48,561 (40.81)	123,718 (37.39)	
60–79	93,338 (19.7)	7,753 (32.44)	29,926 (25.15)	55,659 (16.82)	
More than 79	6,706 (1.42)	487 (2.04)	2,118 (1.78)	4,101 (1.24)	
Family size					
Mean (SD)	4.2 (1.54)	4.01 (1.62)	4.15 (1.59)	4.23 (1.51)	350.63**
Distance from CH (km)					
Mean (SD)	30.7 (15.15)	29.34 (15.48)	30.16 (15.44)	31 (15.01)	234.87**
Arrival time to CH (min)					
Mean (SD)	46.5 (18.03)	44.88 (18.6)	45.95 (18.64)	46.81 (17.76)	200.56**
Capacity of TH					
First level (strong)	56,866 (11.9)	3,633 (15.19)	15,868 (13.28)	37,365 (11.17)	690.05*
Second level (general)	276,701 (57.88)	13,304 (55.61)	69,057 (57.79)	194,340 (58.07)	
Third level (weak)	144,484 (30.22)	6,985 (29.2)	34,567 (28.93)	102,932 (30.76)	
Geography					
Plain	175,767 (36.77)	9,753 (40.77)	47,877 (40.07)	118,137 (35.3)	2,701.41*
Hilly	117,197 (24.52)	5,792 (24.21)	31,915 (26.71)	79,490 (23.75)	
Hilly and mountainous	126,341 (26.43)	5,641 (23.58)	25,665 (21.48)	95,035 (28.4)	
Mountainous	58,746 (12.29)	2,736 (11.44)	14,035 (11.75)	41,975 (12.54)	
Traffic condition					
National standard roads	214,571 (44.88)	10,499 (43.89)	53,544 (44.81)	150,528 (44.98)	885.97*
Provincial standard roads	113,392 (23.72)	6,561 (27.43)	31,008 (25.95)	75,823 (22.66)	
County roads	150,088 (31.4)	6,862 (28.68)	34,940 (29.24)	108,286 (32.36)	
Disease category					
Cancer	4,665 (0.98)	2,686 (11.23)	1,452 (1.22)	527 (0.16)	87,301.73*
ENT disease	41,772 (8.74)	2,475 (10.35)	7,114 (5.95)	32,183 (9.62)	
Respiratory disease	181,929 (38.06)	3,312 (13.84)	30,142 (25.23)	148,475 (44.37)	
Circulatory disease	24,721 (5.17)	3,781 (15.81)	9,701 (8.12)	11,239 (3.36)	
Digestive disease	45,723 (9.56)	2,838 (11.86)	9,539 (7.98)	33,346 (9.96)	
Urinary disease	8,048 (1.68)	1,360 (5.69)	3,364 (2.82)	3,324 (0.99)	

Endocrinology disease	5,785 (1.21)	133 (0.56)	792 (0.66)	4,860 (1.45)
Haematological disorders	2,995 (0.63)	427 (1.78)	739 (0.62)	1,829 (0.55)
Bones and muscles	41,675 (8.72)	2,988 (12.49)	11,993 (10.04)	26,694 (7.98)
Obstetrics and gynaecology	15,844 (3.31)	2,223 (9.29)	6226 (5.21)	7,395 (2.21)
Others	104,894 (21.95)	1,699 (7.1)	38,430 (32.16)	64,765 (19.36)

251 *Pearson's chi-square test.

252 **ANOVA.

253 Table 4 shows the distribution of the healthcare utilisation of all Macheng residents in
 254 2014. The residents in all three groups showed a decreasing trend in their average
 255 number of inpatient cases (2.36, 0.45 and 0.02 for the HC, MC and LC groups,
 256 respectively), average number of inpatient cases and annual LOS (25.69, 3.28 and
 257 0.1). However, the opposite trend was observed in the annual LOS (0.07, 62.94 and
 258 99.95) and average number of outpatient cases (8.28, 13.57 and 5.05) of those
 259 residents without prior hospitalisation experience in the HC, MC and LC groups. The
 260 healthcare utilisation of these residents also showed an inverted V-shaped distribution
 261 at the town and clinic levels. Specifically, the residents in the MC group had a high
 262 average number of outpatient cases at the town (3.69) and clinic (8.61) levels, but the
 263 residents in all three groups showed a decreasing trend in their average number of
 264 outpatient cases (1.76, 1.27 and 0.21 for the HC, MC and LC groups, respectively).

