

## Appendix A Measurement—Quality

We developed two sets of quality measures: 1) Non-disease specific measure, and 2) disease specific measures, based on a review of international practices and consultation with domestic experts.

“Non-disease specific measures” include: infection rate of intestinal incision (laparotomy), readmission rate for all hospital admissions, readmission rate at 30 days for patients with cholecystectomy, readmission rate at 30 days for patients with appendectomy, proportion of fracture surgery for patients who underwent long bone open fracture within 24 hours of admission, and hospital-acquired pneumonia. These measures use information from hospital discharge data and hospital medical records, hospital infection records and details of New Rural Cooperative Medical Scheme (NCMS) inpatient claims data. Table 1 summarizes the data sources of the non-disease specific measures and the formulas used to calculate them.

Table 1. Data sources and formulas of non-disease specific measures

Indicators	Data Sources	Formulas
Infection rate of the intestinal incision (laparotomy)	Infection records; hospital discharge data	No. of cases with infection of intestinal incision/No. of cases with intestinal laparotomy
Readmission rate of all inpatients	NCMS inpatient claims	No. of readmission cases within 30 days after discharge (including those readmitted to other hospitals)/No. of all inpatients
Readmission rate at 30 days in patients with cholecystectomy	NCMS inpatient claims	No. of readmission cholecystectomy cases within 30 days after discharge/No. of cases with cholecystectomy
Readmission rate at 30 days in patients with appendectomy	NCMS inpatient claims	No. of readmission cholecystectomy cases within 30 days after discharge/No. of cases with appendectomy
Proportion of fracture surgery on patients with long bone open fracture within 24 hours after admission	Hospital discharge	No. of fracture cases given surgery within 24 hours after admission/No. of cases with long bone open fracture
Hospital-acquired pneumonia	Infection records; hospital discharge data	No. of cases diagnosed with hospital acquired pneumonia/No. of cases treated within the statistical period

We initially included hospital’s adherence to a surgical safety checklist (Gawande et al. 1999; Kable et al. 2002) as another measure. However, our pilots showed that all hospitals used surgical checklist but varied substantially in how complete and accurate their data entries were. Due to budget limitation, we excluded that measure.

For “disease-specific measures”, we examined hospital admission data from Guizhou and identified the following diseases to focus on: Pneumonia, chronic obstructive pulmonary disease (COPD), ischemic cerebral infarction, and acute myocardial infarction (AMI). Using the US

Hospital Quality Alliance framework (HQA) and UK Advancing Quality framework (see Appendix 4A and 4B) as bases, we selected a subset to use as quality indicators. Table 2 presents the disease-specific measures, obtained using randomly sampled medical records.

Table 2. Disease specific measures

Diseases	Indicators (yes/no)
Pneumonia	Assessment of oxygenation index Effective dose of sputum culture prior to initial antibiotic treatment First antibiotic treatment within 6 hours after admission Influenza vaccination Pneumococcus vaccination Adult suggestion/advice on quitting smoking
Chronic obstructive pulmonary disease	Assessment of oxygenation index Influenza vaccination Pneumococcus vaccination Adult suggestion/advice on quitting smoking
Ischemic cerebral infarction	Aspirin treatment within 24 hours after admission Aspirin treatment at discharge Statins treatment at discharge Adult suggestion/advice on quitting smoking
Acute myocardial infarction	Aspirin treatment at admission Aspirin treatment at discharge $\beta$ -blocker treatment at discharge Adult suggestion/advice on quitting smoking

The process for quality assessment is as follow:

- 1) **Data Preparation:** Medical records and hospital infection records of discharged patients were obtained from hospitals in 2015 and 2016, and details of NCMS inpatients claims were obtained from local NCMS office for the same time period.
- 2) **Sampling:** For each of the four conditions – pneumonia, COPD, ischemic cerebral infarction, and AMI, 40 cases were randomly drawn from hospital medical records using the case ID. Then, hospitals scanned and submitted electronic copies of the selected medical records.
- 3) **Review Case History:** Clinical experts with at least ten years of clinical experience reviewed all “non-disease specific measures” and all “disease-specific Measures” and extracted relevant information from the review process. .

## Appendix B County Household Survey

The two surveys were conducted in 2015 and 2018 respectively. The survey in 2015 collects the information of hospitals in 2014, and the survey in 2018 collects the information of hospitals in 2017.

We match the data from the two surveys in 2015 and 2018 and include the hospitals that had been surveyed in both two periods. Of 75 hospitals (75\*2 records) are included in the analysis.

Table 1 shows the descriptive statistics of the hospital's general information, staffing, financial statement, governance, leadership, and provider payment method. From 2014 to 2017, there were no obviously changes in the general information of hospitals. For example, most of them are the general hospital. 80% of hospitals are secondary. About 70% of hospitals are the public hospital. In addition, we can see from Table 1 that the staffing, financial statement, governance and leadership have changed obviously from 2014 to 2017. For example, the number of staffs, bed and medical equipment of hospitals in 2017 is higher than in 2014.

**Table 1 Descriptive statistics**

Variables	2014	2017	Total
<b>Hospital's general information:</b>			
Is the hospital a pilot hospital in the national public hospital reform pilot?			
Yes	34 (45.33)	52 (69.33)	86 (57.33)
No	36 (48.00)	21 (28.00)	57 (38.00)
Missing	5 (6.67)	2 (2.67)	7 (4.67)
Type of hospital			
General hospital	52 (69.33)	51 (68.00)	103 (68.67)
Traditional Chinese hospital	23 (30.67)	23 (30.67)	46 (30.67)
Others (specify)	0 (0)	1 (1.33)	1 (0.67)
Level of hospital			
Primary	9 (12.00)	6 (8.00)	15 (10.00)
Secondary	60 (80.00)	60 (80.00)	120 (80.00)

	1	4	5
Tertiary	(1.33)	(5.33)	(3.33)
Unassigned	5	5	10
	(6.67)	(6.67)	(6.67)
Ownership of hospital			
Public (MOH)	53	53	106
	(70.67)	(70.67)	(70.67)
Public (non-MOH)	5	4	9
	(6.67)	(5.33)	(6.00)
Non-public (specify)	17	18	35
	(22.67)	(24.00)	(23.33)
Profit status			
For profit	12	13	25
	(16.00)	(17.33)	(16.67)
Not-for-profit	63	62	125
	(84.00)	(82.67)	(83.33)
<b>Staffing</b> <sup>x(f)</sup> :			
Official quota on number of posts that receive government subsidies	248.60	246.30	247.40
	(191.40)	(177.80)	(183.80)
Number of staffs on duty	272.50	369.60	321.10
	(191)	(264.40)	(235)
Number of medical professionals	221.40	313.30	267.30
	(161.50)	(230.80)	(203.80)
Number of retired staffs	43.29	43.60	43.45
	(46.62)	(49.72)	(48.08)
In the past 3 years, number of newly recruited medical practitioners	9.69	13.10	11.39
	(12.97)	(17.24)	(15.3)
In the past 3 years, number of newly recruited assistant medical practitioners	6.32	2.59	4.45
	(15.47)	(5.40)	(11.66)
Official quota on number of beds	221.50	261.40	241.30
	(155.30)	(178.70)	(168)
Number of beds in operation	278.40	352.80	315.6
	(189)	(248.90)	(223.4)
Number of ICU beds	4.25	5.07	4.68
	(5.06)	(5.67)	(5.39)
Total value of medical equipment with value greater than 10,000 RMB	2,076	4,643	3,351
	(1,949)	(7,433)	(5,549)
Number of medical equipment with value greater than 10,000 RMB	137.40	366.90	249.80
	(159.70)	(943.70)	(677.40)
<b>Financial statement</b> <sup>x(f)</sup> :			

Total revenue	5,714 (5,035)	10,247 (10,022)	7,981 (8,225)
Government subsidy	863.60 (792.10)	1,721 (1,590)	1,311 (1,339)
Revenue from medical service	4,916 (4,407)	7,441 (6,929)	6,187 (5,932)
Revenue from outpatient services	1,364 (1,234)	2,258 (2,146)	1,817 (1,807)
Revenue from inpatient services	3,571 (3,281)	5,183 (4,906)	4,388 (4,248)
Revenue from insurance	3,226 (4,617)	4,015 (3,898)	3,626 (4,272)
Total expense	5,737 (5,136)	8,978 (7,959)	7,379 (6,888)
In past year, total surplus as share of total medical service revenue	9.83 (9.86)	21.92 (109.20)	16.26 (79.93)
<b>Governance of the hospital:</b>			
How many departments are there in your hospital excluding administrative and logistics departments <sup>*(S)</sup>	15.56 (7.05)	17.48 (8.94)	16.52 (8.08)
Do the departments in your hospital have separate financial accounts?			
Yes	32 (42.67)	38 (50.67)	70 (46.67)
No	27 (36.00)	28 (37.33)	55 (36.67)
Partially independent	11 (14.67)	9 (12.00)	20 (13.33)
Missing	5 (6.67)	0 (0.00)	5 (3.33)
Has your hospital implemented director-in-charge management system?			
Yes	70 (93.33)	73 (97.33)	143 (95.33)
No	4 (5.33)	2 (2.67)	6 (4.00)
Missing	1 (1.33)	0 (0.00)	1 (0.67)
How to appoint the director of your hospital?			
Direct appointment by the government	64	69	133

	(85.33)	(92.00)	(88.67)
Open Competition	9	5	14
	(12.00)	(6.67)	(9.33)
Missing	2	1	3
	(2.67)	(1.33)	(2.00)
Is there a set of assessment criteria on appointing directors?			
Yes	53	53	106
	(70.67)	(70.67)	(70.67)
No	16	22	38
	(21.33)	(29.33)	(25.33)
Missing	6	0	6
	(8.00)	(0.00)	(4.00)
Does your hospital have discretion in recruiting staffs holding officially budget posts?			
Yes	10	9	19
	(13.33)	(12.00)	(12.67)
No	50	52	102
	(66.67)	(69.33)	(68.00)
Not applied	14	14	28
	(18.67)	(18.67)	(18.67)
Missing	1	0	1
	(1.33)	(0.00)	(0.67)
Does your hospital have discretion in recruiting staffs not holding officially budget posts?			
Yes	67	65	132
	(89.33)	(86.67)	(88.00)
No	2	3	5
	(2.67)	(4.00)	(3.33)
Not applied	6	7	13
	(8.00)	(9.33)	(8.67)
Does your hospital have discretion in designing annual development plan?			
Yes	74	72	146
	(98.67)	(96.00)	(97.33)
No	1	3	4
	(1.33)	(4.00)	(2.67)

---

<b>Leadership of the hospital:</b>			
Does your hospital face competition from other hospitals?			
Intense	50 (66.67)	41 (54.67)	91 (60.67)
Some	23 (30.67)	32 (42.67)	55 (36.67)
No	2 (2.67)	2 (2.67)	4 (2.67)
Overall, does your hospital use centralized or decentralized management approach?			
Centralized management mode at the hospital level	44 (58.67)	47 (62.67)	91 (61.67)
Decentralized management to department level	4 (5.33)	0 (0.00)	4 (2.67)
Combination of both of the above	27 (36.00)	27 (36.00)	54 (36.00)
Missing	0	1 (1.33)	1 (0.67)
<b>Provider payment method:</b>			
Is your hospital contracted with New Cooperative Medical Scheme?			
Yes	75 (100.00)	74 (98.67)	149 (99.33)
No	0 (0.00)	1 (1.33)	1 (0.67)
Is your hospital contracted with Urban Resident Basic Medical Insurance?			
No	69 (92.00)	69 (92.00)	138 (92.00)
Yes	4 (5.33)	6 (8.00)	10 (6.67)
Missing	2 (2.67)	0 (0.00)	2 (1.33)
Is your hospital contracted with Urban Employee Basic Medical Insurance?			
Yes	71 (94.67)	72 (96.00)	143 (95.33)
No	2 (2.67)	3 (4.00)	5 (3.33)

---

Missing	2 (2.67)	0 (0.00)	2 (1.33)
Total	75	75	150

Note: (1) The two surveys were conducted in 2015 and 2018 respectively. The survey in 2015 collects the information of the hospital in 2014, and the survey in 2018 collects the information of the hospital in 2017. (2)  $\bar{X}(S)$ : Mean (Standard deviation). (3) The unit of costs is Yuan. (4) Unless otherwise indicated, data are expressed as weighted number (percentage) of column totals for each group. Estimated counts were rounded to the nearest unit, and thus, totals across categories may differ from the calculated sums.

Table 2, Table 3 and Table 4 show the results of DID regression. Before the DID regression, we tested the basic characteristics of the treatment and control group in the baseline, such as the hospital's level, type, ownership, profit status, medical revenue and etc. We find that there is no significant difference in the basic characteristics of the two groups. We think that the prerequisite of DID, that is, the common trend is valid.

The basic DID model is set as follows:

$$Y_{ict} = \beta_0 + \beta_1 Time + \beta_2 Treatment + \delta Time * Treatment + \beta_3 Pairwise + \varepsilon_{ict}$$

Where  $i$  is the hospital subscript;  $c$  is the county subscript;  $t$  is the time subscript;  $Y_{ihm}$  is the explained variable, which denotes the hospital's staffing, financial statement, governance, leadership, and provider payment method;  $Time$  is a dummy variable for the survey year, equal to 1 when time=2018;  $Treatment$  is a dummy variable for group membership, equal to 1 when the hospital implements the payment reform.  $Pairwise$  is a set of dummies of pair groups.  $\delta$  is the coefficients of interest.  $\varepsilon$  is the disturbance term.

All analyses are conducted using Stata 14.0.  $P < 0.10$  is used to determine statistical significance.

