BMJ Open

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

Journal:	*BMJ Open*
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

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

Yan Zhang[1,2], Shan Lu[1,2], Yadong Niu[1,2], Liang Zhang[1,2 §]

[1] School of Medicine and Health Management, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, 430030, China

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

Email addresses:

Y.Z.	Yan Zhang	Ph.D.	yanzhang@hust.edu.cn
S.L.	Shan Lu	M.D.	shanlu@hust.edu.cn
Y-D. N.	Yadong Niu	M.D.	nyadong@126.com
L.Z.	Liang Zhang	M.D.	zhangliang@mails.tjmu.edu.cn

§Corresponding author:

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

Word Count: 3276

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 Lorentz 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

medical expenditures.

Conclusions: The medical expenditures of rural residents are clustering at a remarkably high level and HC patients are suffering from high economic burden. Therefore, policymakers must prioritise guiding HC patients to seek healthcare and strengthening healthcare quality management to reduce unnecessary healthcare utilisation.

Keywords: medical expenditure; high-cost; clustering; patient flow; rural China

Strengths and limitations of this study

- This study is the first to introduce medical expenditure clustering, which can supplement the findings of other studies on high-cost (HC) patients.
- Population-level data on the annual medical expenditure of residents are seldom reported. This study was conducted in a city with 889,160 residents according to a processed New Rural Cooperative Medical System (NRCMS) database, which contained the inpatient and outpatient records of these residents.
- The Lorentz curve and Gini index were prescribed for medical expenditure clustering and a three-level linear regression model was used for the aggregation at the residential and town levels.
- Age, gender, hospitalisation information, geographic factors and diseases were drawn into the regression model, but some individual factors were absent.

Background

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

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

Qianjiang District utilised inpatient services 27 times in a single year. Another study involving 12,600 families in Jiangsu revealed that the HC families accounted for 44.9% of the total medical expenditures of the entire population.[8] Moreover, clustering at the patient level is much higher than that at the family level. We also suspect that the degree of medical expenditure clustering in rural China is very high and that HC patients incur unnecessary medical expenditures because of the fragmented healthcare delivery system in rural China. Residents in rural China also seek healthcare services from a three-tier (village–town–county) healthcare delivery system and a higher level of the institution corresponds to a better service capability and higher medical costs.[9] Given that patient choice neither follows a specified order nor has limitations and that general practitioners or consultants are unavailable in most parts of rural China, the residents make uninformed decisions when choosing amongst hospitals and various types of healthcare services, thereby resulting in the inappropriate utilisation of healthcare services. For instance, some of these residents, especially HC patients, may be given inpatient services when they actually require outpatient services and may be unnecessarily admitted to higher-level hospitals, thereby increasing their medical expenditures.[10] Yingchun revealed that the inappropriate admission rate in five counties reached as high as 27.6% in 2014.[11]

To reduce the economic burden of rural residents in China, several policies and strategies, such as the Tiered Healthcare System and Serious Illness Medical Insurance (2016), have been proposed. However, the feedback on the effectivity of these programmes is mixed and the object of focus of these policies remains

unsubstantiated. Such lack of evidence may be attributed to the difficulty in identifying high-demand and HC patients.[12] Waxmonsky J A descripted HC patients with multiple or complex conditions, often combined with behavioural health problems or socioeconomic challenges, they are difficult to monitor.[13] Moreover, only few studies have investigated the characteristics and determinants of HC patients based on population-level data because of the lack of necessary data for exploring annual medical expenditures. So the critical values, average expenditure and inpatient service utilisation of HC patients are all unclear in rural China.

Previous studies reveal that HC patients often maintain a high level of medical expenditure for the following year. For instance, Robst pointed out that 49.2% of HC patients continuously were classified HC patient, under Florida Medical Assistance Program (Florida Medicaid) from 2005 to 2010.[14] Meanwhile, Wodchis et al. found the ratio is 1/3 based 2009 to 2011 of public health insurance in Ontario, Canada.[15] Therefore, HC patients are assumed to possess certain characteristics even though their distribution is unclear.

Based on the above findings, identifying HC patients is a necessary procedure. Clarifying the medical expenditure clustering of a given population must be taken as the first step towards such goal. This research focuses on medical expenditure clustering in particular and distinguishes the distribution and characteristics of HC patients and the determinants of the annual medical expenditures of residents in general. This study also aims to guide policymakers and health planners in predicting and planning the future needs of these patients.

Methods

Study setting

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

[Figure 1]

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

Data processing

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

recorded. Fourthly, the annual healthcare utilisation cases were inputted into a new database, where each case represents the annual healthcare utilisation of a resident. Finally, the residents were sorted in a descending order according to their annual medical expenditures. Those residents that occupied the top 5% of the sample were included in the HC group, those in the top 6% to 30% were included in the moderate cost (MC) group and the other patients were included in the low cost (LC) group. These three groups represent the various degrees of medical expenditure clustering. The main programming techniques include Excel formulas (e.g. COUNTIF, SUMPRODUCT, LOOKUP and IF) and case processing technologies (e.g. split columns and removal of duplicates). We marked the outstanding diseases of each resident and adjusted the original ICD-10 disease code to a broader one (for example, the disease code for COPD was adjusted from J44.900 to J44). Township hospitals were divided into four levels according to their scale and service capacity. The distance from and arrival time to county hospitals were individually captured by using Google Maps.

The land form of towns, the healthcare capacity of township hospitals and the sociological characteristics (e.g. gender and age), arrival time to the county hospital, disease category, healthcare utilisation (e.g. annual length of stay (LOS)) and annual medical expenditures of the patients were collected to build the final database.[16]

The study protocol conformed to the guidelines of the Ethics Committee of the Tongji Medical College of Huazhong University of Science and Technology. The patient information was anonymised and deidentified before the analysis.

Statistical analysis

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

$$\begin{aligned}\mathrm{Log10}(\text{Annual medical expense}) &= \beta_{0jk}\mathrm{cons} + \beta_1\mathrm{Gender}_{ijk} + \beta_2\mathrm{Age}_{ijk} + \beta_3\mathrm{Family}_{ijk} + \beta_4\mathrm{Distance}_{ijk} + \beta_5\mathrm{Time}_{ijk} + \beta_6\mathrm{Capacity}_{ijk} \\ &\quad + \beta_7\mathrm{Disease}_{ijk} + \beta_8\mathrm{LOS}_{ijk} + \beta_9\mathrm{Coutinp}_{ijk} + \beta_{10}\mathrm{Outinp}_{ijk} + \beta_{11}\mathrm{Coutinp}_{ijk} + \beta_{12}\mathrm{Towinp}_{ijk} \\ &\quad + \beta_{13}\mathrm{Cououtp}_{ijk} + \beta_{14}\mathrm{Towoutp}_{ijk} + \beta_{15}\mathrm{Clioutp}_{ijk} + \mathrm{More}_{ijk} \\ \beta_{0jk} &= \beta_0 + \mathrm{u}_{0j} + \mathrm{w}_{0jk}\end{aligned} \tag{187}$$

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

Results

Medical expenditure clustering in sample area

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

[Figure 2]

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

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

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)

[Figure 3]

Characteristics of HC patients at various clustering levels

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

were observed amongst the three groups for nine demographic items (all $P < 0.001$). Specifically, females accounted for 51.5% and 47.42% of the residents in the HC and LC groups, respectively. Average age also varied across these groups. Specifically, those residents aged over 60 years mostly belonged to the HC group. The residents in the HC and LC groups generally had small (4.01) and large (4.23) family sizes, respectively. Distance from and arrival time to county hospital showed the same change trends in terms of family size. Those residents living nearby a high-capacity township hospital mostly belonged to the HC group (15.19%), by contrast, the residents living nearby a low-capacity township hospital accounted for 30.76% of all members of the LC group. These three groups showed similar distributions across varying geographic and traffic conditions. Specifically, in the LC group, those residents living in poor areas (mountains and county roads) had high accident rates (12.54% and 32.36%, respectively). In addition, HC patients were highly likely to develop ailments, such as cancer, circulatory, digestive and urinary diseases as well as haematological disorders, whilst LC patients frequently developed respiratory diseases.