265 **Table 4. Distribution of the healthcare utilisation characteristics of the three groups (n =**
 266 **478,051)**

Variable	All n (%)	Cost Groups n (%) / mean(SD)			F
		HC	MC	LC	
Outpatient number					
Mean (SD)	7.34(8.26)	8.28(9.1)	13.57(11.63)	5.05(4.89)	58,639.94**
Outpatient level (time)					
County	0.55 (1.51)	1.76 (2.85)	1.27 (2.26)	0.21 (0.66)	34,430.07**
Township	1.96 (4.14)	2.16 (4.5)	3.69 (6.48)	1.32 (2.57)	15,398.27**
Clinic	4.83 (7.71)	4.36 (7.72)	8.61 (11.78)	3.52 (4.94)	20,854.66**
Inpatient number					
Mean (SD)	0.23 (1.21)	2.36 (4.33)	0.45 (1)	0 (0.02)	55,409.19**

0	409,682 (85.70)	17 (0.07)	75,207 (62.94)	334,458 (99.95)	274,894.92
1–3	64,955 (13.59)	20,840 (87.12)	43,936 (36.77)	179 (0.05)	*
4–7	2,736 (0.57)	2,447 (10.23)	289 (0.24)	0 (0)	
More than 8	678 (0.14)	618 (2.58)	60 (0.05)	0 (0)	
Inpatient level (time)					
Outside the county	0.04 (0.38)	0.58 (1.53)	0.04 (0.22)	0.01 (0.01)	28,920.75**
County	0.14 (1.07)	1.49 (4.09)	0.25 (0.90)	0.01 (0.01)	24,788.76**
Township	0.06 (0.32)	0.29 (0.96)	0.16 (0.45)	0.01 (0.02)	18,940.21**
Annual LOS					
Mean (SD)	2.09 (10.66)	25.46 (36.69)	3.28 (7.91)	0.01 (0.1)	88,485.86**

267 *Pearson's chi-square test.

268 **ANOVA.

269 **Determinants of the annual medical expenditures of residents**

270 A three-level linear regression was performed where the patient, village and town
 271 were assigned to levels 1, 2 and 3, respectively. Table 5 displays the results for the
 272 explanatory variables that are used to fit the three variance component models. Age,
 273 disease category and healthcare utilisation were identified as the major determinants
 274 of the annual medical expenditures of residents. No significant relationship was
 275 observed amongst gender, family size, distance to county hospitals, arrival time to
 276 county hospitals, geography, traffic condition and capacity of township hospitals. The
 277 medical expenditures of these residents increased along with age, number of
 278 outpatient cases, LOS and healthcare utilisation level whilst the other factors were
 279 kept constant. Those patients with cancer, circulatory, digestive, urinary, obstetric and
 280 gynaecological diseases or haematological disorders had a higher probability to incur
 281 high annual medical expenditures compared with those patients with respiratory, ENT,
 282 endocrinal, skeletal and muscular diseases.

283 **Table 5 Three-level linear regression model analysis of annual medical expenditures (n =**
 284 **478,051)**

	Parameter estimate	Standard error	χ^2	P
Fixed factors				

Constant	1.842	0.063	862.479	<0.001
Gender (baseline: male)				
Female	0.003	0.002	5.051	0.025
Age (years)	0.003	0.001	4690.838	<0.001
Family size (person)	0.001	0.001	1.536	0.215
Distance to CH (km)	-0.001	0.001	0.276	0.599
Time to CH (min)	0.001	0.001	0.006	0.938
Capacity of TH	-0.073	0.030	5.797	0.016
Geography (baseline: plains)	0.013	0.020	0.423	0.515
Traffic condition (baseline: national)	0.004	0.022	0.036	0.849
Disease category (baseline: respiratory)				
Cancer	0.759	0.008	8286.833	<0.001
ENT disease	-0.095	0.003	1084.129	<0.001
Circulatory disease	0.303	0.004	4948.426	<0.001
Digestive disease	0.024	0.004	11.12	0.001
Urinary disease	0.304	0.006	2197.804	<0.001
Endocrinology disease	-0.070	0.007	6.756	0.009
Haematological disorders	0.113	0.010	127.488	<0.001
Bones and muscles	0.030	0.008	10.168	0.001
Obstetrics and gynaecology	0.390	0.005	6056.819	<0.001
Else	0.126	0.003	0.294	0.588
Annual LOS (days)	0.015	0.001	18287.060	<0.001
Inpatient level (times)				
Outside the county	0.275	0.002	15149.683	<0.001
County	0.064	0.001	5024.366	<0.001
Township	0.260	0.003	8241.838	<0.001
Outpatient level (times)				
County	0.134	0.001	62795.795	<0.001
Township	0.049	0.001	56609.642	<0.001
Clinic	0.031	0.001	82166.996	<0.001
Random factors				
Town variance	0.005	0.002	9.114	0.003
Village variance	0.007	0.000	298.486	<0.001
Patient scale parameter	1	0.00	-	-

285 Discussion

286 Clustering of the medical expenditures of residents in rural China

287 The medical expenditures of rural residents showed an extremely uneven distribution.

288 A Gini coefficient of 0.814 was recorded in 2014, which indicated that the medical

289 expenditures of this population was clustered at the minority level. The HC group

1
2
3
4 290 accounted for 68.01% of the total medical expenditures, which was much higher than
5
6 291 that recorded in the US (52.3%).^[5] Meanwhile, the LC group, which included 70% of
7
8 292 the residents in the sample, accounted for 2.42% of the total medical expenditures.