Table 2 shows the effects of payment reform on the hospital's staffing. We select some indicators reflecting hospital staffing as dependent variables. Firstly, we estimate the DID model without controlling other covariates and find that the hospital's staffing is not influenced by the implementation of payment reform.

In addition, the basic characteristics of hospitals, including hospital's level, type, ownership, profit status, and whether the pilot of public hospital reform, are controlled in the DID regression model as covariates. We find that the coefficient for the variable of MEV<sup>1</sup> is 2,304.988 and significant in 5% level, suggesting that the implementation of payment reform treatment would increase the MEV of hospitals. Since the medical equipment usually associate with the diagnostic services, we can infer that the payment reform would improve the diagnostic ability of hospitals.

<sup>1</sup> MEV: Total value of medical equipment with value greater than 10,000 RMB.

**Table 2 The effects of payment reform on hospital's staffing**

Variable	OQNP	NSD	NMP	NRS	NRMP	NRAMP	OQNB	NBO	NIB	MEV	NME
DID model without control variables											
Time*Treatment	-38.342 (47.506)	-3.311 (32.489)	-15.484 (9.594)	-1.983 (6.337)	-3.246 (4.679)	4.478 (3.748)	0.962 (54.181)	6.980 (38.244)	-0.302 (1.364)	2,765.684 (2,139.058)	431.002 (356.460)
DID model that controls covariables											
Time*Treatment	-48.471 (71.250)	18.792 (21.842)	6.199 (20.057)	-3.616 (6.396)	-2.434 (6.253)	6.169 (4.464)	4.246 (50.968)	17.746 (47.354)	1.031 (1.484)	2,304.988** (820.699)	411.580 (312.999)

Note: (1) covariables include as following: hospital level, type, ownership, profit status, and whether the pilot of public hospital reform. (2) Cluster standard errors in city are in parentheses. (3) \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . (4) Abbreviation list: OQNP: Official quota on number of posts that receive government subsidies for basic salary; NSD: Number of staffs on duty; NMP: Number of medical professionals; NRS: Number of retired staffs; NRMP: In the past 3 years, number of newly recruited medical practitioners; NRAMP: In the past 3 years, number of newly recruited assistant medical practitioners; OQNB: Official quota on number of beds; NBO: Number of beds in operation; NIB: Number of ICU beds; MEV: Total value of medical equipment with value greater than 10,000 RMB; NME: Number of medical equipment with value greater than 10,000 R

Table 3 displays the effects of payment reform on the hospital's financial statement. We also choose some indicators reflecting the hospital's financial statement. We used a similar model and methods as the hospital's staffing to estimate the effects of payment reform competition on the hospital's financial statement. The coefficient for the variable of total expense and STMSR2 are significant. The results suggest that hospital's total expense was increased after the implementation of payment reform. At the same time, the total surplus as share of total medical service revenue was decreased after the implementation of payment reform. The results suggest that this reform would improve the utilization rate of hospital's revenue.

**Table 3 The effects of payment reform on the hospital's financial statement**

Variable	Total revenue	Government subsidy	Revenue from medical services	Revenue from outpatient services	Revenue from inpatient services	Revenue from insurance	Total expense	STMSR
DID model without control variables								
Time*Treatment	3,095.378 (18,09.429)	411.975 (517.424)	1,129.003 (696.011)	206.884 (145.739)	378.801 (694.868)	898.492 (856.743)	2,551.774** (831.110)	-7.632** (2.543)
DID model that controls covariables								
Time*Treatment	3,818.027 (2,244.401)	422.137 (495.495)	1,685.925 (1,242.848)	251.817 (308.336)	803.394 (1,120.225)	738.008 (918.949)	3,076.699** (1,102.186)	-9.088* (3.609)

Note: (1) covariables include as following: hospital level, type, ownership, profit status, and whether the pilot of public hospital reform. (2) Cluster standard errors in city are in parentheses. (3) \* p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01. (4) Abbreviation list: STMSR: In the past year, total surplus as share of total medical service revenue.

Table 4 list the effects of payment reform on the hospital's governance, leadership and provider payment method. The method and model are similar to above. We find that the implementation of payment method does not affect the hospital's governance, leadership, and provider payment method.

<sup>2</sup> STMSR: In the past year, total surplus as share of total medical service revenue.

**Table 4 The effects of payment reform on hospital's governance, leadership and provider payment method**

Variable	ND	SFA	DICMS	AD	ACAD	RSBP	RSNBP	DADP	Competition	MA	NCMS	URBMI	UEBMI
DID model without control variables													
Time*Treatment	1.773 (2.070)	-0.177 (0.228)	-0.074 (0.082)	-0.005 (0.079)	0.195 (0.159)	0.023 (0.105)	-0.096 (0.150)	0.032 (0.065)	0.121 (0.205)	-0.009 (0.343)	-0.032 (0.035)	0.007 (0.061)	-0.032 (0.036)
DID model that controls covariables													
Time*Treatment	1.980 (2.933)	-0.159 (0.249)	-0.116 (0.111)	-0.019 (0.891)	0.117 (0.140)	0.008 (0.103)	-0.089 (0.164)	-0.059 (0.056)	0.029 (0.264)	-0.021 (0.379)	-0.039 (0.044)	-0.081 (0.460)	-0.040 (0.045)

(1) covariables include as following: hospital level, type, ownership, profit status, and whether the pilot of public hospital reform. (2) Cluster standard errors in city are in parentheses. (3) \* p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01. (4) Abbreviation list: ND: How many departments are there in your hospital excluding administrative and logistics departments; SFA: Do the departments in your hospital have separate financial accounts? DICMS: Has your hospital implemented director-in-charge management system? AD: How to appoint the director of your hospital? ACAD: Is there a set of assessment criteria on appointing directors? RSBP: Does your hospital have discretion in recruiting staffs holding officially budget posts? RSNBP: Does your hospital have discretion in recruiting staffs not holding officially budget posts? DADP: Does your hospital have discretion in designing annual development plan? Competition: Does your hospital face competition from other hospitals? MA: Overall, does your hospital use centralized or decentralized management approach? NCMS: Is your hospital contracted with New Cooperative Medical Scheme? URBMI: Is your hospital contracted with Urban Resident Basic Medical Insurance? UEBMI: Is your hospital contracted with Urban Employee Basic Medical Insurance?

## **Appendix C World Management Survey**

Good management is regarded as a key instrument to achieve good outcomes in various kinds of organizations, including hospitals. But measuring hospital management is never an easy job. Despite theories put managerial ability at the heart of understanding the heterogeneity of organization's productivity ([Kaldor 1934](#); [Woodward 1958](#)), hospital management measurement, however, has been held back by a lack of appropriate data. The data is particularly scant in low- and middle-income countries (LMICs). Little is known about how to improve hospital management effectively and efficiently, even though hospital management has been attached importance to in recent years in these countries. Using the World Management Survey (WMS) instrument to assess the management practices of county-level hospitals in Guizhou Province in western China, this study aims to fill this gap.

### ***Management matters in Healthcare***

Healthcare systems are under severe pressure due to aging populations, rising costs of medical technologies, budget austerity and increasing patient expectations. One potential way to tackle these cost pressures is through improving hospital performance, i.e. efficiency and quality. Management practices are regarded as a means to upgrade the performance ([Lega et al. 2013](#)) as evidence has shown that the management practices of hospitals were highly correlated with the admissions, process-of-care and 30-day mortality for acute myocardial infarction patients ([Bloom et al. 2014b](#); [McConnell et al. 2013](#); [McConnell et al. 2016](#)). It was found a positive link between improved management and higher survival rates for emergency surgery, less number on waiting list, reduced expenditure per patient and lower staff turnover in hospitals in the UK ([Bloom et al. 2015](#)). Evidence from the US and the UK showed that a one-standard-deviation increase in hospital management score was associated with a 20% increase in the probability of being a high-quality hospital ([Tsai et al. 2015](#)). In the US, financial performance was also significantly better in hospitals with higher management scores ([Bloom et al. 2014b](#)). Thus, it is necessary to assess accurately and try to improve hospital management practices since they are vital to healthcare outcomes.

### ***Hospital management measurement in China***

Like other LMICs, China does not have much knowledge or experience in hospital management measurement. Back to 1989, when China introduced hospital accreditation system, management among different domains, for example patient safety management or financial management, was inspected as one aspect of the accreditation criteria to determine hospital levels. In 2005, the former Ministry of Health issued the "the hospital management assessment guidelines" and

revised it in 2008 to help hospitals improve their management practices. The guidelines made general suggestions for hospitals of different types and levels to achieve good practices in overall organizational management, quality improvement, hospital safety, and service management. Recently, there is a trend towards big hospitals applying “Joint Commission International” (JCI) accreditation to prove their excellence in clinical quality and safety. More than 80 hospitals nationwide have applied and passed the accreditation until the end of 2017. All the three practices mentioned above include somewhat hospital management assessment, but they are more of review of working standards and focusing on building blocks. Therefore, China has few empirical scientific evidence of quantitative measurement on hospital management which shows variations between hospitals and has clear links to management fundamentals: setting targets, establishing incentives, and monitoring performance ([Bloom et al. 2014c](#)).

### ***China’s hospital reform and the links to hospital management***

As 40% of outpatient and 77% of inpatient services take place in hospitals (*China’s annual statistical yearbook for health and family planning, 2017*), reform of its hospitals has been a priority for China’s national health reform, especially in recent years. The key areas of hospital reform include altering *public* hospitals by revamping hospital governance, improving hospital management, strengthening monitoring on hospital performances, and changing provider payment methods (PPMs) ([He 2011](#)). Among them, improving internal management is not only one of the main levers of performance change, but the improvement of which can make other reform initiatives more effective. In fact, other reform innovations such as capitation, DRGs, or team-based care require organizational change, rather than individual practice change, which challenges hospital managerial capabilities ([Bohmer 2010](#)). The public hospital reform was designed to be carried out step-by-step, and it started with county hospitals because they made up the largest proportion of public hospitals (approximately 40% of all public hospitals) and bore 16.9% of the first contact outpatient and 51.6% of total inpatient services nationally (*China Center for Health Statistics and Information, 2015*). As the chief reform measures, they were designed to act as gatekeepers for the residents within the county for most common diseases and the head of the medical treatment partnership system to align downwards with township health center and village clinics, which demanded an increased level of managerial expertise. Another strategy of hospital reform in China was to attract *private* capital to invest in hospitals. The rationale behind it was to introduce competition among public and private hospitals and to bring in innovative management through entrepreneurship. Thus, there is a great need to robustly evaluate hospital management practices in China, especially county-level hospitals as a first step, to establish a baseline to target the poor aspects within management to shed light on effective policy interventions and for further reform evaluation.

## ***Global practices of measuring hospital management***

Worldwide, management data used to be collected in closed-ended questions ([Bloom & Van Reenen 2010a](#)). Lately, there has been a new development in empirical economics of management: the World Management Survey (WMS) tool, designed to elicit more accurate measures using open-ended questions approach with independent assessors (analysts) scoring organizational management practices ([Bloom et al. 2016](#); [Bloom et al. 2010](#); [Bloom & Van Reenen 2007](#); [Bloom & Van Reenen 2010a](#)). The WMS, initially applied to the manufacturing industry has since been adapted to over 2000 hospitals in seven high-income countries (HICs) and two LMICs ([Bloom et al. 2014a](#); [Bloom et al. 2014b](#)). It was found that poor management prevailed, and that hospitals were even more poorly managed than manufacturing companies. They also documented considerable variation in hospital management practices both between and within countries.

Our research hereby used the WMS tool to measure hospital management in Chinese county-level hospitals, to add knowledge of hospital management measurement in LMICs and to understand in detail what particular aspects of management matter the most to provide insights into how to improve hospital management quality in China.

## **Methods**

### ***Study site***

We set our study site in Guizhou province. Guizhou is a low-income province in the south-west part of China with a large rural population of 20 million, but rich in natural, cultural and environmental resources. The ethnic composition is 64% Han and 36% of more than 17 minority groups. With a per capita annual income of 26,743 CNY (4,179 USD) in urban areas and 8,090 CNY (1,264 USD) in rural areas, Guizhou was the seventh poorest out of 32 provinces in the mainland of China in 2016. The infant mortality rate was 7.9 ‰ and the maternal mortality rate was 27.3/100 000 (Guizhou Statistical Yearbook, 2017). Guizhou province has 88 county level units under nine municipalities (sub-provincial level). Within each of the counties, there is at least one public general hospital, in some places an additional traditional Chinese hospital, and a few specialty public hospitals. The private sector predominates in number in Guizhou; on average each county has six private general or specialty hospitals. But most private hospitals are small-scaled and low-level, so public hospitals still dominate in services provision.

## ***Sampling***

Our sample randomly covered 58 counties under eight out of nine municipalities. In each of the counties, we tried to include all the public non-specialty hospitals, i.e. county general hospitals and county traditional Chinese hospitals. We also sampled the private non-specialty hospitals above a certain size based on number of beds, market share of volume or revenue. Finally, we had a list of 146 hospitals, 106 public (100% included) and 40 private (11% of the total 379 private hospitals in the eight municipalities).

In each sampled hospital, the survey was completed by one director and one charge nurse from the Department of Orthopedics, or the Department of Cardiology or the Department of Surgery in order (based on their existence). In accordance with the WMS methodology, this level of middle managers was purposely selected because they were senior enough to have an overview of management practices but not so senior as to be detached from day-to-day operations. These three departments were also purposely selected; they were relatively more likely to be standardized and procedural than other more contingent departments.