Table 3. Distribution of the demographic characteristics of residents in the three cost groups (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)	

*Pearson's chi-square test.

**ANOVA.

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

The average number of inpatient cases demonstrated a decreasing trend for the HC, MC and LC groups (2.36, 0.45 and 0.02, respectively). Similar results were obtained

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

Table 4. Distribution of the healthcare utilisation characteristics of the three groups (n = 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**

*Pearson's chi-square test.

**ANOVA.

Determinants of the annual medical expenditure of residents

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

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

Table 5 Three-level linear regression model analysis of annual medical expenditures (n = 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	–	–

Discussion

Clustering of the medical expenditures of residents in rural China

The medical expenditures of rural residents showed an extremely uneven distribution. With a Gini coefficient of 0.814 in 2014, medical expenditure was clustered in the minority population. The HC group accounted for 68.01% of the total medical expenditures; this proportion is much higher than that observed in the US, where the HC group accounted for 52.3% of the total medical expenditures in the same year.[3] Meanwhile, the LC group, 70% of population, accounted for 2.42% of the total medical expenditures. The annual medical expenditure per capita for the entire population was 1,222.49 ¥, which is nearly similar to that for the MC group (1261.36¥) and 7.35% for the HC group. The maximum annual medical expenditure of this population was 424,962.1¥, which is more than four times larger than the annual reimbursement amount provided by NRCMS (100,000¥).

The three groups showed obvious differences in their expenditure structures. Hospitalisation expenditures accounted for over 95% of the medical expenditures of the HC group, thereby supporting the findings of Driessen et al.[18] Meanwhile, outpatient expenditures accounted for most of the medical expenditures of the LC group, medical expenditure of the MC group is made up of outpatient and hospitalization expenditures coequally. The medical expenditures of the HC group showed a dispersed distribution with a standard deviation that was much higher than the mean (22,817.95 vs. 16,618.78). By contrast, the medical expenditures of the MC and LC groups showed a relatively concentrated distribution.

The HC group faces a very high economic burden such that those residents with over 5,000¥ in medical costs are identified as members of this group. The annual medical expenditure per capita of the HC group is over 16,000¥ whereas the actual reimbursement ratio of NRCMS is approximately 50%. Therefore, the average Out of Pocket (OOP) of HC patients is over 8,000¥, which is nearly 80% of the total consumer spending per capita of rural residents (10,129.8¥).[7] This ratio can lead to a catastrophic health spending easily.

Aggregating annual medical expenditures at the village and town level

The multilevel linear regression revealed that the resident data followed a hierarchical structure (town–village-residents). The annual medical expenditures of residents were clustered at the town and village levels, the annual medical expenditures for the same village and town tend to be approximated. The distribution of annual medical

expenditures in Table 3 significantly differs from that in Table 5 in terms of home geography, traffic condition and capacity of township hospitals. Such differences were mostly observed at the town level because the residents were living in a town with the same home geography, traffic condition and township hospital. In addition, 30 towns in the sample showed differences in their social customs, geographic location and capacity of their township hospitals whereas the other township hospitals in the sample had different operating capacities. Taken together, those residents living in areas with favourable geographic conditions (e.g. plains and national standard roads) and hospitals with high capacity tend to accumulate high medical expenditures. Similarly, the effects of distance and arrival time to county hospital were mostly observed at the village level, shorter distance from and arrival time of these residents to the county hospital, higher probability of availing county healthcare services, and higher medical expenditures. The decision of these residents to pursue the most convenient option can be easily understood.

Determinants of annual medical expenditures

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

for those residents aged above 60 years (comprising 8.5% of the HC group) because of their poor physical condition and high tendency to develop one or multiple diseases. A WHO report revealed that an aging population would increase healthcare spending but the extent of such increase is less than expected. Although health-related needs often increase with age, the relationship between healthcare utilisation and health expenditure is variable.[19] In fact, the medical expenditures of people from several high-income countries gradually declined after reaching the age of 75. Other studies even show that those people aged 80 years and above are major contributors of medical resources, that is, their share in the consumption of medical resources is much lower than its share of the population. In terms of gender, the average medical expenditures of female residents do not differ from those of male residents, there is no conclusive research on the relationship between gender and medical expenditures.[20]

Respiratory, urinary, endocrinology, skeletal and muscular diseases have the same degree of medical expenditure clustering. By contrast, other diseases, such as cancer, circulatory, digestive, obstetrical and gynaecological diseases or haematological disorders, have a significantly high annual medical expenditures, with cancer having the highest regression coefficient of 0.759. Those diseases that are difficult to assess and can easily lead to death greatly contribute to high medical expenditures.[21] This finding can be attributed to the ability of healthcare professionals at the township level, most of which do not specialise in urinary and cardiovascular diseases or haematological disorders; therefore, those residents suffering from such diseases tend to be admitted to county hospitals or hospitals outside their respective counties.[22]

Redesigning the health delivery system for the HC group

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

Conclusion

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

of medical expenditure clustering. Policymakers must consider patient prioritisation to guide HC patients in seeking for healthcare services. They must also improve healthcare quality management to discourage the unnecessary utilisation of healthcare services. Doctors must also be given motivation to deliver continued care to their patients.

Limitation

This study has two limitations. Firstly, hospitalisation information, geographic factors, referral status and diseases were included in the regression model whilst other individual factors, such as economic status, education and preference, were ignored. Secondly, several studies point out that HC residents with a high tendency to develop multiple chronic diseases tend to utilise multidisciplinary and multi-institutional services.[25] However, we only considered main disease, the diagnoses from the NRCMS database. These limitations may affect the stability of our findings and should be addressed in further studies.

Competing interests

The authors have declared that no competing interests exist.

Authors' contributions

Y.Z. and S.L. participated in conception and design, and the analyses, and wrote the manuscript. Y-D.N. participated in data collection and performed the statistical analysis. L.Z. helped to draft the manuscript, reviewed the manuscript and made final changes. All authors have given their final approval of the version to be published.

Funding statement

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

Disclosure

No past publication history, no past presentation history

Provenance and peer review

Not commissioned; externally peer reviewed.

Data sharing statement

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

References:

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

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

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

[4] Coughlin TA, Long SK. Health Care Spending and Service Use among High-Cost Medicaid Beneficiaries, 2002-2004. INQUIRY-THE JOURNAL OF HEALTH CARE ORGANIZATION PROVISION AND FINANCING. 2009;**46**(4):405-417 'doi': 10.5034/inquiryjrnl_46.4.405['.