9
10 293 The annual medical expenditure per capita of the entire population was
11
12
13 294 1,222.49¥ (199.01\$), which was nearly similar to that of the MC group (1261.36¥,
14
15 295 205.18\$). However, this value was only 7.35% of the annual medical expenditure per
16
17 296 capita of the HC group. The maximum annual medical expenditure of this population
18
19 297 was 424,962.1¥ (69,180.52\$), which is more than four times larger than the annual
20
21 298 reimbursements provided by NRCMS (100,000¥, 16,279.22\$).

22
23
24
25
26 299 The three groups showed obvious differences in their expenditure structures.
27
28 300 Hospitalisation expenditures accounted for over 95% of the medical expenditures of
29
30 301 the HC group, thereby supporting the findings of Driessen et al.^[18] Outpatient
31
32 302 expenditures accounted for most of the medical expenditures of the LC group whilst
33
34 303 outpatient and hospitalization expenditures equally accounted for the medical
35
36 304 expenditures of the MC group. The medical expenditures of the HC group showed a
37
38 305 dispersed distribution and a standard deviation that was much higher than the mean
39
40 306 (22,817.95¥ vs. 16,618.78¥, 3,714.58\$ vs. 2,705.41\$). By contrast, the medical
41
42 307 expenditures of the MC and LC groups showed a relatively concentrated distribution.

43
44
45
46
47
48 308 The HC group faces a very high economic burden. In fact, those residents with
49
50 309 medical expenditures of over 5,000¥ (813.96\$) were included in this group. This
51
52 310 group had an annual medical expenditure per capita of over 16,000¥ (2,604.68\$), of
53
54
55

1
2
3
4 311 which only around 50% were reimbursed by NRCMS. Therefore, HC patients had an
5
6 312 average out-of-pocket (OOP) expenditures over 8,000¥ (1,302.34\$), which was nearly
7
8 313 80% of the total consumer spending per capita of rural residents (10,129.8¥,
9
10 314 1,649.05\$).^[1] This ratio can easily lead to a catastrophic health spending.

13 315 **Aggregating annual medical expenditures at the village and town levels**

16 316 The multilevel linear regression revealed a hierarchical structure (town–village–
17
18 317 residents) in the collected data. The annual medical expenditures of residents were
19
20 318 clustered at the town and village levels whilst the annual medical expenditures of
21
22 319 residents living in the same village and town tend to be approximated. The
23
24 320 distribution of annual medical expenditures as shown in Table 3 significantly differs
25
26 321 from that shown in Table 5 in terms of geography, traffic condition and capacity of
27
28 322 township hospitals. Such differences were mostly observed at the town level because
29
30 323 the residents were living in towns with the same geography, traffic conditions and
31
32 324 capacity of township hospitals. Meanwhile, 30 towns in the sample showed
33
34 325 differences in their social customs, geographic locations and capacity of township
35
36 326 hospitals. These findings altogether show that those residents living in areas with
37
38 327 favourable geographic conditions (e.g. plains and national standard roads) and
39
40 328 high-capacity hospitals tend to accumulate high medical expenditures. Similarly, the
41
42 329 effects of distance from and arrival time to county hospitals were mostly observed at
43
44 330 the village level. Given their convenient locations, those residents who are living near
45
46 331 county hospitals have a higher probability to avail healthcare services in these
47
48 332 institutions and accumulate higher medical expenditures compared with those who are
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 333 living far from these hospitals.
5

6 334 **Determinants of annual medical expenditures**
7

8
9 335 Annual medical expenditures are directly determined by healthcare utilisation. Table 5
10
11 336 shows that a higher utilisation of healthcare services corresponds to higher medical
12
13 337 expenditures. The village, town and county levels had regression coefficients of 0.031,
14
15 338 0.049 and 0.134 for outpatient cases, which were lower than the corresponding
16
17 339 coefficients for inpatient cases. Similar to hospital utilisation, an increase in LOS
18
19 340 corresponds to an increase in medical expenditures.
20
21