## ***Measurement instrument***

In order to measure management practices across hospitals, we drew on an expanded evaluation tool based on the original WMS Questionnaire ([Bloom & Van Reenen 2007](#)). The original WMS provided metrics to measure hospital management practices along four broad dimensions: (i) standardizing operations; (ii) monitoring performance; (iii) setting targets; and (iv) incentivizing personnel. A set of 20 basic management practices were evaluated and scored from one (“worst practice”) to five (“best practice”) on a pre-defined scoring grid, in increments of one point ([Bloom et al. 2016](#)).

The research team found that the distributions in developing countries had large tails at the lower end of the distribution, so the original WMS tool missed some important variation at the left tail. To address this issue, Lemos and Scur developed an expanded scoring grid which allowed for systematic inclusion of half points in the scale, effectively extending it without compromising the backwards comparability of the “Development” WMS (D-WMS) with the “original” WMS ([Lemos & Scur 2016](#)).

Compared with the original one, the D-WMS further divided each management practice into three key processes: (i) process implementation (formulating, adopting and putting into effect management practices); (ii) process usage (carrying out and using management practices frequently and efficiently); (iii) process monitoring (monitoring the appropriateness and efficient use of management practices). Therefore, the D-WMS tool increased the number of

scores from 20 in the original WMS to 60 and allowed us to understand what part of the process was driving the results. ([Lemos & Scur 2016](#)).

Our research team translated the English questionnaire and scoring grid into Chinese and adapted it into the Chinese context.

### ***Interviewers and interview process***

To collect the data, we hired and trained a team of 40 public health and health management teachers and students from the two chief medical universities in Guizhou (and so could interview managers in their native languages), who had medical or management knowledge to conduct the interviews. Our interviewers were extensively and intensively trained on hospital management, what information he/she was seeking and how to probe for further information. Calibration checks were conducted during the training and the survey as well.

Following the WMS “double-blind” and “double-scored” methodology<sup>1</sup>, each survey was conducted face-to-face on one manager by two interviewers ([Bloom et al. 2016](#); [Lemos & Scur 2016](#)). The two interviewers would then discuss their individual scores to correct for any misinterpretation of responses. We mixed pairs of interviewers as much as possible throughout the survey, conditional on geographic limitations.

Additionally, we used a variety of procedures to obtain a high success rate and to remove potential sources of bias from our estimates, including obtaining government endorsements, avoiding asking interviewees for performance or financial data, and collecting characteristics of the interviewee (e.g. department) and the identity of the interviewer to include in the regressions to improve the precision of our estimates.

### ***Analysis strategy***

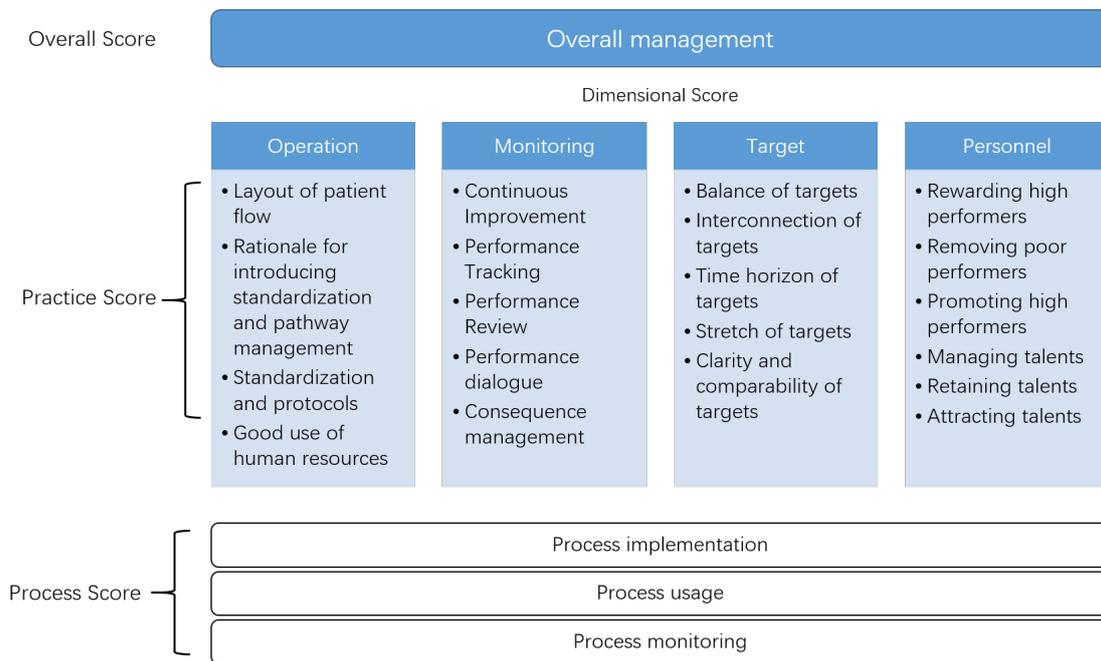
First, we calculated the overall and dimensional scores. As we interviewed each interviewee about information referring to all three processes for each of the 20 practices, we first took a simple average of these three to build a single score for each of the practices for each interviewee. As in each hospital, we interviewed two managers, so we then averaged out their scores for each practice to assign to each hospital. Based on the individual hospital score data, we created indices for

---

1 “Double-blind” means that managers were not told in advance they were being scored or shown the scoring grid while the interviewers were not told in advance about the organization’s performance. “Double-scored” refers to that the first interviewer was accompanied by a second interviewer whose main role was to monitoring the quality of the interview being conducted by taking notes and separately scoring the responses after the interviews had ended.

overall management (average of all 20 practices), and dimensional management –operation, monitoring, target, and personnel management.

Then, we looked at the scores in each process – implementation, usage, monitoring. To build these, we omitted the above first step of averaging across the three processes for each practice and re-organized the dataset into three new sets of 20 practices along the lines of each process. We took the score of each of the 60 processes and built average indices for overall management, every dimensional management for each of the process types. Figure 1 demonstrated the four dimensions, the practices included in each dimension as well as the processes within each practice, each dimension or across the board.



Note: The process score could be calculated for overall management, each dimension and each practice.

**Figure 1 conceptual model of D-WMS**

To enable the results to be comparable to those of other countries assessed by the original WMS, we converted the scores of D-WMS, by coding down the half-point scores. For example, 1.5 for the D-WMS to 1 for the original WMS ([Lemos & Scur 2016](#)).

Finally, we also compared the overall, dimensional management scores as well as process scores by different hospital ownerships.

## Results

### *Sample characteristics*

Our sample finally covered 139 hospitals (response rate 95.21%) in 58 counties in Guizhou. Approximately 71% of the sampled hospitals were general hospitals while the others were traditional Chinese ones. More than 81 % were secondary hospitals, and 76% were public hospitals. On average, each hospital had 285 employees and 275 operating beds, with an annual revenue of 61 million CNY (9.5 million USD). Almost all (91%) of the hospital directors had medicine backgrounds, and 50% of them had never been involved in any management training (Table 1).

**Table 1 Characteristics of sample hospitals and interviewees**

	N	Mean	S.D.
<b>Sampled hospitals</b>			
Hospital type			
General hospitals	139	0.676	0.470
Traditional Chinese hospitals	139	0.324	0.470
Hospital level			
Primary	139	0.173	0.379
Secondary	139	0.813	0.391
Undefined	139	0.014	0.120
Public hospitals	139	0.755	0.431
Number of staffs			
Number of medical staffs	137	226	164
Number of beds in operation			
Number of ICU beds	123	3.829	4.832
Annual revenue (million RMB)	136	60.937	53.085
Annual expenditure (million RMB)	135	59.169	51.610
Medicine background of director	103	0.913	0.284
Management training of director	104	0.500	0.502
<b>Sampled interviewees</b>			
Age	254	39.398	8.601
Gender	273	0.473	0.500
Duration in hospital (years)	266	14.240	9.461
Duration in positon (years)	252	4.565	4.341
Management training	257	0.128	0.335
Departments			
Orthopedics	273	0.604	0.490
Cardiology	273	0.055	0.228
Surgery	273	0.234	0.424
Others	273	0.106	0.309

Job titles				
Department director	273	0.465	0.500	
Charge nurse	273	0.498	0.501	
Vice director	273	0.026	0.158	
Others	273	0.011	0.104	
Background				
Medicine	260	0.562	0.497	
Nursing	260	0.408	0.492	
Others	260	0.031	0.173	
Education level				
High school and below	259	0.100	0.301	
College	259	0.328	0.470	
Bachelor	259	0.525	0.500	
Master	259	0.039	0.193	
Others	259	0.008	0.088	

Note: Hospital characteristics were obtained from institution surveys in 139 county-level hospitals in 8 municipalities in Guizhou. Interviewees data were reported by 273 interviewed managers in these hospitals.

We interviewed 273 managers from the 139 hospitals. Among them, 60% were from Department of Orthopedics, 23% from Surgery and 5% from Cardiology. Most of the interviewees were charge nurses (50%) or department directors (47%). An average manager was 39-years-old, working in his/her position for 4.56 years and nearly 14 years in the hospital. Nearly all (95%) held at least bachelor's clinical degree, over 97% from medicine or nursing specialty. Only a few (12.8%) had had any formal management training (Table 1).

### ***Overall score***

Our results (Figure 2) showed that the overall management practices (D-WMS score) of the Guizhou sample scored 2.57 (S.D. = 0.462). It could be transferred to the internationally comparable score (WMS score) as 2.43 (Table 2). We compared our survey data with the existing international management practice for seven HICs, i.e. Canada, France, Italy, Germany, Sweden, U.S and U.K. (in 2009); two other LMICs, i.e. India (2012) and Brazil (2013). Guizhou hospitals generally ranked the seventh, significantly worse managed than the average and most of the HICs except France (Table 2), but better than each of the two other LMICs (Figure 3). When looking at the distribution of scores, we did observe a tighter distribution of poorly managed hospitals than well-managed ones, when compared our sample with the better-off countries like the US or the UK, though it was nowhere near as dire as the differences in distribution in Brazil and India (Figure 4).

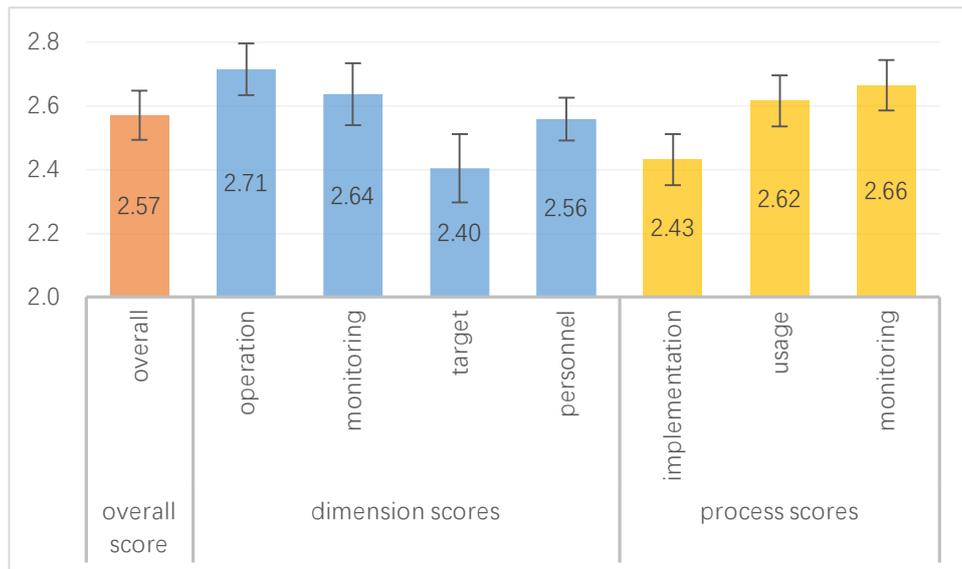
**Table 2 International comparisons on hospital overall and dimensional management scores (comparable)**

	Overall	Operation	Monitoring	Target	Personnel	Generalized operations (overall excl. personnel)
<b>US</b>	3.00	3.03	3.21	2.87	2.92	3.04
<b>UK</b>	2.69	2.91	2.99	2.55	2.37	2.81
<b>SW</b>	2.68	2.52	2.99	2.75	2.46	2.77
<b>GE</b>	2.64	2.78	2.85	2.55	2.45	2.72
<b>CA</b>	2.52	2.78	2.82	2.44	2.17	2.67
<b>IT</b>	2.48	2.85	2.67	2.33	2.20	2.60
<b>Guizhou (China)</b>	2.43 (7)	2.56 (7)	2.52 (8)	2.27 (8)	2.40 (4)	2.44 (8)
<b>FR</b>	2.40	2.87	2.59	2.29	2.03	2.56
<b>BR</b>	2.19	2.38	2.47	1.99	1.98	2.27
<b>IN</b>	1.90	2.11	2.03	1.55	1.93	1.88

Note: Other samples except Guizhou (China) were surveyed and scored using WMS and Guizhou data are converted to comparable scores. The rank of the Guizhou (China) is in the parentheses.

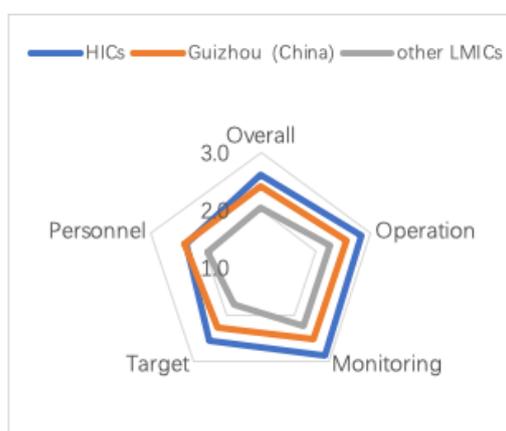
### ***Dimensional scores***

When looking at the four dimensions of management practices, Guizhou performed better in operation and monitoring than target and personnel by scores, similarly to other country samples. Conversely, Guizhou performed significantly better in personnel than target, a pattern only seen in the US and India (Table 2), meaning that Guizhou's worst practices, on average, were in the target dimension. Besides, the target management held the largest variation among the four dimensions, while personnel the smallest (Figure 2).