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

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

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

[8] Miao Y, Qian D, Sandeep S, Ye T, et al. Exploring the characteristics of the high-cost population from the family perspective: a cross-sectional study in Jiangsu Province, China. BMJ open.

2017;**7**(11):e17185 'doi': 10.1136/bmjopen-2017-017185['.

[9] Li X, Lu JP, Hu S, Cheng KK, et al. The primary health-care system in China. LANCET. 2017;**390**(10112):2584-2594ELSEVIER SCIENCE INC.

[10] Yip W, Hsiao W. Harnessing the privatisation of China's fragmented health-care delivery. LANCET. 2014;**384**(9945):805-818 'doi': 10.1016/S0140-6736(14)61120-X['.

[11] Zhang Y, Chen Y, Zhang X, Zhang L. Current level and determinants of inappropriate admissions to township hospitals under the new rural cooperative medical system in China: a cross-sectional study. BMC HEALTH SERVICES RESEARCH. 2014;**14**(649).

[12] Bleich SN. Systematic Review of Programs Treating High-Need and High-Cost People With Multiple Chronic Diseases or Disabilities in the United States, 2008-2014 (vol 12, E197, 2015). PREVENTING CHRONIC DISEASE. 2016;**13** 'doi': 10.5888/pcd13.150275e['.

[13] Waxmonsky JA, Giese AA, McGinnis GF, Reynolds RT, et al. Colorado access' enhanced care management for high-cost, high-need Medicaid members: preliminary outcomes and lessons learned. The Journal of ambulatory care management. 2011;**34**(2):183-191 'doi': 10.1097/JAC.0b013e31820f64be['.

[14] Robst J. Developing Models to Predict Persistent High-Cost Cases in Florida Medicaid. POPULATION HEALTH MANAGEMENT. 2015;**18**(6):467-476 'doi': 10.1089/pop.2014.0174['.

[15] Wodchis WP, Austin PC, Henry DA. A 3-year study of high-cost users of health care. CANADIAN MEDICAL ASSOCIATION JOURNAL. 2016;**188**(3):182-188 'doi': 10.1503/cmaj.150064['.

[16] Herrin J, St Andre J, Kenward K, Joshi MS, Audet AJ, Hines SC. Community Factors and Hospital Readmission Rates. HEALTH SERVICES RESEARCH. 2015;**50**(1):20-39.

[17] Longford NT. Improved approximations for multilevel models with binary responses. JOURNAL OF THE ROYAL STATISTICAL SOCIETY SERIES A-STATISTICS IN SOCIETY. 1997;**160**(3):593.

[18] Driessen RJB, Bisschops LA, Adang EMM, Evers AW, van de Kerkhof PCM, de Jong EMGJ. The economic impact of high-need psoriasis in daily clinical practice before and after the introduction of biologics. BRITISH JOURNAL OF DERMATOLOGY. 2010;**162**(6):1324-1329 'doi': 10.1111/j.1365-2133.2010.09693.x['.

[19] Beard JR, Officer AM, Cassels AK. The World Report on Ageing and Health. GERONTOLOGIST. 2016;**562**:S163-S166 'doi': 10.1093/geront/gnw037['.

[20] Owens GM. Gender differences in health care expenditures, resource utilization, and quality of care. JOURNAL OF MANAGED CARE PHARMACY. 2008;**143**(3):S2-S6.

[21] Ross JS, Chen J, Lin Z, Bueno H, et al. Recent National Trends in Readmission Rates After Heart Failure Hospitalization. CIRCULATION-HEART FAILURE. 2010;**3**(1):97-103.

[22] Hartl S, Luis Lopez-Campos J, Pozo-Rodriguez F, Castro-Acosta A, et al. Risk of death and readmission of hospital-admitted COPD exacerbations: European COPD Audit. EUROPEAN RESPIRATORY JOURNAL. 2016;**47**(1):113-121.

[23] Feder JL. Predictive Modeling And Team Care For High-Need Patients At HealthCare Partners. HEALTH AFFAIRS. 2011;**30**(3):416-418 'doi': 10.1377/hlthaff.2011.0080['.

[24] Ye T, Sun XW, Tang WX, Miao Y, Zhang YD, Zhang L. Effect of continuity of care on health-related quality of life in adult patients with hypertension: a cohort study in China. BMC HEALTH SERVICES RESEARCH. 2016;**16**(674)BIOMED CENTRAL LTD.

[25] Zulman D, Yoon J, Cohen DM, Wagner TH, Ritchie C, Asch S. MULTIMORBIDITY AND

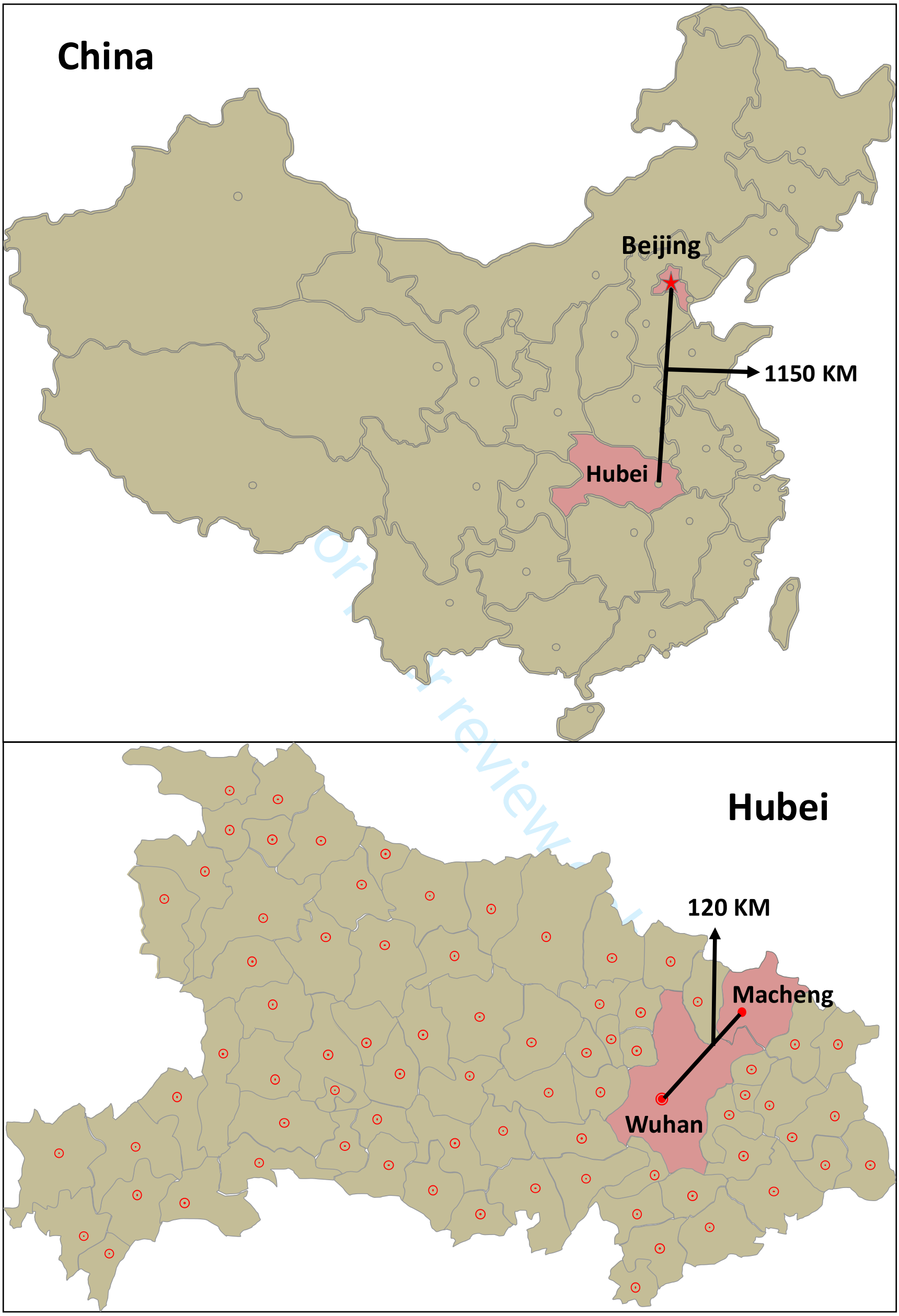
HEALTH CARE UTILIZATION AMONG HIGH-COST PATIENTS: IMPLICATIONS FOR CARE COORDINATION. JOURNAL OF GENERAL INTERNAL MEDICINE. 2013;**281**:S123-S124.