22
23 341 In addition, a higher age corresponds to a higher expenditure. This case is particularly
24
25 342 true for those residents aged above 60 years, who account for 34.48% of HC group, in
26
27 343 other words, 8.5% of the elderly population was defined as HC, while only 5% of
28
29 344 total population was defined as HC. Such high expenditures can be attributed to their
30
31 345 poor physical condition and high tendency to develop one or multiple diseases. A
32
33 346 WHO report revealed that an aging population would increase the healthcare
34
35 347 expenditures of a country, but the extent of such increase is less than expected.
36
37

38
39 348 Although health-related needs often increase along with age, the relationship between
40
41 349 healthcare utilisation and health expenditure varies as a person grows older.^[19] In fact,
42
43 350 the medical expenditures of people from high-income countries gradually decline
44
45 351 after they reach the age of 75. Other studies even show that those people aged 80
46
47 352 years and above have a much lower share in the consumption of medical resources
48
49 353 compared with the total population. In terms of gender, the average medical
50
51 354 expenditures of female residents do not differ from those of male residents. The
52
53
54
55
56

1
2
3
4 355 relationship between gender and medical expenditures also warrant further study.^[20]
5
6
7 356 Those patients with respiratory, urinary, endocrinology, skeletal and muscular diseases
8
9 357 have the same degree of medical expenditure clustering. By contrast, those patients
10
11 358 with cancer, circulatory, digestive, obstetrical and gynaecological diseases or
12
13
14 359 haematological disorders have significantly high annual medical expenditures, with
15
16 360 cancer patients having the highest regression coefficient of 0.759. Developing
17
18 361 diseases that are fatal and difficult to assess can easily lead to high medical
19
20
21 362 expenditures.^[21] These findings can be attributed to the limited ability of healthcare
22
23 363 professionals at townships, most of whom lack specialisation in treating urinary and
24
25
26 364 cardiovascular diseases or haematological disorders; therefore, those residents
27
28 365 suffering from such diseases tend to be admitted to county hospitals or hospitals
29
30
31 366 outside their respective counties.^[22]

367 **Redesigning the healthcare delivery system for the HC group**

368 The intense concentration of medical expenditures reveals a great imbalance in the
369 healthcare utilisation of Macheng City residents. Therefore, the rural healthcare
370 delivery system as well as the current efforts in reducing the costs and promoting the
371 cost efficiency of healthcare services should focus on the HC patients.

372 Firstly, a monitoring mechanism for HC patients must be established and the NRCMS
373 database can be used as a source of information that can facilitate the monitoring of
374 these patients. Although HC patients are identified based on their medical expenditure
375 or health claims, a patient can be predicted as HC high probability in advance based

1
2
3
4 376 on several risk factors. Robst et al. and Wodchis et al. found that a patient identified as
5
6 377 HC in a year is more than 40% likely to be identified as an HC patient in the
7
8 378 following year given that these patients often maintain a high level of medical
9
10 379 expenditure for the following year. [14, 15] Therefore, those residents with a
11
12 380 remarkably high healthcare utilisation, are exposed to many risk factors and have
13
14 381 been identified as HC patients in the previous year warrant special attention.
15
16 382 Moreover, HC patients need guidance when choosing healthcare services and must be
17
18 383 given priority access to primary healthcare.
19
20
21
22
23 384 Secondly, enhancing the healthcare quality management of patients plays a vital role
24
25 385 in discouraging their unnecessary utilisation of healthcare services, such as
26
27 386 inappropriate admission to hospitals and excessive utilisation of inpatient services. [23]
28
29
30
31 387 Thirdly, exploring new mechanisms, such as the comprehensive management of
32
33 388 patients with chronic diseases, setting a global budget for multilevel institutions,
34
35 389 integrated management programmes and integrated delivery of medicine and nursing
36
37 390 services to aged residents, can motivate doctors to deliver continued care to HC
38
39 391 patients. In doing so, healthcare professionals can avoid unnecessary service
40
41 392 duplications, improve their healthcare efficacy and help their patients take advantage
42
43 393 of primary healthcare services. [24]
44
45
46
47
48

49 **Conclusion**

50
51 395 From the city-level population perspective, the medical expenditures of rural residents
52
53 396 in China have an intense clustering level. Apart from demographic characteristics (e.g.
54
55
56
57
58
59
60

1
2
3 397 age and disease), healthcare utilisation was identified as a primary determinant of
4
5
6 398 medical expenditure clustering. Therefore, policy makers must guide HC patients in
7
8
9 399 choosing healthcare services and improve their healthcare quality management to
10
11 400 discourage their unnecessary utilisation of healthcare services. Doctors must also be
12
13 401 motivated to deliver continued care to this group of patients.