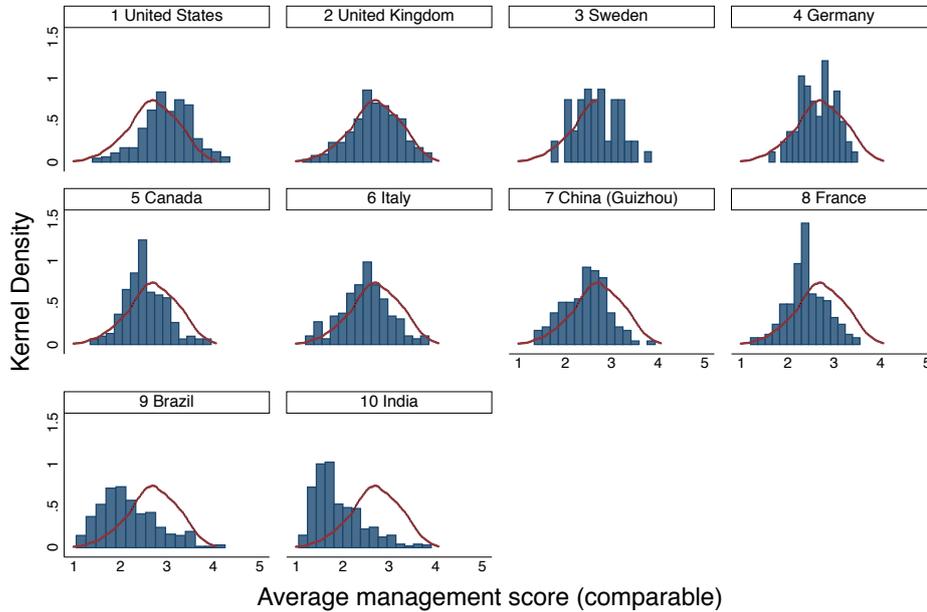
**Figure 2 Average overall, dimension and process management scores (with 95% confidence interval)**

Meanwhile, Guizhou surpassed each of the two other LMICs in all four dimensions. Our sample lagged behind the average of HICs in the first three, but overtook in the personnel dimension (Figure 3). The personnel management of Guizhou scored better than France, Italy, Canada and even UK, and was the best among the four dimensions by ranks. If we excluded the personnel dimension from the overall score, which reflected the generalized operations management ([Lemos & Scur 2016](#)), China then fell one rank down, after France (Table 2).



**Figure 3 International comparisons on hospital overall and dimensional management scores (comparable)**

Note: Other samples except Guizhou (China) are surveyed and scored using WMS and Guizhou data are converted to comparable scores.



Note: China (Guizhou) data is for Guizhou province only, pilot data. The rest of the countries are random samples from the countries and come from the full WMS international dataset.

Bars are the histogram of firms in each country. Line is the smoothed U.S. density, shown for the comparison to the U.S.

**Figure 4 Distribution of management across countries**

The Guizhou hospitals worked best in operation management, which referred to Lean operations in hospitals (description of each practice is in Table 3). The sample performed best in presenting the good rationale for introducing operational changes but worst in implementing standardization of protocols and clinical processes. Within performance monitoring, the sampled hospitals did well in consequence management, especially for managers quickly identified and started to solve problems, but poor in performance tracking, review and dialogue (Table 3), three practices of which were highly correlated. As in the worst target dimension, hospitals performed particularly poor in setting clarified and comparable goals using both qualitative and quantitative measures. Within personnel management, on average, the hospitals were good at managing talent; nevertheless, the weakest practice was to retain talent. In fact, retaining talent was the worst among all the 20 practices. It seemed that the managers did very little to ensure top clinicians wanted to stay or to retrieve good staff when he/she

wanted to leave. Table 3 showed the detailed scores and ranks of practices within each dimension.

Dimensions and rank within the dimension	Management Practice	Description	Scores			
			Mean	S.D.	Min	Max
operation-1	Rationale for introducing standardization and pathway management	Tests the motivation/impetus behind changes to operation, what change story was communicated	2.83	0.56	1.00	4.00
operation-2	Layout of Patient Flow	Tests how well the patient pathway is configured and improved	2.72	0.47	1.50	4.08
operation-3	Good use of human resources	Tests whether staff are deployed to do what they are best qualified for	2.71	0.72	1.00	4.33
operation-4	Standardization and protocols	Tests if there are standardized procedures that are applied and monitored systematically	2.60	0.73	1.00	4.67
monitoring-1	Consequence Management	Tests whether differing levels of performance (NOT personal but plan/ process based) lead to different consequences	3.00	0.81	1.00	4.83
monitoring-2	Continuous Improvement	Tests processes for and attitudes towards continuous improvement, and whether learnings are captured and documented	2.71	0.73	1.00	4.00
monitoring-3	Performance Review	Tests whether performance is reviewed with appropriate frequency and communicated to staff	2.55	0.76	1.00	4.00
monitoring-4	Performance Tracking	Tests whether performance is tracked using meaningful metrics and with appropriate regularity	2.50	0.62	1.00	4.33
monitoring-5	Performance Dialogue	Tests the quality of review conversations	2.42	0.77	1.00	4.50
target-1	Stretch of Targets	Tests whether targets are appropriately difficult to achieve	2.57	0.71	1.00	3.83
targes-2	Interconnection of Targets	Tests whether targets are tied to hospital objectives and	2.57	0.76	1.00	4.67

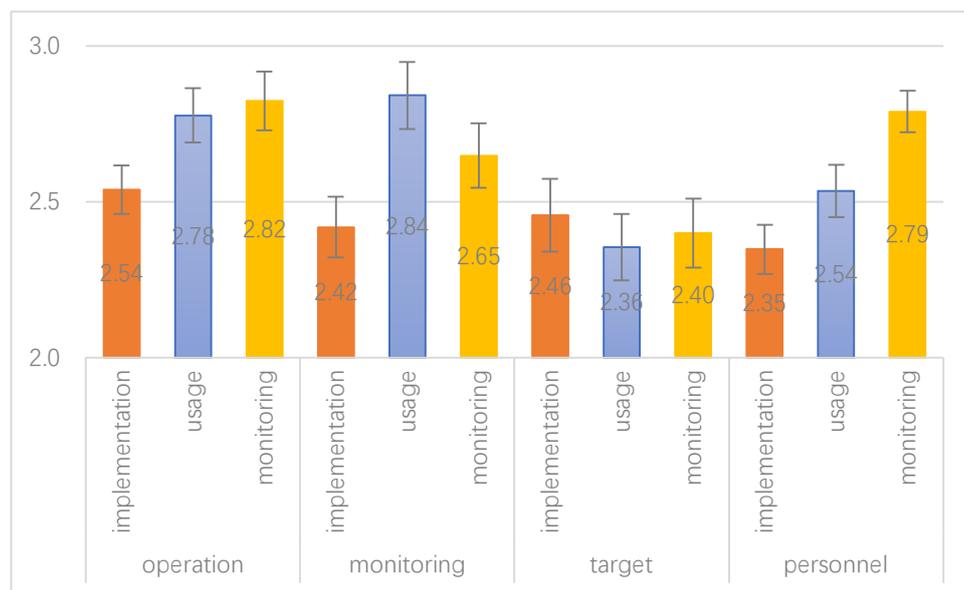
		how well they cascade down the organization				
target-3	Balance of Targets Metrics	Tests whether targets cover a sufficiently broad set of metrics	2.46	0.73	1.00	4.25
target-4	Time Horizon of Targets	Tests whether hospital has a '3 horizons' approach to planning and targets	2.45	0.82	1.00	4.33
target-5	Clarity and Comparability of Targets	Tests how easily understandable performance measures are and whether performance is openly communicated	1.98	0.65	1.00	3.42
personnel-1	Managing Talents	Tests what emphasis is put on talent management	3.03	0.58	1.67	5.00
personnel-2	Removing Poor Performers	Tests whether hospital is able to deal with underperformers	2.93	0.61	1.17	4.50
personnel-3	Promoting High Performers	Tests whether promotion is performance based	2.73	0.54	1.08	4.00
personnel-4	Rewarding High Performers	Tests whether good performance is rewarded proportionately	2.58	0.70	1.00	4.17
personnel-5	Attracting talents	Tests the strength of the employee value proposition	2.23	0.46	1.17	3.17
personnel-6	Retaining talents	Tests whether hospital will go out of its way to keep its top talent	1.85	0.39	1.00	3.33



### Scores of Processes

In the D-WMS, three key processes – implementation, usage, and monitoring – were defined to measure the strength and emphasis of each practice, as well as systematically across all the practices, within an organization. Hospitals in Guizhou generally performed slightly better in process monitoring (2.66) than process usage (2.62) while worst in process implementation (2.43). All the three processes had similar variations (Figure 2).

We could also score the variation across the practices within each dimension. In target management, hospitals performed similarly poor in all the three processes. Within the second poorest personnel dimension, hospitals scored lowest in implementation. In fact, the worst implementation across the dimensions was seen in personnel dimension (Figure 5).



**Figure 5 Hospital management scores of overall and dimensional processes**

We presented the correlation matrix for processes in Table 4 and noticed that all correlations were positive (above 0.5) and significant at the 1% level. All three processes were highly correlated across the broad, and in operation, monitoring and target dimensions, with the coefficients above 0.8 (Table 4).

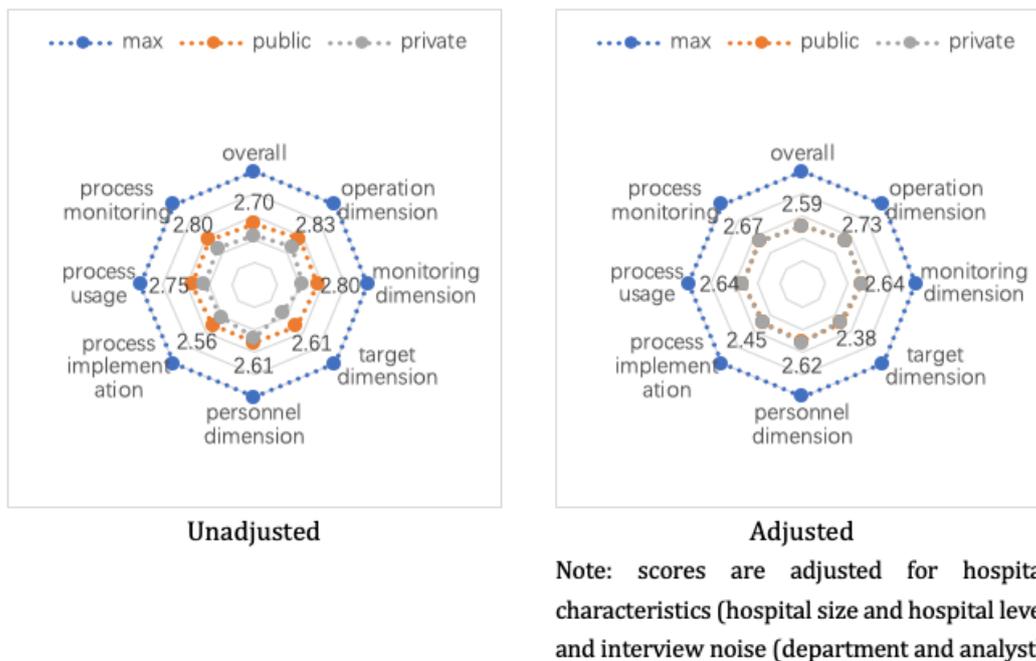
**Table 4 Correlations of overall and dimensional processes**

		implementation	usage	monitoring
<b>Overall</b>	implementation	1.00		
	Usage	0.90	1.00	
	monitoring	0.89	0.94	1.00

<b>Operation</b>	implementation	1.00		
	Usage	0.80	1.00	
	monitoring	0.81	0.80	1.00
<b>Monitoring</b>	implementation	1.00		
	Usage	0.84	1.00	
	monitoring	0.84	0.88	1.00
<b>Target</b>	implementation	1.00		
	Usage	0.87	1.00	
	monitoring	0.91	0.90	1.00
<b>Personnel</b>	implementation	1.00		
	Usage	0.68	1.00	
	monitoring	0.51	0.69	1.00

### ***Comparison between public and private hospitals***

We then compared the management quality between the public and private hospitals. Figure 6 showed that the public hospitals sample performed better than the private sample overall, and in all dimensional and process scores ( $p < 0.005$ ). Nevertheless, when controlling for the hospital characteristics (hospital size and hospital level) and interview noise (interviewee's department and analyst identity), there was no significant difference between public and private hospitals in any of the domains.



**Figure 6 Overall, dimensional and process scores between public hospitals and private hospitals, unadjusted and adjusted data**

## Discussion

By interviewing 273 department heads from 139 county-level hospitals (105 public and 34 private) in Guizhou via the D-WMS instrument, we were the first<sup>2</sup> to use an internationally validated survey tool to measure carefully the management practices in county-level hospitals in China. Our study added evidences to the insufficient measurement of hospital management practices in LMICs, enabled us to compare China data to others and detect the bottlenecks to shed light on effective policy interventions: The sampled county-level hospitals in western China ranked the seventh overall among ten country samples – after most of the HICs except France and before the only other two LMICs (Brazil and India); they ranked lower in three management dimensions (operation, monitoring, and target) except personnel management, but surpassed the other LMICs in all of the domains; they comparatively performed worst in target management and process implementation.