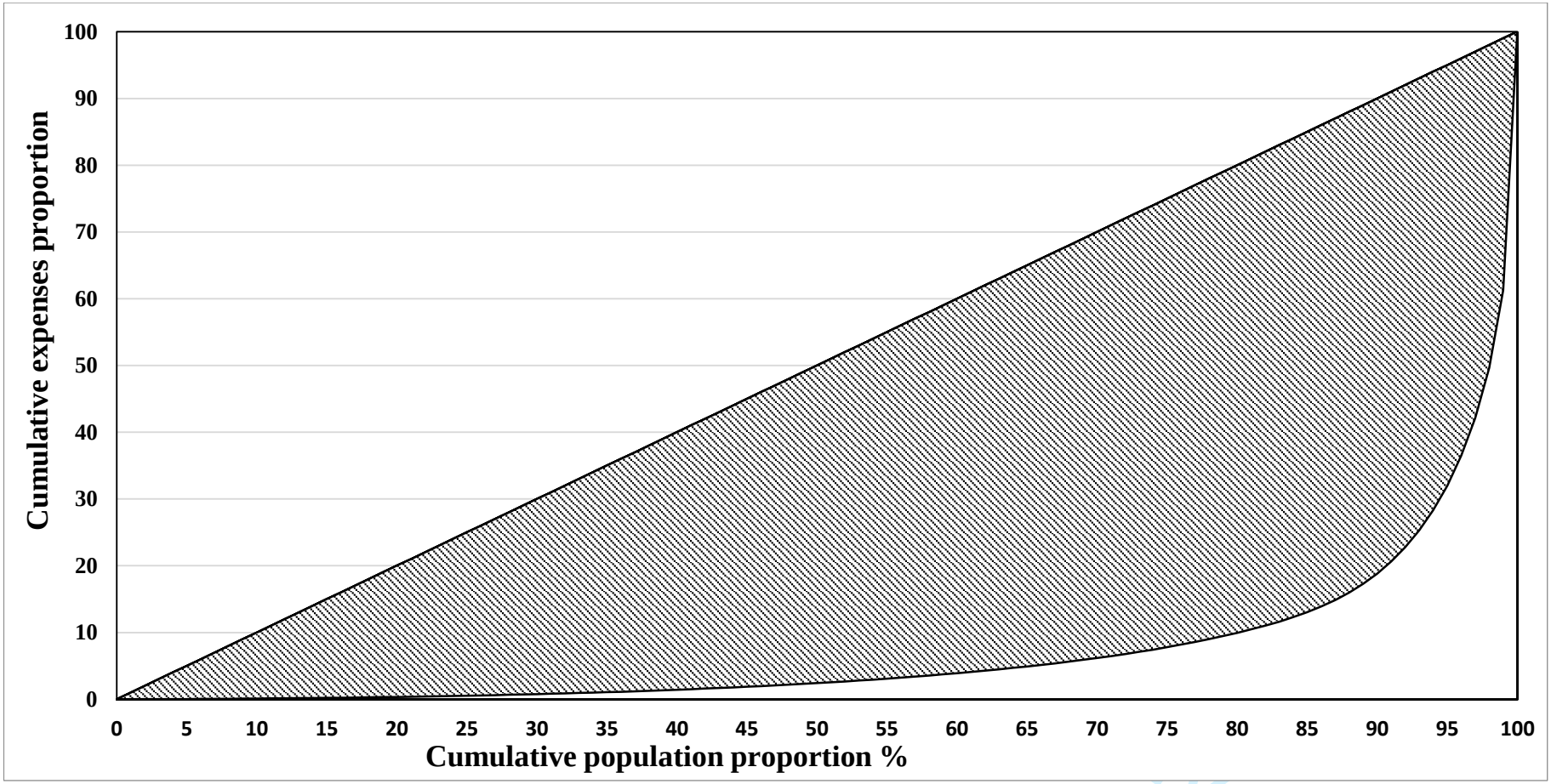
Figure Legend

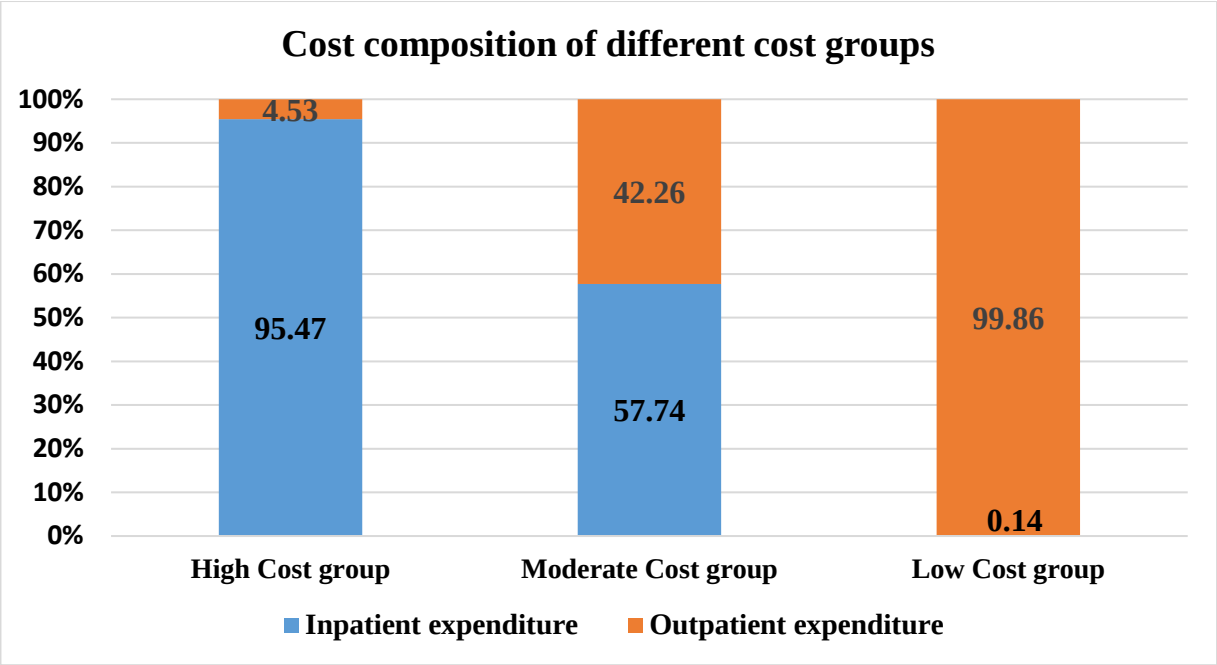
Figure 1. Map of Macheng and geographic distribution of the residents.