402 **Limitation**

14
15
16
17
18 403 This study has two limitations. Firstly, hospitalisation information, geographic factors,
19
20
21 404 referral status and diseases were included in the regression model whilst other
22
23
24 405 individual factors, such as economic status, education and preference, were ignored.
25
26 406 Secondly, several studies show that HC residents with a high tendency to develop
27
28 407 multiple chronic diseases tend to utilise multidisciplinary and multi-institutional
29
30
31 408 services.^[25] However, this work only considered the main diseases of these patients
32
33
34 409 as captured in the NRCMS database. These limitations may affect the stability of the
35
36 410 findings and should be examined in further studies.

411 **Competing interests**

40
41 412 The authors declare no competing interests.

413 **Authors' contributions**

42
43
44 414 Y.Z. and S.L. participated in the conception, design, analyses and writing of the
45
46
47 415 manuscript. Y-D.N. participated in the data collection and statistical analysis. L.Z.
48
49
50 416 helped draft, review and revise the manuscript. All authors gave their approval to
51
52
53 417 publish this version of the manuscript.
54
55
56
57
58
59
60

418 Funding statement

419 This research is supported by the National Youth Natural Science Foundation of
420 China (Grant No.: 71603088).

421 Disclosure

422 This work has never been published or presented at any time in the past.

423 Provenance and peer review

424 This work is not commissioned and is externally peer reviewed.

425 Data sharing statement

426 The anonymised dataset can be requested by sending an email to the corresponding
427 author.

428

429

References:

430 [1] NHaFPCoC. *2017 China Statistics Yearbook of Health and Family Planning*. Beijing: China
431 union medical university press, 2017.

432 [2] Coughlin TA, Long SK. Health Care Spending and Service Use among High-Cost Medicaid
433 Beneficiaries, 2002-2004. *INQUIRY-THE JOURNAL OF HEALTH CARE ORGANIZATION
434 PROVISION AND FINANCING*. 2009;**46**(4):405-417 'doi': 10.5034/inquiryjrn1_46.4.405['].

435 [3] Murakami Y, Okamura T, Nakamura K, Miura K, Ueshima H. The clustering of cardiovascular
436 disease risk factors and their impacts on annual medical expenditure in Japan: community-based cost
437 analysis using Gamma regression models. *BMJ open*. 2013;**3**(3) 'doi': 10.1136/bmjopen-2012-002234['].

438 [4] Leech RM, McNaughton SA, Timperio A. The clustering of diet, physical activity and sedentary
439 behavior in children and adolescents: a review. *INTERNATIONAL JOURNAL OF BEHAVIORAL
440 NUTRITION AND PHYSICAL ACTIVITY*. 2014;**11** 'doi': 10.1186/1479-5868-11-4['].

441 [5] Zulman DM, Chee CP, Wagner TH, Yoon J, et al. Multimorbidity and healthcare utilisation
442 among high-cost patients in the US Veterans Affairs Health Care System. *BMJ OPEN*. 2015;**5**(4) 'doi':
443 10.1136/bmjopen-2015-007771['].

444 [6] Zulman DM, Ezeji-Okoye SC, Shaw JG, Hummel DL, et al. Partnered Research in Healthcare
445 Delivery Redesign for High-Need, High-Cost Patients: Development and Feasibility of an Intensive
446 Management Patient-Aligned Care Team (ImPACT). *JOURNAL OF GENERAL INTERNAL
447 MEDICINE*. 2014;**294**:S861-S869 'doi': 10.1007/s11606-014-3022-7['].

448 [7] Hong CS, Siegel AL, Ferris TG. Caring for high-need, high-cost patients: what makes for a
449 successful care management program? Issue brief (Commonwealth Fund). 2014;**19**:1-19.