Our sample only included county-level hospitals in western China and they might not be generalizable to national samples as in other countries. They were smaller in size and lower in level (average employment was 285, no teaching hospitals), compared to the intentional samples (average employment was from 558 to 2344, 9%~42% were teaching hospitals). From previous studies, we understood management scores were positively associated with hospital scale and size ([Bloom et al. 2017a](#); [Bloom et al. 2014b](#)) and levels of development ([Lemos & Scur 2013](#)).

Target management was the weakness of hospitals in Chia and in other LMICs as well. Among the four management dimensions, we observed the sample performed noticeably worse in the targeting - similarly badly in implementation, usage and monitoring on average, and the variations among hospitals were particularly large. The same trend was found in India, and not only in hospitals but retails and schools as well ([Lemos & Scur 2013](#)). Brazil hospitals performed likewise poor in target setting, if not the poorest. A previous review showed there was a dearth of empirical work on hospital level priority setting practices and more so in smaller, rural hospitals in developing country contexts ([Barasa et al. 2014](#)).

Our interviews also confirmed this. Quite a few hospitals surveyed stated their

---

<sup>2</sup> There are other efforts in China to measure hospital management recently, to introduce the Chinese hospital management survey (CHMS) which brought in WMS concepts, but different in questions, calibration and training.

goals as “being the best or one of the best in the area”, “better save lives and serve people” or “achieve progress in technical capacity”. These goals were ambiguous and difficult to break down into explicit individual goals. Even with some measurable targets, they were usually related directly to government-imposed targets or set as reform goals that must be filled, especially for public hospitals. However, the targets were rarely set based on internal factors that can reflect realistic improvements on previous years’ outcomes, and it was easy to result in misunderstanding or complaint of the targets or motivation to manipulate or forge rather than real improvement. What was even worse was that among the current goals mentioned most, quality and patient outcomes were not the priority. Moreover, the goals in a hospital were usually short- or mid-term rather than long-term. Plus, the goals of different time-spans were often set independently which make short-term targets little chance to become a “staircase” to reach mid-term or long-term goal.

As setting targets is part of the fundamental management function of planning, we argue that county-level hospitals in China need first deal with the target management, in which more emphasis be put on making goals of different time spans clarified and measurable, balancing all the clinical, efficiency, financial and operational goals, based upon both external and internal factors, and tricking down the hospital goals into department and individual’s targets. As setting and tracing performance metrics for public hospitals by the government was a notable measure to be carried out in the Chinese hospital reform, it is valuable to study how the targets were set and how hospitals improve target management within organizations when facing the policy change.

The sample scored second poorest in personnel management. Hospitals scored lowest in “retaining talent practice”, “attracting talent” and “rewarding high performers”. This was suggested by a survey covering 86 county hospitals that 9392 physicians, 76.6% of whom were young, flew out from these hospitals during last 4 years and more than 70% of them went to upper level hospitals or hospitals in more developed areas (Chinese Hospital Association, 2014). As county-level hospitals act as the crucial role in the Chinese health delivery system, especially the leading role in rural area and meanwhile talents are critical to an organization, we strongly suggest China take active and effective actions in improving managerial expertise in attracting, retaining and rewarding the staff at the county-level hospitals.

Using the D-WMS, we were able to see what processes led to average scores. From the results, we understood the gaps from best practices mainly remained in process implementation across the board, which meant a range of management practices haven’t been adopted or have been adopted very reactively. This gap was particularly so in personnel management. As interviewees acknowledged

that they barely tried to understand the reasons for staff leaving or there was seldom formal and clear set of criteria to distinguish good performers, we confirmed that poor scores were received because hospitals did not put some basic management practices into effect. As the three processes were highly correlated, we suggested first strengthening the implementation of management practices by educating and incentivizing the managers to implement effective management practices and tools, particularly in personnel management dimension.

But one surprising finding was Guizhou ranked fourth and performed relatively better in this personnel management, compared to other country samples. In fact, seven out of ten countries performed worst in this dimension. The sampled county-level hospitals comparatively showed best practice of “managing talent” and “removing poor performers”. Currently there are two parallel talent-recruiting and managing systems inside Chinese public hospitals – one is controlled strictly by the local human resource authority to decide the type and number of staff the hospitals need – the other is led by hospitals themselves to absorb staff they want as a supplement since the scale of talents recruited through the first system is shrunk. Hospitals also have more autonomy in dismissing and disposing staffs within the second system. The survey did not distinguish these two systems and might overestimate the autonomy and capability of personnel recruiting and discharging.

Besides the focus on the bottlenecks of management, how best to motivate management improvement overall? Previous studies showed ownership structures, competition, education, governance, autonomy and institutions seem to be important drivers of the quality of management practices ([Bloom et al. 2012](#); [Bloom et al. 2014b](#); [Dorgan et al. 2010](#); [Lemos & Scur 2013](#)).

Private hospitals usually achieved better management than public hospitals across the countries, same pattern as in industry ([Bloom et al. 2014b](#); [Bloom & Van Reenen 2010b](#); [Lemos & Scur 2013](#); [Pillay 2008](#)). It was interesting to see the result in our study inconsistent as public hospitals performed significantly better than private ones in all domains and no significantly different when controlling hospital characteristics and interview noises. It illustrated that the majority of private hospitals in western counties in China were small-scaled and low-level, and did not have more advanced management than public hospitals. Though the sample size of private hospitals in our survey was small (34), the result raised the concern about the effectiveness of introducing small private hospitals on strengthening competition and boosting innovative management if we could not attract and incentivize high-grade private hospitals in county areas .

From the survey data, we discovered the hospital middle-level managers are all

clinical background and with scarce training of management. Only a half of the hospital directors had Management training experiences. Although it was approved that having physicians in leadership positions was valuable for hospital performance ([Dorgan et al. 2010](#); [Goodall 2011](#)), evidence also suggested that the bundle of managerial and clinical skills had an positive impact on hospital management quality and clinical outcomes ([Bloom et al. 2017b](#)). Given that management and leadership competence was complex to included knowledge, skills, attitudes and abilities while the interplay between management education and policy reform was recognized ([Frenk et al. 2010](#); [Kebede et al. 2009](#)), we argued the potential of strengthening the hospital management practices by intensifying trainings on modern management ideas and tools among physician managers, and encouraging dual degree or elective courses of management in the medical schools in the longer term.

As researchers argued that health care reform might be impeded if underestimating the nature and extent of the managerial work needed to make it happen ([Bohmer 2010](#)), others suggested accountability and governance played an important role for better management ([Bloom et al. 2014b](#)). Now that China has been initiating a comprehensive hospital reform which includes innovations and improvements in governance, autonomy, accountability, management and incentives, it has many opportunities for hospitals to accelerate management changes, interact with other policy changes, and thus affect reform outcomes. We need to pay close attention to these changes and concern their interactions in policy design and reform evaluations.

Last but not the least, though the correlations between WMS measurement and hospital outcomes were internationally validated, it is still worthwhile linking the measurement data with the performance data to empirically test the most useful practices in the Chinese context.

## **Conclusion**

While improving hospital management is a promising approach to increase health care performance, there is very little empirical evidence on how to measure management practices of hospitals and what efforts need to be made to improve them effectively. Using the WMS instrument, this research critically measured the management of county-level hospitals in Western China and sought for the particularly weak aspects that could be improved. We found that quality of management was still low in county-level hospitals in China. While target management was the worst dimension, implementation was the poorest management process and particularly in personnel management. Trainings on modern management among the clinical managers need to be intensified. A major research effort is needed to empirically examine how improvement in

management is effective in improving service efficiency and quality in LMICs, and how it interacts with other health reform measures such as governance, accountability or competition in affecting service outcomes.

## Reference

- Barasa EW, Molyneux S, English M, Cleary S. 2014. Setting healthcare priorities in hospitals: a review of empirical studies. *Health policy and planning*, **30**: 386-396.
- Bloom N, Brynjolfsson E, Foster L, et al. 2017a. What drives differences in management? National Bureau of Economic Research.
- Bloom N, Genakos C, Sadun R, Van Reenen J. 2012. Management practices across firms and countries. *The Academy of Management Perspectives*, **26**: 12-33.
- Bloom N, Lemos R, Sadun R, Scur D, Van Reenen J. 2014a. JEEA-FBBVA Lecture 2013: The new empirical economics of management. *Journal of the European Economic Association*, **12**: 835-876.
- Bloom N, Lemos R, Sadun R, Scur D, Van Reenen J. 2016. International Data on Measuring Management Practices. *American Economic Review*, **106**: 152-156.
- Bloom N, Lemos R, Sadun R, Van Reenen J. 2017b. Healthy Business? Managerial Education and Management in Healthcare. National Bureau of Economic Research.
- Bloom N, Propper C, Seiler S, Van Reenen J. 2015. The impact of competition on management quality: evidence from public hospitals. *The Review of Economic Studies*, **82**: 457-489.
- Bloom N, Sadun R, Van Reenen J. 2010. Recent Advances in the Empirics of Organizational Economics. *Annual Review of Economics*, **2**: 105-137.
- Bloom N, Sadun R, Van Reenen J. 2014b. Does management matter in healthcare. *Boston, MA: Center for Economic Performance and Harvard Business School*.
- Bloom N, Sadun R, Van Reenen J. 2014c. Does Management Really Work? How three essential practices can address even the most complex global problems. *WINNING IDEAS*: 36.
- Bloom N, Van Reenen J. 2007. Measuring and explaining management practices across firms and countries. *The Quarterly Journal of Economics*, **122**: 1351-1408.
- Bloom N, Van Reenen J. 2010a. New Approaches to Surveying Organizations. *American Economic Review*, **100**: 105-109.
- Bloom N, Van Reenen J. 2010b. Why do management practices differ across firms and countries? *The Journal of Economic Perspectives*, **24**: 203-224.
- Bohmer RM. 2010. Managing the new primary care: the new skills that will be needed. *Health Affairs*, **29**: 1010-1014.
- Dorgan S, Layton D, Bloom N, et al. 2010. Management in healthcare: why good practice really matters. *London: McKinsey and Company/London School of Economics*.
- Frenk J, Chen L, Bhutta ZA, et al. 2010. Health professionals for a new century: transforming education to strengthen health systems in an interdependent world. *The lancet*, **376**: 1923-1958.
- Goodall AH. 2011. Physician-leaders and hospital performance: is there an association? *Social science & medicine*, **73**: 535-539.
- He JA. 2011. China's ongoing public hospital reform: initiatives, constraints and prospect. *Journal*

- of Asian Public Policy*, **4**: 342-349.
- Kaldor N. 1934. The Equilibrium of the Firm. *The Economic Journal*, **44**: 60-76.
- Kebede S, Abebe Y, Wolde M, et al. 2009. Educating leaders in hospital management: a new model in Sub-Saharan Africa. *International Journal for Quality in Health Care*, **22**: 39-43.
- Lega F, Prenestini A, Spurgeon P. 2013. Is management essential to improving the performance and sustainability of health care systems and organizations? A systematic review and a roadmap for future studies. *Value in Health*, **16**: S46-S51.
- Lemos R, Scur D. 2013. Could poor management be holding back development? Describing practices in the public and private sectors in India. IGC Working Paper.
- Lemos R, Scur D. 2016. Developing Management: An expanded evaluation tool for developing countries. RISE Working Paper 16/007.
- McConnell KJ, Lindrooth RC, Wholey DR, Maddox TM, Bloom N. 2013. Management practices and the quality of care in cardiac units. *JAMA internal medicine*, **173**: 684-692.
- McConnell KJ, Lindrooth RC, Wholey DR, Maddox TM, Bloom N. 2016. Modern management practices and hospital admissions. *Health economics*, **25**: 470-485.
- Pihlainen V, Kivinen T, Lammintakanen J. 2016. Management and leadership competence in hospitals: a systematic literature review. *Leadership in Health Services*, **29**: 95-110.
- Pillay R. 2008. Managerial competencies of hospital managers in South Africa: a survey of managers in the public and private sectors. *Hum Resour Health*, **6**: 4.
- Tsai TC, Jha AK, Gawande AA, et al. 2015. Hospital board and management practices are strongly related to hospital performance on clinical quality metrics. *Health Affairs*, **34**: 1304-1311.
- Woodward J. 1958. *Management and technology*. Cambridge University Press, Cambridge, UK.

## Appendix D Sensitivity analysis using different cutoffs for NCMS revenue share

### General hospitals

Table D1. Difference-in-differences (DD) estimates for general hospitals

NCMS cutoff = 20%	ln(NCMS-eligible exp)	ln(Non-NCMS-eligible exp)	ln(OOP)	LOS
Model 3 DD	0.056 (0.045)	0.309 (0.388)	0.179 (0.106)	-0.592 (0.663)
DD x staff (in 100)	0.019 (0.013)	-0.107 (0.128)	0.007 (0.034)	0.265 (0.245)
DD x high NCMS share	-0.077* (0.036)	-0.09 (0.471)	-0.189 (0.150)	1.823 (1.035)
DD x WMS score	-0.063 (0.067)	0.045 (0.730)	0.356 (0.201)	-2.277 (1.258)
DD x Wave2	0.063 (0.041)	0.706 (0.482)	0.385* (0.097)	3.295* (1.407)
DD x Wave2 x staff (in 100)	0.049 (0.025)	0.958* (0.307)	0.395* (0.036)	2.625* (1.215)
DD x Wave2 x high NCMS share	0.094 (0.048)	-1.330* (0.632)	-0.453* (0.151)	-4.374* (2.085)
DD x Wave2 x WMS score	0.08 (0.106)	2.269 (1.311)	-0.772* (0.206)	3.367 (4.894)
<i>N</i>	1989738	1899585	1983206	1920648

NCMS = New Rural Cooperative Medical Scheme; WMS = World Management Survey; LOS = length of stay; OOP = out-of-pocket payment

Notes: This admission level analysis uses 13 pairs of general hospitals. DD refers to the interaction term of treatment dummy and post dummy, "staff" refers to the demean number of medical staff in 100, "high NCMS share" refers to being in the top 20% of the distribution of revenue from NCMS. All models control for pair fixed effects and baseline hospital characteristics, including number of NCMS enrollees, total revenue, revenue from inpatient services, number of beds, number of medical equipment worth more than 10,000 RMB, dummies for management style (decentralization to departments, some decentralization, no decentralization), and dummies for self-perceived degree of competition (fierce, some, none). Standard errors are clustered at the county level and shown in parentheses.