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

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




Cumulative expenses proportion
100
90
80
70
60
50
40
30
20
10
0
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Cumulative population proportion %



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:	*BMJ Open*
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

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

Yan Zhang[1,2], Shan Lu[1,2], Yadong Niu[1,2], Liang Zhang[1,2 §]

[1] School of Medicine and Health Management, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, 430030, China

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

Email addresses:

Y.Z.	Yan Zhang	Ph.D.	yanzhang@hust.edu.cn
S.L.	Shan Lu	M.D.	shanlu@hust.edu.cn
Y-D. N.	Yadong Niu	M.D.	nyadong@126.com
L.Z.	Liang Zhang	M.D.	zhangliang@mails.tjmu.edu.cn

§Corresponding author:

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

Word Count: 3276

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

medical expenditures.

Conclusions: The medical expenditures of rural residents are clustered at a remarkably high level and HC patients are suffering from high economic burden. Therefore, policy makers must guide these patients in seeking appropriate healthcare services and improve their management of healthcare quality to reduce the unnecessary healthcare utilisation of these patients.

Keywords: medical expenditure; high-cost; clustering; patient flow; rural China

Strengths and limitations of this study

- This study is the first to introduce the medical expenditure clustering technique, which findings can supplement the results of previous research on HC patients.
- The annual medical expenditures of residents are seldom reported at the population level. This study was conducted in Macheng, a city in Hubei Province with 889,160 residents according to the New Rural Cooperative Medical System (NRCMS) database, which also stores the inpatient and outpatient records of these residents.
- The Lorentz curve and Gini index were used to cluster the annual medical expenditures data and a three-level linear regression model was used to aggregate these data at the residential and town levels.
- The age, gender, hospitalisation, geographic and disease data of the sample were included in the regression model. However, some individual factors were not included in the model.

Background

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

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

reduce their health expenditures, conserve social health insurance funds and promote horizontal equity.

Medical expenditure clustering has also become a major concern in developing countries. In rural China, the rapid development of NRCMS significantly promoted the healthcare utilisation rates of residents. For instance, the annual hospitalisation rate in rural China increased from 8.7% in 2008 to 14.9% in 2017.[1] An empirical analysis of seven counties in China revealed that 78.6% of inpatient services in 2015 were distributed amongst one-third of all inpatients in the area. In addition, one patient in Qianjiang District utilised inpatient services 27 times in 2014. Another study involving 12,600 families in Jiangsu revealed that HC patients accounted for 44.9% of the total medical expenditures of the entire population.[8] Moreover, the medical expenditures clustered at the patient level are much higher than those clustered at the family level. We also suspect that China has a very high degree of medical expenditure clustering and that HC patients incur unnecessary medical expenditures because of the fragmented healthcare delivery system in rural China. Residents in rural China also seek healthcare services from a three-tier (village–town–county) healthcare delivery system where the higher tiers provide better services and charge higher medical costs.[9] Given that patients neither follow a specified order nor have limitations when seeking healthcare services and that general practitioners or consultants are unavailable in most parts of rural China, the residents of rural areas tend to make uninformed decisions when choosing amongst hospitals and various types of healthcare services, thereby leading to their inappropriate utilisation of

healthcare services. For instance, some of these residents, especially HC patients, may be given inpatient services when they actually require outpatient services and may be unnecessarily admitted to higher-level hospitals, thereby incurring higher medical costs.[10] Yingchun revealed that the inappropriate hospital admission rate in five counties reached as high as 27.6% in 2014.[11]

To reduce the economic burden of rural residents in China, several policies and strategies, such as the Tiered Healthcare System and Serious Illness Medical Insurance (2016), have been proposed. However, the effectivity of these programmes has received mixed feedback and the focus of these policies remains unsubstantiated, which may be ascribed to the difficulty in identifying high-demand and HC patients.[12] Waxmonsky J. A. argued that those HC patients with multiple or complex conditions, behavioural disorders or socioeconomic problems are particularly difficult to monitor.[13] Moreover, only few studies have investigated the characteristics and determinants of HC patients by using population-level data because of the lack of necessary data for exploring their annual medical expenditures. Therefore, the critical values, average expenditures and inpatient service utilisation of HC patients in rural China remain unclear.

Previous studies reveal that HC patients often maintain a high level of medical expenditure for the following year. For instance, Robst found that 49.2% of patients under the Florida Medical Assistance Program (Florida Medicaid) were continuously classified as HC patients from 2005 to 2010.[14] Meanwhile, Wodchis et al. found that one-third of residents with public health insurance in Ontario, Canada continuously

incurred high medical expenditures from 2009 to 2011.[15] Therefore, whilst HC patients are assumed to possess certain characteristics, the distribution of their medical expenditures remains unclear.

The growth of health expenditures must be controlled and the efficiency of insurance funds must be enhanced to improve the health insurance system of a specific area. Medical expenditure clustering can be used to monitor the cost efficiency of healthcare services. Based on the above findings, identifying HC patients is a necessary procedure and clustering the medical expenditures of a given population must be taken as the first step towards achieving such goal.

This research focuses on medical expenditure clustering in particular as well as on the distribution and characteristics of HC patients and the determinants of the annual medical expenditures of residents in general. This study also aims to guide policy makers and health planners in predicting and planning the future needs of these patients.

Methods

Study setting

We calculated the annual medical expenditures of residents based on the outpatient and inpatient services that they have availed within a calendar year. A population-based retrospective study was performed in Macheng, a typical rural area in Hubei and county-level city in central China (Fig. 1). Macheng has a total population of 889,160 and a GDP per capita of 22,758¥ (3,704.83$).

[Figure 1]

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

Data processing

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

included in the low cost (LC) group. These three groups represent the various degrees of medical expenditure clustering.

The main programming techniques employed in this paper include Excel functions (e.g. COUNTIF, SUMPRODUCT, LOOKUP and IF) and case processing technologies (e.g. split columns and removal of duplicates). The outstanding diseases of each resident were marked and the original ICD-10 disease codes were adjusted to broader ones (for example, the disease code for COPD was adjusted from J44.900 to J44). Township hospitals were divided into four levels according to their scale and service capacity. The distance from and arrival time to county hospitals were individually captured by using Google Maps. The exchange rate of the US dollar against the RMB in 2014 was 6.1428.