- 1
2
3 450 [8] Miao Y, Qian D, Sandeep S, Ye T, et al. Exploring the characteristics of the high-cost population
4 451 from the family perspective: a cross-sectional study in Jiangsu Province, China. *BMJ open*.
5 452 2017;**7**(11):e17185 'doi': 10.1136/bmjopen-2017-017185['].
- 6
7 453 [9] Li X, Lu JP, Hu S, Cheng KK, et al. The primary health-care system in China. *LANCET*.
8 454 2017;**390**(10112):2584-2594ELSEVIER SCIENCE INC.
- 9 455 [10] Yip W, Hsiao W. Harnessing the privatisation of China's fragmented health-care delivery.
10 456 *LANCET*. 2014;**384**(9945):805-818 'doi': 10.1016/S0140-6736(14)61120-X['].
- 11 457 [11] Zhang Y, Chen Y, Zhang X, Zhang L. Current level and determinants of inappropriate admissions
12 458 to township hospitals under the new rural cooperative medical system in China: a cross-sectional study.
13 459 *BMC HEALTH SERVICES RESEARCH*. 2014;**14**(649).
- 14 460 [12] Bleich SN. Systematic Review of Programs Treating High-Need and High-Cost People With
15 461 Multiple Chronic Diseases or Disabilities in the United States, 2008-2014 (vol 12, E197, 2015).
16 462 *PREVENTING CHRONIC DISEASE*. 2016;**13** 'doi': 10.5888/pcd13.150275e['].
- 17 463 [13] Waxmonsky JA, Giese AA, McGinnis GF, Reynolds RT, et al. Colorado access' enhanced care
18 464 management for high-cost, high-need Medicaid members: preliminary outcomes and lessons learned.
19 465 *The Journal of ambulatory care management*. 2011;**34**(2):183-191 'doi':
20 466 10.1097/JAC.0b013e31820f64be['].
- 21 467 [14] Robst J. Developing Models to Predict Persistent High-Cost Cases in Florida Medicaid.
22 468 *POPULATION HEALTH MANAGEMENT*. 2015;**18**(6):467-476 'doi': 10.1089/pop.2014.0174['].
- 23 469 [15] Wodchis WP, Austin PC, Henry DA. A 3-year study of high-cost users of health care.
24 470 *CANADIAN MEDICAL ASSOCIATION JOURNAL*. 2016;**188**(3):182-188 'doi':
25 471 10.1503/cmaj.150064['].
- 26 472 [16] Herrin J, St Andre J, Kenward K, Joshi MS, Audet AJ, Hines SC. Community Factors and
27 473 Hospital Readmission Rates. *HEALTH SERVICES RESEARCH*. 2015;**50**(1):20-39.
- 28 474 [17] Longford NT. Improved approximations for multilevel models with binary responses. *JOURNAL*
29 475 *OF THE ROYAL STATISTICAL SOCIETY SERIES A-STATISTICS IN SOCIETY*.
30 476 1997;**160**(3):593.
- 31 477 [18] Driessen RJB, Bisschops LA, Adang EMM, Evers AW, van de Kerkhof PCM, de Jong EMGJ.
32 478 The economic impact of high-need psoriasis in daily clinical practice before and after the introduction
33 479 of biologics. *BRITISH JOURNAL OF DERMATOLOGY*. 2010;**162**(6):1324-1329 'doi':
34 480 10.1111/j.1365-2133.2010.09693.x['].
- 35 481 [19] Beard JR, Officer AM, Cassels AK. The World Report on Ageing and Health.
36 482 *GERONTOLOGIST*. 2016;**56**2:S163-S166 'doi': 10.1093/geront/gnw037['].
- 37 483 [20] Owens GM. Gender differences in health care expenditures, resource utilization, and quality of
38 484 care. *JOURNAL OF MANAGED CARE PHARMACY*. 2008;**14**3(3):S2-S6.
- 39 485 [21] Ross JS, Chen J, Lin Z, Bueno H, et al. Recent National Trends in Readmission Rates After Heart
40 486 Failure Hospitalization. *CIRCULATION-HEART FAILURE*. 2010;**3**(1):97-103.
- 41 487 [22] Hartl S, Luis Lopez-Campos J, Pozo-Rodriguez F, Castro-Acosta A, et al. Risk of death and
42 488 readmission of hospital-admitted COPD exacerbations: European COPD Audit. *EUROPEAN*
43 489 *RESPIRATORY JOURNAL*. 2016;**47**(1):113-121.
- 44 490 [23] Feder JL. Predictive Modeling And Team Care For High-Need Patients At HealthCare Partners.
45 491 *HEALTH AFFAIRS*. 2011;**30**(3):416-418 'doi': 10.1377/hlthaff.2011.0080['].
- 46 492 [24] Ye T, Sun XW, Tang WX, Miao Y, Zhang YD, Zhang L. Effect of continuity of care on
47 493 health-related quality of life in adult patients with hypertension: a cohort study in China. *BMC*

1
2
3 494 HEALTH SERVICES RESEARCH. 2016;**16**(674)BIOMED CENTRAL LTD.
4 495 [25] Zulman D, Yoon J, Cohen DM, Wagner TH, Ritchie C, Asch S. MULTIMORBIDITY AND
5 496 HEALTH CARE UTILIZATION AMONG HIGH-COST PATIENTS: IMPLICATIONS FOR CARE
6 497 COORDINATION. JOURNAL OF GENERAL INTERNAL MEDICINE. 2013;**281**:S123-S124.
7

8
9 **498 Figure Legend**

10
11 **499 Figure 1.** Map of Macheng City and geographic distribution of the residents.
12