\*  $p < 0.05$

Table D2. Difference-in-differences (DD) estimates for general hospitals

NCMS cutoff = 40%		ln(NCMS-eligible exp)	ln(Non-NCMS-eligible exp)	ln(OOP)	LOS
Model 3	DD	0.056 (0.045)	0.309 (0.388)	0.179 (0.106)	-0.592 (0.663)
	DD x staff (in 100)	0.019 (0.013)	-0.107 (0.128)	0.007 (0.034)	0.265 (0.245)
	DD x high NCMS share	-0.077* (0.036)	-0.09 (0.471)	-0.189 (0.150)	1.823 (1.035)
	DD x WMS score	-0.063 (0.067)	0.045 (0.730)	0.356 (0.201)	-2.277 (1.258)
	DD x Wave2	0.081* (0.035)	-0.147 (0.408)	0.053 (0.103)	2.525* (0.624)
	DD x Wave2 x staff (in 100)	0.060* (0.014)	0.257 (0.158)	0.091* (0.041)	1.541* (0.269)
	DD x Wave2 x high NCMS share	0.066 (0.036)	-0.102 (0.492)	0.015 (0.154)	-3.452* (1.049)
	DD x Wave2 x WMS score	0.012 (0.084)	4.232* (1.404)	-0.14 (0.352)	2.861 (1.893)
<i>N</i>		1989738	1899585	1983206	1920648

NCMS = New Rural Cooperative Medical Scheme; WMS = World Management Survey; LOS = length of stay; OOP = out-of-pocket payment

Notes: This admission level analysis uses 13 pairs of general hospitals. DD refers to the interaction term of treatment dummy and post dummy, "staff" refers to the demean number of medical staff in 100, "high NCMS share" refers to being in the top 40% of the distribution of revenue from NCMS. All models control for pair fixed effects and baseline hospital characteristics, including number of NCMS enrollees, total revenue, revenue from inpatient services, number of beds, number of medical equipment worth more than 10,000 RMB, dummies for management style (decentralization to departments, some decentralization, no decentralization), and dummies for self-perceived degree of competition (fierce, some, none). Standard errors are clustered at the county level and shown in parentheses.

\*  $p < 0.05$

## TCM hospitals

Table D3. Difference-in-differences (DD) estimates for TCM hospitals

NCMS cutoff = 20%		ln(NCMS- eligible exp)	ln(Non-NCMS- eligible exp)	ln(OOP)	LOS
Model 3	DD	0.089*	-0.496	-0.119	0.727
		(0.019)	(0.328)	(0.202)	(0.409)
	DD x staff (in 100)	-0.041*	0.675	0.381	-1.701*
		(0.008)	(0.730)	(0.199)	(0.507)
	DD x high NCMS share	0.043*	-0.004	-0.296	2.054*
		(0.009)	(0.714)	(0.183)	(0.477)
	DD x WMS score	-0.156*	0.443	0.335	-3.761*
		(0.025)	(0.747)	(0.402)	(0.760)
	DD x Wave2	-0.004	0.472	0.244	-2.314*
		(0.010)	(0.814)	(0.256)	(0.602)
	DD x Wave2 x staff (in 100)	0.073*	-0.479	-0.469	3.951*
		(0.017)	(1.386)	(0.390)	(0.968)
	DD x Wave2 x high NCMS share	-	-	-	-
		-	-	-	-
	DD x Wave2 x WMS score	-	-	-	-
		-	-	-	-
<i>N</i>		514737	489640	514093	492288

NCMS = New Rural Cooperative Medical Scheme; WMS = World Management Survey; LOS = length of stay; OOP = out-of-pocket payment

Notes: This admission level analysis uses eight pairs of Traditional Chinese Medicine hospitals. DD refers to the interaction term of treatment dummy and post dummy, "staff" refers to the demean number of medical staff in 100, "high NCMS share" refers to being in the top 20% of the distribution of revenue from NCMS. All models control for pair fixed effects and baseline hospital characteristics, including number of NCMS enrollees, total revenue, revenue from inpatient services, number of beds, number of medical equipment worth more than 10,000 RMB, dummies for management style (decentralization to departments, some decentralization, no decentralization), and dummies for self-perceived degree of competition (fierce, some, none). Standard errors are clustered at the county level and shown in parentheses.

\*  $p < 0.05$

Table D4. Difference-in-differences (DD) estimates for TCM hospitals

NCMS cutoff = 40%		ln(NCMS-eligible exp)	ln(Non-NCMS-eligible)	ln(OOP)	LOS
Model 3	DD	0.077* (0.022)	-0.086 (0.170)	0.170 (0.159)	-0.256 (0.558)
	DD x staff (in 100)	-0.033 (0.025)	-0.682* (0.183)	-0.522* (0.137)	0.733 (0.897)
	DD x high NCMS share	0.019 (0.028)	1.525* (0.261)	0.766* (0.160)	-1.133 (1.182)
	DD x WMS score	-0.109* (0.040)	-0.815* (0.357)	-0.609* (0.241)	-0.338 (1.712)
	DD x Wave2	0.022 (0.034)	-1.423* (0.336)	-1.016* (0.200)	1.333 (1.493)
	DD x Wave2 x staff (in 100)	0.021 (0.032)	1.184* (0.262)	0.877* (0.182)	-0.804 (1.254)
	DD x Wave2 x high NCMS share	-	-	-	-
	DD x Wave2 x WMS score	-	-	-	-
		-	-	-	-
<i>N</i>		514737	489640	514093	492288

NCMS = New Rural Cooperative Medical Scheme; WMS = World Management Survey; LOS = length of stay; OOP = out-of-pocket payment

Notes: This admission level analysis uses eight pairs of Traditional Chinese Medicine hospitals. DD refers to the interaction term of treatment dummy and post dummy, "staff" refers to the demean number of medical staff in 100, "high NCMS share" refers to being in the top 40% of the distribution of revenue from NCMS. All models control for pair fixed effects and baseline hospital characteristics, including number of NCMS enrollees, total revenue, revenue from inpatient services, number of beds, number of medical equipment worth more than 10,000 RMB, dummies for management style (decentralization to departments, some decentralization, no decentralization), and dummies for self-perceived degree of competition (fierce, some, none). Standard errors are clustered at the county level and shown in parentheses.

\*  $p < 0.05$

## Appendix E Exploratory analysis using quarter dummies

Model A1 is specified as follows:

$$Y_{ihct} = \alpha_0 + \alpha_1 T_{ihc} + \sum_{t=1}^{12} \beta_t q_{ihct} + \sum_{l=1}^{12} \mu_l T q_{ihct}^l + X_{ihc} \gamma + u_{ihc} + \varepsilon_{ihct} \quad (a1)$$

where  $\sum_{l=1}^{12} q_{ihct}^l$  is a set of post-intervention dummies representing the first, the second, and up to the twelfth quarter after the intervention, and  $\sum_{l=1}^{12} T q_{ihct}^l$  are the interaction terms of the treatment dummy with  $\sum_{l=1}^{12} q_{ihct}^l$ . Average treatment effect on the treated in each quarter is captured by the corresponding  $\beta_3^l$  ( $l = 1, 2, \dots, 12$ ).

Model A2 is specified as follows:

$$Y_{ihct} = \alpha_0 + \alpha_1 T_{ihc} + \sum_{t=1}^{12} \beta_t q_{ihct} + \sum_{t=1}^{12} \mu_t T q_{ihct} + \sum_{l=P,R,M} \delta_1^l Z_{ihc}^l + \sum_{l=P,R,M} \sum_{t=1}^{12} \eta_t^l T q_{ihct} \times Z_{ihc}^l + \alpha_3 Wave2_{ihc} + \sum_{t=1}^{12} \theta_t T q_{ihct} \times Wave2_{ihc} + \sum_{l=P,R,M} \sum_{t=1}^{12} \varphi_t^l T q_{ihct}^l \times Wave2_{ihc} \times Z_{ihc}^l + X_{ihc} \gamma + u_{ihc} + \varepsilon_{ihct} \quad (a2)$$

where  $\sum_{l=P,R,M} \sum_{t=1}^{12} \eta_t^l$  capture the heterogeneous response from wave 1 intervention hospitals with different characteristics in each quarter, and  $\sum_{l=P,R,M} \sum_{t=1}^{12} \varphi_t^l$  capture how wave 2 intervention hospitals behave differently from wave 1 intervention hospitals with the same characteristics in each quarter.

Table E1. Difference-in-differences (DD) estimates for general hospitals

		ln(NCMS-eligible exp)	ln(Non-NCMS-eligible exp)	ln(OOP)	LOS
Model A1	DD1	0.064	-0.047	0.004	-0.212
	DD2	0.047	-0.095	-0.004	-0.581
	DD3	0.058	0.108	-0.002	-0.264
	DD4	0.086	0.245	0.148	0.069
	DD5	0.083*	0.069	0.342	-0.184
	DD6	0.068	-0.027	0.362*	-0.016
	DD7	-0.005	0.182	0.470	-0.462
	DD8	-0.043	0.343	0.255	-0.946
	DD9	0.025	0.179	-0.198	-0.418
	DD10	0.060	-0.208	-0.202	-0.512
	DD11	0.069	0.091	-0.241	-0.285
	DD12	0.082	-0.143	-0.258	-0.823
Model A2	DD1	0.058	0.459*	0.116	1.135*
	DD2	0.007	0.058	-0.022	0.390
	DD3	0.045	-0.138	-0.071	0.545
	DD4	0.101	-0.171	0.178	0.280
	DD5	0.199*	-0.513	0.440	1.347*
	DD6	0.125*	-0.655*	0.221	1.124*
	DD7	-0.112*	-0.755	0.089	0.345
	DD8	-0.018	-0.870	-0.460	0.155
	DD9	0.145*	-0.632	-0.542	1.351*
	DD10	0.132*	-0.616	-0.271	0.597
	DD11	0.119	-0.671	0.105	0.210
	DD12	0.497	0.354	0.760	1.460
	DD1 x staff (in 100)	0.017	0.012	0.247	-2.759*
	DD2 x staff (in 100)	-0.005	0.214	0.447	-2.055*
	DD3 x staff (in 100)	0.015	0.540	0.534*	-1.308*
	DD4 x staff (in 100)	-0.029	0.564	0.517	-0.265
	DD5 x staff (in 100)	-0.164*	0.875	0.296	-2.473*
	DD6 x staff (in 100)	-0.045	0.887	0.379	-1.703*
	DD7 x staff (in 100)	0.043	0.952	0.438	-1.379*
	DD8 x staff (in 100)	-0.101*	1.031	0.636*	-1.899*
	DD9 x staff (in 100)	-0.174*	1.045	0.564*	-2.455*
	DD10 x staff (in 100)	-0.104*	0.405	0.252	-1.771*
	DD11 x staff (in 100)	-0.010	0.663	-0.133	-0.954
	DD12 x staff (in 100)	-0.206	-0.289	-0.625	-2.096

Table E1. Difference-in-differences (DD) estimates for general hospitals (continued)

		ln(NCMS-eligible exp)	ln(Non-NCMS-eligible exp)	ln(OOP)	LOS
Model A2	DD1 x high NCMS share	-0.085	-0.294	-0.142	-0.113
	DD2 x high NCMS share	-0.021	-0.601	-0.192	0.358
	DD3 x high NCMS share	-0.011	-0.635	-0.256*	0.180
	DD4 x high NCMS share	0.078	-0.592	-0.194	0.885
	DD5 x high NCMS share	-0.005	-0.458	-0.194	0.419
	DD6 x high NCMS share	-0.034	-0.041	-0.171	0.374
	DD7 x high NCMS share	-0.064	0.101	-0.157	2.969*
	DD8 x high NCMS share	0.013	0.090	-0.163	4.101*
	DD9 x high NCMS share	-0.123*	-0.070	-0.331	3.047
	DD10 x high NCMS share	-0.134*	0.039	-0.250	1.810
	DD11 x high NCMS share	-0.059	0.132	-0.277	0.741
	DD12 x high NCMS share	-0.138	0.016	-0.176	0.181
	DD1 x WMS score	0.036	-0.310	-0.060	-0.569
	DD2 x WMS score	-0.122	-0.069	-0.112	-0.631
	DD3 x WMS score	-0.144	0.089	-0.064	-0.815
	DD4 x WMS score	-0.267*	0.338	-0.167	-1.056
	DD5 x WMS score	-0.139	0.721	0.162	-0.409
	DD6 x WMS score	-0.165	0.276	0.328	-0.810
	DD7 x WMS score	-0.092	0.724	0.546*	0.053
	DD8 x WMS score	-0.347*	0.175	1.075*	-3.812
	DD9 x WMS score	-0.069	0.611	1.057*	-4.244
	DD10 x WMS score	-0.084	0.332	0.874*	-4.447*
	DD11 x WMS score	-0.172	0.584	0.804*	-0.916
	DD12 x WMS score	-0.206	1.550	1.123*	5.011*
	DD1 x Wave2	-0.127*	0.603	-0.054	2.360
	DD2 x Wave2	0.043	0.257	-0.167	3.026*
	DD3 x Wave2	0.074	0.582	-0.086	3.277*
	DD4 x Wave2	0.282*	0.696	0.093	4.604*
	DD5 x Wave2	0.164*	0.346	0.258*	3.062
	DD6 x Wave2	0.021	0.048	0.344*	2.175
	DD7 x Wave2	0.241*	0.246	0.588*	4.103*
	DD8 x Wave2	0.323*	0.802	0.738*	3.917
	DD1 x Wave2 x staff	-0.143*	0.705*	0.094	2.273
	DD2 x Wave2 x staff	-0.025	0.701	0.031	2.555*
	DD3 x Wave2 x staff	0.018	1.198*	0.155	2.713*
DD4 x Wave2 x staff	0.172*	1.350*	0.345*	3.502*	
DD5 x Wave2 x staff	0.134*	1.149*	0.678*	2.575	
DD6 x Wave2 x staff	0.013	1.061*	0.815*	1.576	
DD7 x Wave2 x staff	0.162*	0.885*	0.826*	2.487	
DD 8 x Wave2 x staff	0.158*	0.875*	0.538*	1.651	