The land form of towns, the healthcare capacity of township hospitals and the sociological characteristics (e.g. gender and age), arrival time to county hospitals, disease categories, healthcare utilisation (e.g. annual length of stay or LOS) and annual medical expenditures of the residents were collected to build the final database.[16]

Statistical analysis

Firstly, the medical expenditures of the residents were clustered by using the Gini coefficient and Lorentz curve. The Gini coefficient is a digitised representation of medical expenditure clustering. A larger Gini coefficient corresponds to a higher degree of medical expenditure clustering. Then, the characteristics of the residents in the HC, MC and LC groups were compared by conducting T test and chi-square test

in the IBM SPSS Statistics 22.0 software. At last, the determinants of annual medical expenditure were then examined by conducting a linear logistic regression analysis. Two key observations were obtained at this stage. Firstly, the obtained data showed a hierarchical structure. Therefore, the determinants of annual medical expenditure were examined by conducting a multilevel linear logistic regression analysis using MLwiN 2.30.[17] Secondly, the annual medical expenditure showed a skewed distribution as expected. Therefore, this variable was transformed to follow a normal distribution. The patient, village and town were assigned to levels 1, 2 and 3, respectively. After accepting the Log10 (x) translation, the following regression model was obtained:

$$\begin{aligned}\mathrm{Log10}(\text{Annual medical expense}) = {} & \beta_{0jk}\mathrm{cons} + \beta_1\mathrm{Gender}_{ijk} + \beta_2\mathrm{Age}_{ijk} + \beta_3\mathrm{Family}_{ijk} + \beta_4\mathrm{Distance}_{ijk} + \beta_5\mathrm{Time}_{ijk} + \beta_6\mathrm{Capacity}_{ijk} \\ & + \beta_7\mathrm{Disease}_{ijk} + \beta_8\mathrm{LOS}_{ijk} + \beta_9\mathrm{Coutinp}_{ijk} + \beta_{10}\mathrm{Outinp}_{ijk} + \beta_{11}\mathrm{Coutinp}_{ijk} + \beta_{12}\mathrm{Towinp}_{ijk} \\ & + \beta_{13}\mathrm{Cououtp}_{ijk} + \beta_{14}\mathrm{Towoutp}_{ijk} + \beta_{15}\mathrm{Clioutp}_{ijk} + \mathrm{More}_{ijk} \\ \beta_{0jk} = {} & \beta_0 + \mathrm{u}_{0j} + \mathrm{w}_{0jk}\end{aligned} \tag{207}$$

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

Patient and public involvement

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

Ethical approval

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

Results

Medical expenditure clustering in the sample area

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

[Figure 2]

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

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

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)

[Figure 3]

Characteristics of HC patients at various clustering levels

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

developed respiratory diseases.

Table 3. Distribution of the demographic characteristics of residents in the three cost groups (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)

*Pearson's chi-square test.

**ANOVA.

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

Table 4. Distribution of the healthcare utilisation characteristics of the three groups (n = 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**

*Pearson's chi-square test.

**ANOVA.

Determinants of the annual medical expenditures of residents

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

Table 5 Three-level linear regression model analysis of annual medical expenditures (n = 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	–	–

Discussion

Clustering of the medical expenditures of residents in rural China

The medical expenditures of rural residents showed an extremely uneven distribution. A Gini coefficient of 0.814 was recorded in 2014, which indicated that the medical expenditures of this population was clustered at the minority level. The HC group

accounted for 68.01% of the total medical expenditures, which was much higher than that recorded in the US (52.3%).[5] Meanwhile, the LC group, which included 70% of the residents in the sample, accounted for 2.42% of the total medical expenditures. The annual medical expenditure per capita of the entire population was 1,222.49¥ (199.01$), which was nearly similar to that of the MC group (1261.36¥, 205.18$). However, this value was only 7.35% of the annual medical expenditure per capita of the HC group. The maximum annual medical expenditure of this population was 424,962.1¥ (69,180.52$), which is more than four times larger than the annual reimbursements provided by NRCMS (100,000¥, 16,279.22$).

The three groups showed obvious differences in their expenditure structures. Hospitalisation expenditures accounted for over 95% of the medical expenditures of the HC group, thereby supporting the findings of Driessen et al.[18] Outpatient expenditures accounted for most of the medical expenditures of the LC group whilst outpatient and hospitalization expenditures equally accounted for the medical expenditures of the MC group. The medical expenditures of the HC group showed a dispersed distribution and a standard deviation that was much higher than the mean (22,817.95¥ vs. 16,618.78¥, 3,714.58$ vs. 2,705.41$). By contrast, the medical expenditures of the MC and LC groups showed a relatively concentrated distribution.

The HC group faces a very high economic burden. In fact, those residents with medical expenditures of over 5,000¥ (813.96$) were included in this group. This group had an annual medical expenditure per capita of over 16,000¥ (2,604.68$), of

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

Aggregating annual medical expenditures at the village and town levels

The multilevel linear regression revealed a hierarchical structure (town–village–residents) in the collected data. The annual medical expenditures of residents were clustered at the town and village levels whilst the annual medical expenditures of residents living in the same village and town tend to be approximated. The distribution of annual medical expenditures as shown in Table 3 significantly differs from that shown in Table 5 in terms of geography, traffic condition and capacity of township hospitals. Such differences were mostly observed at the town level because the residents were living in towns with the same geography, traffic conditions and capacity of township hospitals. Meanwhile, 30 towns in the sample showed differences in their social customs, geographic locations and capacity of township hospitals. These findings altogether show that those residents living in areas with favourable geographic conditions (e.g. plains and national standard roads) and high-capacity hospitals tend to accumulate high medical expenditures. Similarly, the effects of distance from and arrival time to county hospitals were mostly observed at the village level. Given their convenient locations, those residents who are living near county hospitals have a higher probability to avail healthcare services in these institutions and accumulate higher medical expenditures compared with those who are

living far from these hospitals.

Determinants of annual medical expenditures

Annual medical expenditures are directly determined by healthcare utilisation. Table 5 shows that a higher utilisation of healthcare services corresponds to higher medical expenditures. The village, town and county levels had regression coefficients of 0.031, 0.049 and 0.134 for outpatient cases, which were lower than the corresponding coefficients for inpatient cases. Similar to hospital utilisation, an increase in LOS corresponds to an increase in medical expenditures.

In addition, a higher age corresponds to a higher expenditure. This case is particularly true for those residents aged above 60 years, who account for 34.48% of HC group, in other words, 8.5% of the elderly population was defined as HC, while only 5% of total population was defined as HC. Such high expenditures can be attributed to their poor physical condition and high tendency to develop one or multiple diseases. A WHO report revealed that an aging population would increase the healthcare expenditures of a country, but the extent of such increase is less than expected. Although health-related needs often increase along with age, the relationship between healthcare utilisation and health expenditure varies as a person grows older.[19] In fact, the medical expenditures of people from high-income countries gradually decline after they reach the age of 75. Other studies even show that those people aged 80 years and above have a much lower share in the consumption of medical resources compared with the total population. In terms of gender, the average medical expenditures of female residents do not differ from those of male residents. The

relationship between gender and medical expenditures also warrant further study.[20]

Those patients with respiratory, urinary, endocrinology, skeletal and muscular diseases have the same degree of medical expenditure clustering. By contrast, those patients with cancer, circulatory, digestive, obstetrical and gynaecological diseases or haematological disorders have significantly high annual medical expenditures, with cancer patients having the highest regression coefficient of 0.759. Developing diseases that are fatal and difficult to assess can easily lead to high medical expenditures.[21] These findings can be attributed to the limited ability of healthcare professionals at townships, most of whom lack specialisation in treating urinary and cardiovascular diseases or haematological disorders; therefore, those residents suffering from such diseases tend to be admitted to county hospitals or hospitals outside their respective counties.[22]

Redesigning the healthcare delivery system for the HC group

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

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

on several risk factors. Robst et al. and Wodchis et al. found that a patient identified as HC in a year is more than 40% likely to be identified as an HC patient in the following year given that these patients often maintain a high level of medical expenditure for the following year. [14, 15] Therefore, those residents with a remarkably high healthcare utilisation, are exposed to many risk factors and have been identified as HC patients in the previous year warrant special attention. Moreover, HC patients need guidance when choosing healthcare services and must be given priority access to primary healthcare.