13
14 **500 Figure 2.** Lorentz curve of the medical expenditure clustering for Macheng in 2014.
15

16
17 **501 Figure 3.** Cost composition of the three groups of rural residents in Macheng in 2014.
18

19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

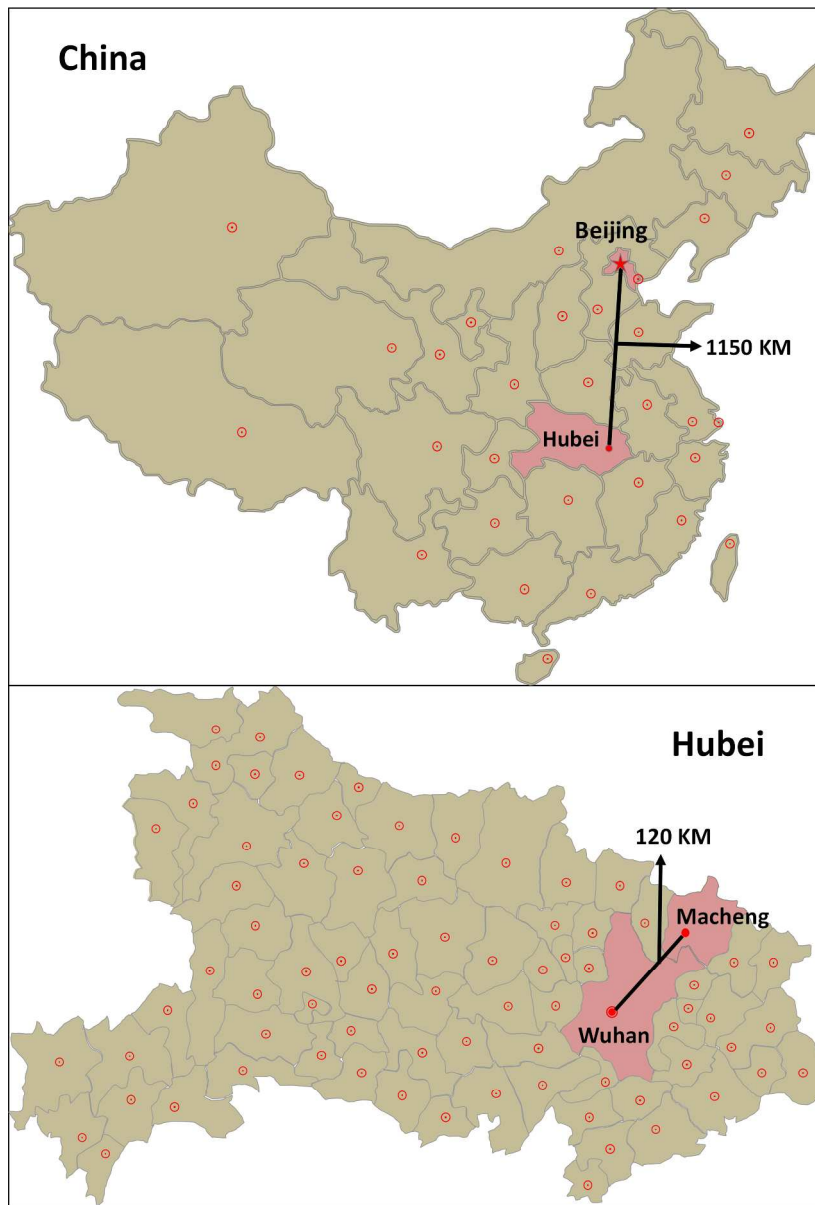


Figure 1. Map of Macheng City and geographic distribution of the residents.