Table E1. Difference-in-differences (DD) estimates for general hospitals (continued)

		ln(NCMS- eligible exp)	ln(Non-NCMS- eligible exp)	ln(OOP)	LOS
Model A2	DD1 x Wave2 x high NCMS share	0.370*	-1.111	0.014	-1.583
	DD2 x Wave2 x high NCMS share	0.129	-0.672	0.173	-2.881
	DD3 x Wave2 x high NCMS share	0.062	-1.199	0.069	-2.718
	DD4 x Wave2 x high NCMS share	-0.266*	-1.434*	-0.575*	-4.897*
	DD5 x Wave2 x high NCMS share	-0.068	-0.995	-0.835*	-2.245
	DD6 x Wave2 x high NCMS share	0.176*	-0.832	-0.942*	-0.968
	DD7 x Wave2 x high NCMS share	-0.080	-1.124	-0.927*	-4.795
	DD8 x Wave2 x high NCMS share	0.000	0.000	0.000	0.000
	DD1 x Wave2 x WMS score	0.140	1.635	0.212	4.714
	DD2 x Wave2 x WMS score	0.231	1.196	0.216	2.940
	DD3 x Wave2 x WMS score	0.226	0.998	-0.017	2.152
	DD4 x Wave2 x WMS score	0.102	1.214	-0.101	2.488
	DD5 x Wave2 x WMS score	0.141	2.400	-0.848*	4.587
	DD6 x Wave2 x WMS score	0.123	4.406*	-1.077*	6.388
	DD7 x Wave2 x WMS score	-0.685*	3.546*	-1.590*	2.411
	DD8 x Wave2 x WMS score	0.167	3.278*	-2.115*	9.746
<i>N</i>		1989738	1899585	1983206	1920648

staff = number of medical staff (in 100 persons); WMS score = World Management Score; NCMS = New Rural Cooperative Medical Scheme; OOP = out-of-pocket payment; LOS = length of stay

Notes: This admission level analysis uses 13 pairs of general hospitals. DD1-DD12 refers to the interaction term of treatment dummy and post-intervention quarter dummy, "staff" refers to the demean number of medical staff in 100, "high NCMS share" refers to being in the top 30% of the distribution of revenue from NCMS. All models control for paired fixed effect and baseline hospital characteristics, including number of NCMS enrollees, total revenue, revenue from inpatient services, number of beds, number of medical equipment worth more than 10,000 RMB, dummies for management style (decentralization to departments, some decentralization, no decentralization), and dummies for self-perceived degree of competition (fierce, some, none). Standard errors are clustered at the county level.

\*  $p < 0.05$

Table E2. Difference-in-differences (DD) estimates for TCM hospitals

		ln(NCMS-eligible exp)	ln(Non-NCMS-eligible exp)	ln(OOP)	LOS
Model A1	DD1	0.064	-0.047	0.004	-0.212
	DD2	0.047	-0.095	-0.004	-0.581
	DD3	0.058	0.108	-0.002	-0.264
	DD4	0.086	0.245	0.148	0.069
	DD5	0.083*	0.069	0.342	-0.184
	DD6	0.068	-0.027	0.362*	-0.016
	DD7	-0.005	0.182	0.470	-0.462
	DD8	-0.043	0.343	0.255	-0.946
	DD9	0.025	0.179	-0.198	-0.418
	DD10	0.060	-0.208	-0.202	-0.512
	DD11	0.069	0.091	-0.241	-0.285
	DD12	0.082	-0.143	-0.258	-0.823
Model A2	DD1	0.058	0.459*	0.116	1.135*
	DD2	0.007	0.058	-0.022	0.390
	DD3	0.045	-0.138	-0.071	0.545
	DD4	0.101	-0.171	0.178	0.280
	DD5	0.199*	-0.513	0.440	1.347*
	DD6	0.125*	-0.655*	0.221	1.124*
	DD7	-0.112*	-0.755	0.089	0.345
	DD8	-0.018	-0.870	-0.460	0.155
	DD9	0.145*	-0.632	-0.542	1.351*
	DD10	0.132*	-0.616	-0.271	0.597
	DD11	0.119	-0.671	0.105	0.210
	DD12	0.497	0.354	0.760	1.460
	DD1 x staff (in 100)	0.017	0.012	0.247	-2.759*
	DD2 x staff (in 100)	-0.005	0.214	0.447	-2.055*
	DD3 x staff (in 100)	0.015	0.540	0.534*	-1.308*
	DD4 x staff (in 100)	-0.029	0.564	0.517	-0.265
	DD5 x staff (in 100)	-0.164*	0.875	0.296	-2.473*
	DD6 x staff (in 100)	-0.045	0.887	0.379	-1.703*
	DD7 x staff (in 100)	0.043	0.952	0.438	-1.379*
	DD8 x staff (in 100)	-0.101*	1.031	0.636*	-1.899*
	DD9 x staff (in 100)	-0.174*	1.045	0.564*	-2.455*
	DD10 x staff (in 100)	-0.104*	0.405	0.252	-1.771*
	DD11 x staff (in 100)	-0.010	0.663	-0.133	-0.954
	DD12 x staff (in 100)	-0.206	-0.289	-0.625	-2.096

Table E2. Difference-in-differences (DD) estimates for TCM hospitals (continued)

		ln(NCMS- eligible exp)	ln(Non-NCMS- eligible exp)	ln(OOP)	LOS
Model A2	DD1 x high NCMS share	0.132*	-0.068	-0.188	3.178*
	DD2 x high NCMS share	0.076	-0.157	-0.489*	2.342*
	DD3 x high NCMS share	0.049	-0.158	-0.609*	1.600*
	DD4 x high NCMS share	0.045	-0.204	-0.604*	0.725
	DD5 x high NCMS share	0.112*	-0.474	-0.353	2.770*
	DD6 x high NCMS share	-0.045	-0.419	-0.467*	1.740*
	DD7 x high NCMS share	0.014	-0.107	-0.346	1.906*
	DD8 x high NCMS share	0.124*	-0.324	-0.471	2.352*
	DD9 x high NCMS share	0.132*	-0.155	-0.437	2.613*
	DD10 x high NCMS share	0.105*	0.977	0.110	2.375*
	DD11 x high NCMS share	0.027	0.851	0.346	1.752*
	DD12 x high NCMS share	0.008	1.042	0.649*	1.899
	DD1 x WMS score	-0.320*	-2.321*	-0.755*	-4.979*
	DD2 x WMS score	-0.176	-0.993	-0.199	-2.987*
	DD3 x WMS score	-0.294*	-0.195	-0.089	-3.626*
	DD4 x WMS score	-0.310	0.714	0.101	-2.076
	DD5 x WMS score	-0.387*	0.707	0.087	-5.324*
	DD6 x WMS score	-0.083	0.572	0.770*	-3.005*
	DD7 x WMS score	0.267*	1.597	1.221*	-3.027*
	DD8 x WMS score	-0.227	3.044*	2.351*	-4.240*
	DD9 x WMS score	-0.373*	1.208*	0.921	-5.803*
	DD10 x WMS score	-0.289*	-0.310	-0.192	-4.073*
	DD11 x WMS score	-0.289	0.457	-1.094	-2.646*
	DD12 x WMS score	-1.184	-1.738	-2.419	-6.669*
	DD1 x Wave2	-0.025	-0.692	0.028	-4.153*
	DD2 x Wave2	0.062	-0.279	0.481	-3.442*
	DD3 x Wave2	0.118*	0.212	0.688*	-1.862*
	DD4 x Wave2	0.017	0.377	-0.013	-0.781
	DD5 x Wave2	-0.145*	0.975	-0.329	-3.309*
	DD6 x Wave2	-0.032	1.274	0.229	-2.797*
	DD7 x Wave2	0.068	1.153	0.516	-2.777*
	DD8 x Wave2	0.148*	-0.002	0.005	0.320*
	DD1 x Wave2 x staff	-0.035	1.073	-0.065	3.901*
DD2 x Wave2 x staff	0.023	0.477	-0.491	3.104*	
DD3 x Wave2 x staff	-0.062	-0.118	-0.650*	1.848*	
DD4 x Wave2 x staff	0.105	-0.351	0.001	1.017	
DD5 x Wave2 x staff	0.193*	-0.776	0.192	2.813*	
DD6 x Wave2 x staff	0.058	-1.114	-0.284	2.057*	
DD7 x Wave2 x staff	-0.077	-1.180	-0.573	2.261*	
DD 8 x Wave2 x staff	-	-	-	-	

Table E2. Difference-in-differences (DD) estimates for TCM hospitals (continued)

		ln(NCMS- eligible exp)	ln(Non-NCMS- eligible exp)	ln(OOP)	LOS
Model A2	DD1 x Wave2 x high NCMS share	-	-	-	-
	DD2 x Wave2 x high NCMS share	-	-	-	-
	DD3 x Wave2 x high NCMS share	-	-	-	-
	DD4 x Wave2 x high NCMS share	-	-	-	-
	DD5 x Wave2 x high NCMS share	-	-	-	-
	DD6 x Wave2 x high NCMS share	-	-	-	-
	DD7 x Wave2 x high NCMS share	-	-	-	-
	DD8 x Wave2 x high NCMS share	-	-	-	-
	DD1 x Wave2 x WMS score	-	-	-	-
	DD2 x Wave2 x WMS score	-	-	-	-
	DD3 x Wave2 x WMS score	-	-	-	-
	DD4 x Wave2 x WMS score	-	-	-	-
	DD5 x Wave2 x WMS score	-	-	-	-
	DD6 x Wave2 x WMS score	-	-	-	-
	DD7 x Wave2 x WMS score	-	-	-	-
	DD8 x Wave2 x WMS score	-	-	-	-
<i>N</i>		514737	489640	514093	492288

staff = number of medical professionals (in 100 persons); WMS score = World Management Score; NCMS = New Rural Cooperative Medical Scheme; OOP = out-of-pocket payment; LOS = length of stay

Notes: This admission level analysis uses eight pairs of Traditional Chinese Medicine hospitals. DD1-DD12 refers to the interaction term of treatment dummy and post-intervention quarter dummy, "staff" refers to the demean number of medical staff in 100, "high NCMS share" refers to being in the top 30% of the distribution of revenue from NCMS. All models control for paired fixed effect and baseline hospital characteristics, including number of NCMS enrollees, total revenue, revenue from inpatient services, number of beds, number of medical equipment worth more than 10,000 RMB, dummies for management style (decentralization to departments, some decentralization, no decentralization), and dummies for self-perceived degree of competition (fierce, some, none). Standard errors are clustered at the county level.

\*  $p < 0.05$

## Appendix F Exploratory analysis using WMS dimensional score

Table F1. Difference-in-differences (DD) estimates for general hospitals using WMS target dimension

WMS dimension = targets		ln(NCMS-eligible exp)	ln(Non-NCMS-eligible exp)	ln(OOP)	LOS
Model 3	DD	0.047 (0.042)	0.193 (0.275)	0.145 (0.080)	-0.454 (0.530)
	DD x staff (in 100)	0.021 (0.015)	-0.081 (0.167)	0.010 (0.052)	0.269 (0.320)
	DD x high NCMS share	-0.066* (0.022)	0.144 (0.226)	-0.084 (0.067)	1.314 (0.642)
	DD x WMS score	-0.073 (0.042)	-0.544 (0.569)	0.018 (0.187)	-0.576 (1.141)
	DD x Wave2	0.071* (0.030)	1.348* (0.261)	0.338* (0.068)	3.119* (0.514)
	DD x Wave2 x staff (in 100)	0.042* (0.016)	1.394* (0.168)	0.328* (0.055)	2.399* (0.540)
	DD x Wave2 x high NCMS share	0.090* (0.023)	-2.497* (0.226)	-0.425* (0.069)	-3.547* (0.756)
	DD x Wave2 x WMS score	0.043 (0.044)	1.363* (0.569)	-0.07 (0.190)	-1.118 (1.429)
	<i>N</i>		1989738	1899585	1983206

NCMS = New Rural Cooperative Medical Scheme; WMS = World Management Survey; LOS = length of stay; OOP = out-of-pocket payment

Notes: This admission level analysis uses 13 pairs of general hospitals. DD refers to the interaction term of treatment dummy and post dummy, "staff" refers to the demean number of medical staff in 100, "high NCMS share" refers to being in the top 30% of the distribution of revenue from NCMS. All models control for pair fixed effects and baseline hospital characteristics, including number of NCMS enrollees, total revenue, revenue from inpatient services, number of beds, number of medical equipment worth more than 10,000 RMB, dummies for management style (decentralization to departments, some decentralization, no decentralization), and dummies for self-perceived degree of competition (fierce, some, none). Standard errors are clustered at the county level and shown in parentheses.