Secondly, enhancing the healthcare quality management of patients plays a vital role in discouraging their unnecessary utilisation of healthcare services, such as inappropriate admission to hospitals and excessive utilisation of inpatient services.[23]

Thirdly, exploring new mechanisms, such as the comprehensive management of patients with chronic diseases, setting a global budget for multilevel institutions, integrated management programmes and integrated delivery of medicine and nursing services to aged residents, can motivate doctors to deliver continued care to HC patients. In doing so, healthcare professionals can avoid unnecessary service duplications, improve their healthcare efficacy and help their patients take advantage of primary healthcare services.[24]

Conclusion

From the city-level population perspective, the medical expenditures of rural residents in China have an intense clustering level. Apart from demographic characteristics (e.g.

age and disease), healthcare utilisation was identified as a primary determinant of medical expenditure clustering. Therefore, policy makers must guide HC patients in choosing healthcare services and improve their healthcare quality management to discourage their unnecessary utilisation of healthcare services. Doctors must also be motivated to deliver continued care to this group of patients.

Limitation

This study has two limitations. Firstly, hospitalisation information, geographic factors, referral status and diseases were included in the regression model whilst other individual factors, such as economic status, education and preference, were ignored. Secondly, several studies show that HC residents with a high tendency to develop multiple chronic diseases tend to utilise multidisciplinary and multi-institutional services.[25] However, this work only considered the main diseases of these patients as captured in the NRCMS database. These limitations may affect the stability of the findings and should be examined in further studies.

Competing interests

The authors declare no competing interests.

Authors' contributions

Y.Z. and S.L. participated in the conception, design, analyses and writing of the manuscript. Y-D.N. participated in the data collection and statistical analysis. L.Z. helped draft, review and revise the manuscript. All authors gave their approval to publish this version of the manuscript.

Funding statement

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

Disclosure

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

Provenance and peer review

This work is not commissioned and is externally peer reviewed.

Data sharing statement

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

References:

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

[2] Coughlin TA, Long SK. Health Care Spending and Service Use among High-Cost Medicaid Beneficiaries, 2002-2004. INQUIRY-THE JOURNAL OF HEALTH CARE ORGANIZATION PROVISION AND FINANCING. 2009;**46**(4):405-417 'doi': 10.5034/inquiryjrnl_46.4.405['.

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

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

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

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

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

[8] Miao Y, Qian D, Sandeep S, Ye T, et al. Exploring the characteristics of the high-cost population from the family perspective: a cross-sectional study in Jiangsu Province, China. BMJ open. 2017;**7**(11):e17185 'doi': 10.1136/bmjopen-2017-017185['.

[9] Li X, Lu JP, Hu S, Cheng KK, et al. The primary health-care system in China. LANCET. 2017;**390**(10112):2584-2594ELSEVIER SCIENCE INC.

[10] Yip W, Hsiao W. Harnessing the privatisation of China's fragmented health-care delivery. LANCET. 2014;**384**(9945):805-818 'doi': 10.1016/S0140-6736(14)61120-X['.

[11] Zhang Y, Chen Y, Zhang X, Zhang L. Current level and determinants of inappropriate admissions to township hospitals under the new rural cooperative medical system in China: a cross-sectional study. BMC HEALTH SERVICES RESEARCH. 2014;**14**(649).

[12] Bleich SN. Systematic Review of Programs Treating High-Need and High-Cost People With Multiple Chronic Diseases or Disabilities in the United States, 2008-2014 (vol 12, E197, 2015). PREVENTING CHRONIC DISEASE. 2016;**13** 'doi': 10.5888/pcd13.150275e['.

[13] Waxmonsky JA, Giese AA, McGinnis GF, Reynolds RT, et al. Colorado access' enhanced care management for high-cost, high-need Medicaid members: preliminary outcomes and lessons learned. The Journal of ambulatory care management. 2011;**34**(2):183-191 'doi': 10.1097/JAC.0b013e31820f64be['.

[14] Robst J. Developing Models to Predict Persistent High-Cost Cases in Florida Medicaid. POPULATION HEALTH MANAGEMENT. 2015;**18**(6):467-476 'doi': 10.1089/pop.2014.0174['.

[15] Wodchis WP, Austin PC, Henry DA. A 3-year study of high-cost users of health care. CANADIAN MEDICAL ASSOCIATION JOURNAL. 2016;**188**(3):182-188 'doi': 10.1503/cmaj.150064['.

[16] Herrin J, St Andre J, Kenward K, Joshi MS, Audet AJ, Hines SC. Community Factors and Hospital Readmission Rates. HEALTH SERVICES RESEARCH. 2015;**50**(1):20-39.

[17] Longford NT. Improved approximations for multilevel models with binary responses. JOURNAL OF THE ROYAL STATISTICAL SOCIETY SERIES A-STATISTICS IN SOCIETY. 1997;**160**(3):593.

[18] Driessen RJB, Bisschops LA, Adang EMM, Evers AW, van de Kerkhof PCM, de Jong EMGJ. The economic impact of high-need psoriasis in daily clinical practice before and after the introduction of biologics. BRITISH JOURNAL OF DERMATOLOGY. 2010;**162**(6):1324-1329 'doi': 10.1111/j.1365-2133.2010.09693.x['.

[19] Beard JR, Officer AM, Cassels AK. The World Report on Ageing and Health. GERONTOLOGIST. 2016;**562**:S163-S166 'doi': 10.1093/geront/gnw037['.

[20] Owens GM. Gender differences in health care expenditures, resource utilization, and quality of care. JOURNAL OF MANAGED CARE PHARMACY. 2008;**143**(3):S2-S6.

[21] Ross JS, Chen J, Lin Z, Bueno H, et al. Recent National Trends in Readmission Rates After Heart Failure Hospitalization. CIRCULATION-HEART FAILURE. 2010;**3**(1):97-103.

[22] Hartl S, Luis Lopez-Campos J, Pozo-Rodriguez F, Castro-Acosta A, et al. Risk of death and readmission of hospital-admitted COPD exacerbations: European COPD Audit. EUROPEAN RESPIRATORY JOURNAL. 2016;**47**(1):113-121.

[23] Feder JL. Predictive Modeling And Team Care For High-Need Patients At HealthCare Partners. HEALTH AFFAIRS. 2011;**30**(3):416-418 'doi': 10.1377/hlthaff.2011.0080['.