331x486mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

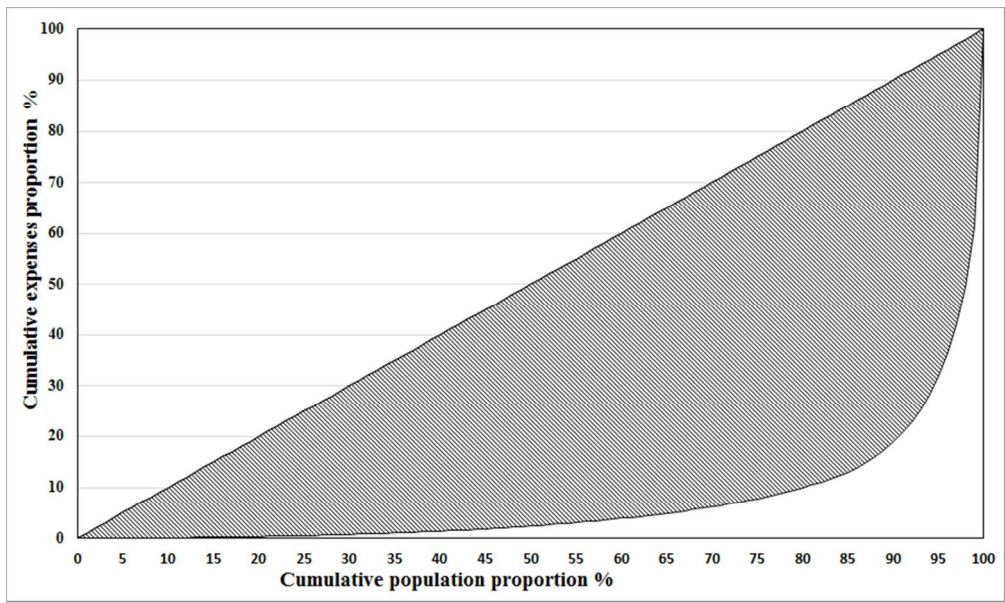


Figure 2. Lorentz curve of the medical expenditure clustering for Macheng in 2014.

review only

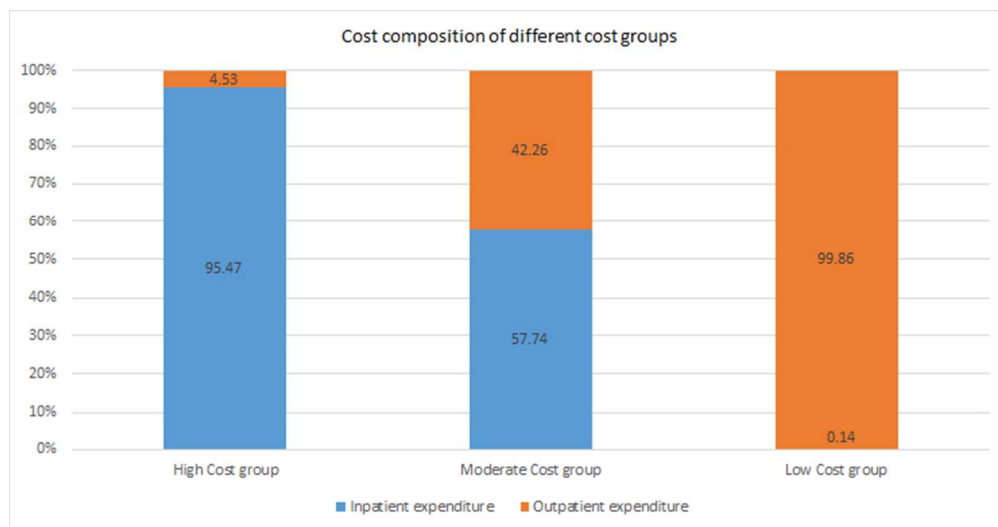


Figure 3. Cost composition of the three groups of rural residents in Macheng in 2014.

STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *cross-sectional studies*

Section/Topic	Item #	Recommendation	Reported on Page#; Line #
Title and abstract	1	(a) Indicate the study's design with a commonly used term in the title or the abstract	P1; L1-2
		(b) Provide in the abstract an informative and balanced summary of what was done and what was found	P2-3; L24-66
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	P4-7; L68-135
Objectives	3	State specific objectives, including any prespecified hypotheses	P7; L136-147
Methods			
Study design	4	Present key elements of study design early in the paper	P2; L31-33
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	P7-8; L149-162
Participants	6	(a) Give the eligibility criteria, and the sources and methods of selection of participants	P8; L163-168
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	P8-9; L172-191
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	P8; L166-172
Bias	9	Describe any efforts to address potential sources of bias	No
Study size	10	Explain how the study size was arrived at	P8; L163-166
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	No
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding	P9-10; L192-209
		(b) Describe any methods used to examine subgroups and interactions	P9; L196-198
		(c) Explain how missing data were addressed	No
		(d) If applicable, describe analytical methods taking account of sampling strategy	No
		(e) Describe any sensitivity analyses	No
Results			

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	P11-12; L218-244
		(b) Give reasons for non-participation at each stage	No
		(c) Consider use of a flow diagram	No
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	P13-15; L233-269
		(b) Indicate number of participants with missing data for each variable of interest	No
Outcome data	15*	Report numbers of outcome events or summary measures	P12-13; L218-232
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	P15-16; L270-285
		(b) Report category boundaries when continuous variables were categorized	No
		(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	No
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	No
Discussion			
Key results	18	Summarise key results with reference to study objectives	P16-21; L287-394
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	P22; L403-411
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	P22; L395-402
Generalisability	21	Discuss the generalisability (external validity) of the study results	No
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	P23;L419-421

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is available at www.strobe-statement.org.