[24] Ye T, Sun XW, Tang WX, Miao Y, Zhang YD, Zhang L. Effect of continuity of care on health-related quality of life in adult patients with hypertension: a cohort study in China. BMC

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

Figure Legend

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

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

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

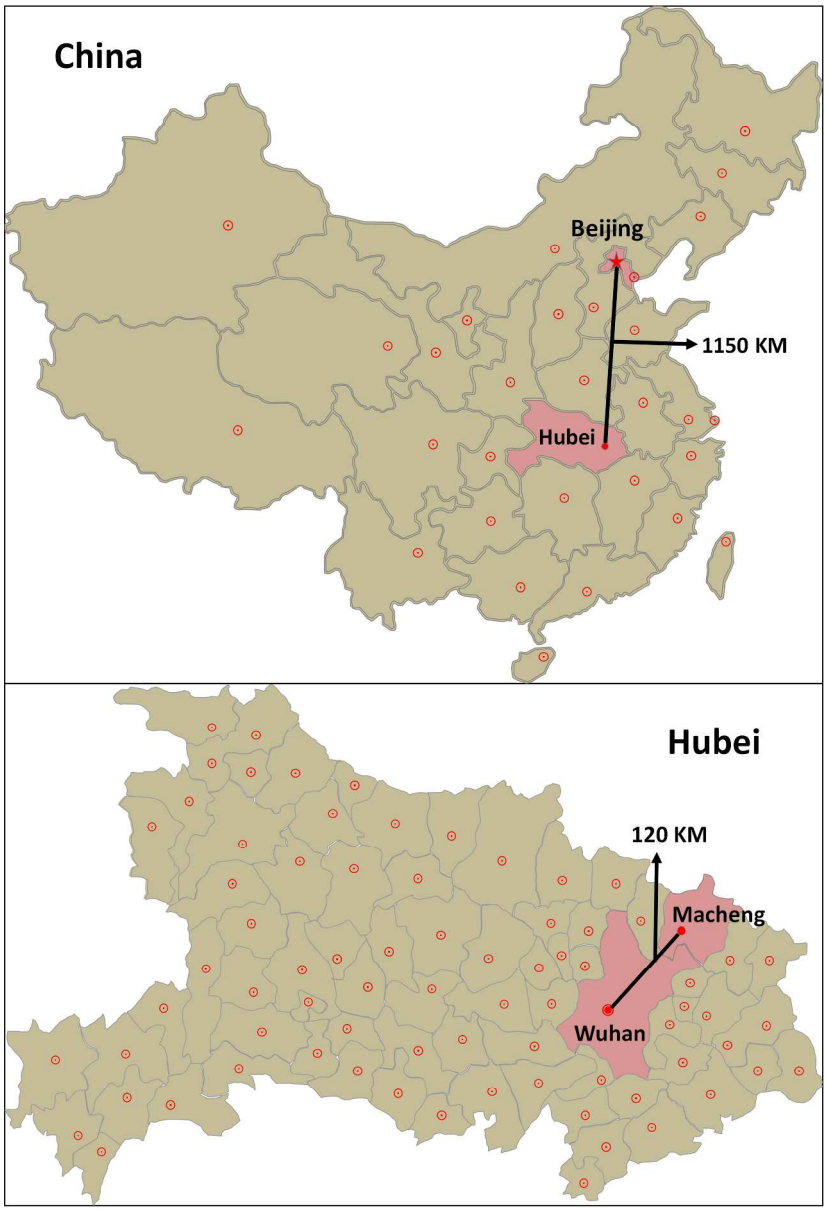


Figure 1. Map of Macheng City and geographic distribution of the residents.

331x486mm (300 x 300 DPI)

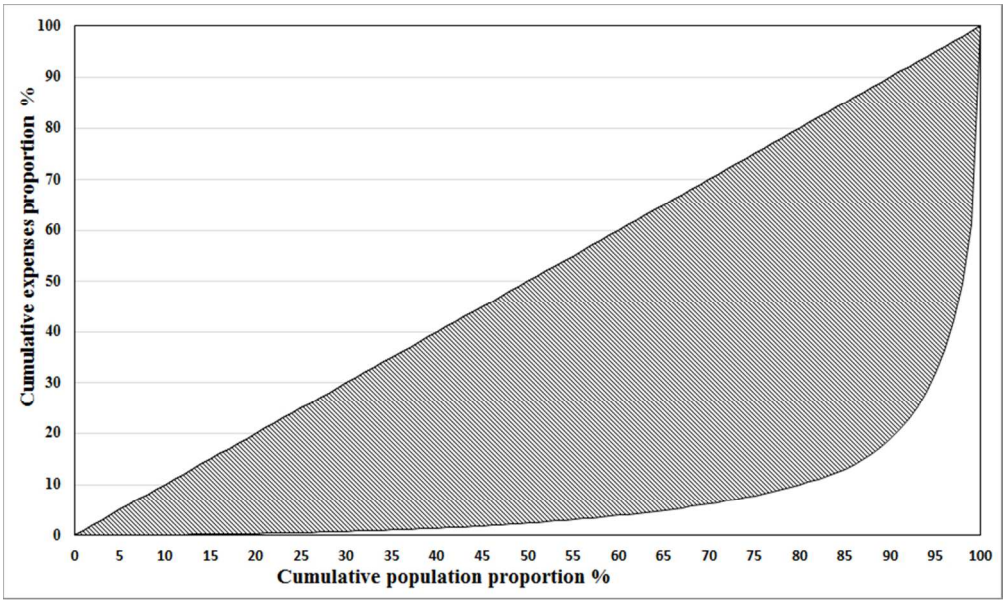


Figure 2. Lorentz curve of the medical expenditure clustering for Macheng in 2014.

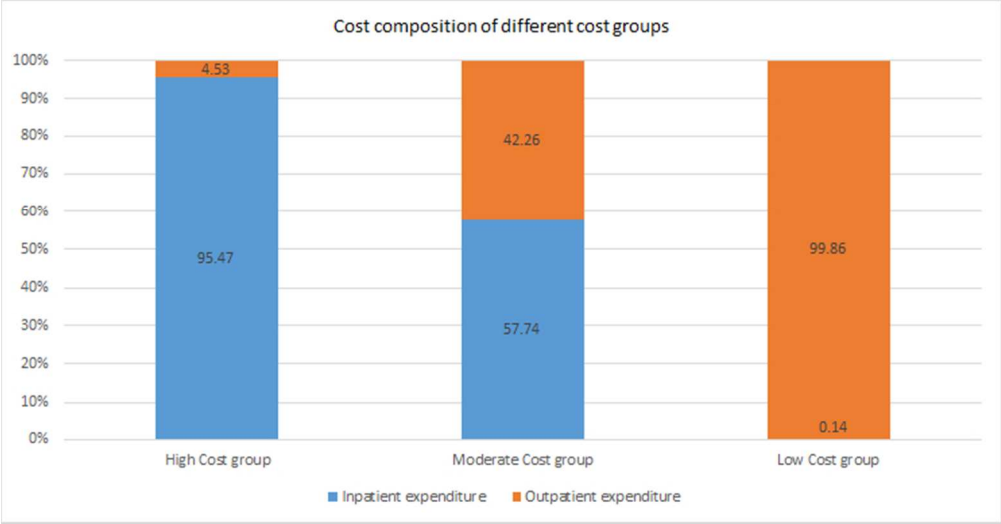


Figure 3. Cost composition of the three groups of rural residents in Macheng in 2014.

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.