\*  $p < 0.05$

Table F2. Difference-in-differences (DD) estimates for general hospitals using WMS monitoring dimension

WMS dimension = monitoring		ln(NCMS-eligible exp)	ln(Non-NCMS-eligible exp)	ln(OOP)	LOS
Model 3	DD	0.059 (0.042)	0.291 (0.338)	0.15 (0.100)	-0.389 (0.625)
	DD x staff (in 100)	0.015 (0.011)	-0.117 (0.110)	0.018 (0.037)	0.201 (0.261)
	DD x high NCMS share	-0.078* (0.029)	-0.011 (0.357)	-0.118 (0.113)	1.335 (0.881)
	DD x WMS score	-0.088 (0.072)	-0.307 (0.758)	0.195 (0.231)	-1.148 (1.661)
	DD x Wave2	0.05 (0.031)	1.682* (0.332)	0.287* (0.091)	2.579* (0.731)
	DD x Wave2 x staff (in 100)	0.046* (0.014)	1.655* (0.127)	0.292* (0.039)	2.309* (0.581)
	DD x Wave2 x high NCMS share	0.102* (0.031)	-2.533* (0.364)	-0.361* (0.114)	-3.551* (1.054)
	DD x Wave2 x WMS score	0.201* (0.096)	-4.556* (0.881)	0.288 (0.251)	7.446 (4.007)
<i>N</i>		1989738	1899585	1983206	1920648

NCMS = New Rural Cooperative Medical Scheme; WMS = World Management Survey; LOS = length of stay; OOP = out-of-pocket payment

Notes: This admission level analysis uses 13 pairs of general hospitals. DD refers to the interaction term of treatment dummy and post dummy, "staff" refers to the demean number of medical staff in 100, "high NCMS share" refers to being in the top 30% of the distribution of revenue from NCMS. All models control for pair fixed effects and baseline hospital characteristics, including number of NCMS enrollees, total revenue, revenue from inpatient services, number of beds, number of medical equipment worth more than 10,000 RMB, dummies for management style (decentralization to departments, some decentralization, no decentralization), and dummies for self-perceived degree of competition (fierce, some, none). Standard errors are clustered at the county level and shown in parentheses.

\*  $p < 0.05$

Table F3. Difference-in-differences (DD) estimates for TCM hospitals using WMS targets dimension

WMS dimension = targets		ln(NCMS-eligible exp)	ln(Non-NCMS-eligible exp)	ln(OOP)	LOS
Model 3	DD	0.092*	-0.277	0.023	0.479
		(0.022)	(0.363)	(0.266)	(0.731)
	DD x staff (in 100)	-0.054*	0.407	0.238	-1.533
		(0.015)	(0.642)	(0.262)	(0.896)
	DD x high NCMS share	0.057*	0.332	-0.107	1.816*
		(0.023)	(0.591)	(0.223)	(0.799)
	DD x WMS score	-0.139*	-0.335	-0.189	-2.361
		(0.049)	(0.623)	(0.444)	(1.342)
	DD x Wave2	-0.011	-0.011	0.014	-1.976
		(0.025)	(0.762)	(0.361)	(1.169)
	DD x Wave2 x staff (in 100)	0.002	-0.135	-0.048	1.15
		(0.011)	(0.617)	(0.230)	(0.817)
	DD x Wave2 x high NCMS share	-	-	-	-
		-	-	-	-
	DD x Wave2 x WMS score	-	-	-	-
		-	-	-	-
<i>N</i>		514737	489640	514093	492288

NCMS = New Rural Cooperative Medical Scheme; WMS = World Management Survey; LOS = length of stay; OOP = out-of-pocket payment

Notes: This admission level analysis uses eight pairs of Traditional Chinese Medicine hospitals. DD refers to the interaction term of treatment dummy and post dummy, "staff" refers to the demean number of medical staff in 100, "high NCMS share" refers to being in the top 30% of the distribution of revenue from NCMS. All models control for pair fixed effects and baseline hospital characteristics, including number of NCMS enrollees, total revenue, revenue from inpatient services, number of beds, number of medical equipment worth more than 10,000 RMB, dummies for management style (decentralization to departments, some decentralization, no decentralization), and dummies for self-perceived degree of competition (fierce, some, none). Standard errors are clustered at the county level and shown in parentheses.

\*  $p < 0.05$

Table F4. Difference-in-differences (DD) estimates for TCM hospitals using WMS monitoring dimension

WMS dimension = monitoring		ln(NCMS-eligible exp)	ln(Non-NCMS-eligible exp)	ln(OOP)	LOS
Model 3	DD	0.062*	-0.495*	-0.093	0.160
		(0.020)	(0.221)	(0.158)	(0.297)
	DD x staff (in 100)	-0.004	0.702	0.296*	-0.888*
		(0.016)	(0.526)	(0.129)	(0.125)
	DD x high NCMS share	0.019*	-0.218	-0.312*	1.752*
		(0.009)	(0.583)	(0.095)	(0.116)
	DD x WMS score	-0.082*	0.707	0.406*	-2.610*
		(0.016)	(0.449)	(0.150)	(0.127)
	DD x Wave2	0.026	0.772	0.344*	-2.242*
		(0.013)	(0.742)	(0.127)	(0.150)
	DD x Wave2 x staff (in 100)	0.009	-0.766	-0.300*	2.100*
		(0.012)	(0.689)	(0.117)	(0.139)
	DD x Wave2 x high NCMS share	-	-	-	-
		-	-	-	-
	DD x Wave2 x WMS score	-	-	-	-
		-	-	-	-
<i>N</i>		514737	489640	514093	492288

NCMS = New Rural Cooperative Medical Scheme; WMS = World Management Survey; LOS = length of stay; OOP = out-of-pocket payment

Notes: This admission level analysis uses eight pairs of Traditional Chinese Medicine hospitals. DD refers to the interaction term of treatment dummy and post dummy, "staff" refers to the demean number of medical staff in 100, "high NCMS share" refers to being in the top 30% of the distribution of revenue from NCMS. All models control for pair fixed effects and baseline hospital characteristics, including number of NCMS enrollees, total revenue, revenue from inpatient services, number of beds, number of medical equipment worth more than 10,000 RMB, dummies for management style (decentralization to departments, some decentralization, no decentralization), and dummies for self-perceived degree of competition (fierce, some, none). Standard errors are clustered at the county level and shown in parentheses.

\*  $p < 0.05$

## Appendix G Exploratory analysis on why wave 1 and 2 effects differ

### 1. Separate year 1 and year 2 effects for wave 1

Table G1. Difference-in-differences (DD) estimates for general hospitals

	ln(NCMS-eligible exp)	ln(Non-NCMS-eligible exp)	ln(OOP)	LOS
Model 1 DD	0.027 (0.042)	0.165 (0.337)	0.175 (0.108)	-0.216 (0.691)
Model 2 DD	0.014 (0.039)	0.156 (0.317)	0.142 (0.095)	-0.078 (0.553)
DD x Year 2	0.028 (0.044)	0.014 (0.173)	0.056 (0.085)	-0.238 (0.698)
Model 3 DD	-0.013 (0.048)	0.409 (0.365)	0.193 (0.107)	-0.286 (0.550)
DD x staff (in 100)	0.057 (0.058)	0.003 (0.160)	0.101 (0.076)	-1.010 (0.577)
DD x high NCMS share	0.030 (0.017)	-0.107 (0.112)	-0.006 (0.040)	0.037 (0.089)
DD x WMS score	-0.042 (0.042)	-0.446 (0.426)	-0.178 (0.126)	0.614 (0.401)
DD x Year2	-0.057 (0.106)	-0.211 (0.637)	-0.141 (0.227)	-1.141* (0.472)
DD x Year2 x staff (in 100)	-0.019 (0.019)	-0.090 (0.061)	-0.046* (0.008)	0.183 (0.211)
DD x Year2 x high NCMS share	-0.010 (0.072)	0.458* (0.204)	0.048 (0.029)	1.877* (0.834)
DD x Year2 x WMS score	-0.093 (0.124)	0.254 (0.378)	0.313* (0.060)	0.638 (1.109)
<i>N</i>	1000401	944758	998417	926541

NCMS = New Rural Cooperative Medical Scheme; WMS = World Management Survey; LOS = length of stay; OOP = out-of-pocket payment

Notes: This admission level analysis separately estimates year 1 and year 2 effect for wave 1 general hospitals. DD refers to the interaction term of treatment dummy and post dummy, "staff" refers to the demean number of medical staff in 100, "high NCMS share" refers to being in the top 30% of the distribution of revenue from NCMS. All models control for pair fixed effects and baseline hospital characteristics, including number of NCMS enrollees, total revenue, revenue from inpatient services, number of beds, number of medical equipment worth more than 10,000 RMB, dummies for management style (decentralization to departments, some decentralization, no decentralization), and dummies for self-perceived degree of competition (fierce, some, none). Standard errors are clustered at the county level and shown in parentheses.

\*  $p < 0.05$

Table G2. Difference-in-differences (DD) estimates for TCM hospitals

	ln(NCMS-eligible exp)	ln(Non-NCMS-eligible exp)	ln(OOP)	LOS
Model 1 DD	0.025 (0.060)	0.304 (0.367)	-0.147 (0.291)	-0.878 (0.857)
Model 2 DD	0.004 (0.096)	0.290 (0.331)	-0.204 (0.286)	-1.073 (1.195)
DD x Year 2	0.024 (0.076)	0.061 (0.258)	0.180 (0.213)	0.403 (0.857)
Model 3 DD	0.042 (0.095)	0.302 (0.431)	-0.277 (0.297)	-0.261 (1.186)
DD x staff (in 100)	0.043 (0.075)	-0.525* (0.236)	0.052 (0.197)	0.876 (0.870)
DD x high NCMS share	-0.004 (0.034)	0.406 (0.742)	0.423 (0.213)	-1.120* (0.399)
DD x WMS score	0.077 (0.042)	-0.240 (0.686)	-0.452* (0.168)	1.432* (0.546)
DD x Year2	-0.274* (0.114)	-0.667 (0.716)	-0.270 (0.334)	-2.989* (0.850)
DD x Year2 x staff (in 100)	-0.075 (0.038)	0.496* (0.153)	-0.142* (0.030)	-0.818 (0.474)
DD x Year2 x high NCMS share	-0.017 (0.035)	-0.048 (0.292)	0.181* (0.055)	0.779 (0.464)
DD x Year2 x WMS score	0.127 (0.082)	1.492 (0.856)	0.751* (0.167)	-0.813 (0.553)
<i>N</i>	220453	200249	220021	197574

NCMS = New Rural Cooperative Medical Scheme; WMS = World Management Survey; LOS = length of stay; OOP = out-of-pocket payment

Notes: This admission level analysis separately estimates year 1 and year 2 effect for wave 1 TCM hospitals. DD refers to the interaction term of treatment dummy and post dummy, "staff" refers to the demean number of medical staff in 100, "high NCMS share" refers to being in the top 30% of the distribution of revenue from NCMS. All models control for pair fixed effects and baseline hospital characteristics, including number of NCMS enrollees, total revenue, revenue from inpatient services, number of beds, number of medical equipment worth more than 10,000 RMB, dummies for management style (decentralization to departments, some decentralization, no decentralization), and dummies for self-perceived degree of competition (fierce, some, none). Standard errors are clustered at the county level and shown in parentheses.

\*  $p < 0.05$

## 2. DID estimates for wave 1 year 1 using wave 1 control counties in 2016 and all wave 2 counties in 2016 as control

Table G3. DID estimates for wave 1 general hospitals, year 1

		ln(NCMS- eligible exp)	ln(Non-NCMS- eligible exp)	ln(OOP)	LOS
Model 1	DD	0.016 (0.037)	0.211 (0.326)	0.149 (0.093)	-0.048 (0.519)
Model 2	DD	-0.008 (0.045)	0.416 (0.358)	0.195 (0.101)	-0.172 (0.513)
	DD x staff (in 100)	0.032 (0.017)	-0.059 (0.121)	0.004 (0.041)	0.065 (0.049)
	DD x high NCMS share	-0.048 (0.040)	-0.474 (0.427)	-0.187 (0.121)	0.227 (0.258)
	DD x WMS score	-0.032 (0.105)	-0.371 (0.682)	-0.159 (0.236)	-0.972* (0.296)
<i>N</i>		1080567	1017209	1078554	1005670

NCMS = New Rural Cooperative Medical Scheme; WMS = World Management Survey; LOS = length of stay; OOP = out-of-pocket payment

Notes: This admission level analysis estimates effect for wave 1 general hospitals in year 1 using wave 1 control counties' general hospitals in 2016 and all wave 2 counties' general hospitals in 2016 (both eventual treatment and control) as a counterfactual. DD refers to the interaction term of treatment dummy and post dummy, "staff" refers to the demean number of medical staff in 100, "high NCMS share" refers to being in the top 30% of the distribution of revenue from NCMS. All models control for pair fixed effects and baseline hospital characteristics, including number of NCMS enrollees, total revenue, revenue from inpatient services, number of beds, number of medical equipment worth more than 10,000 RMB, dummies for management style (decentralization to departments, some decentralization, no decentralization), and dummies for self-perceived degree of competition (fierce, some, none). Standard errors are clustered at the county level and shown in parentheses.

\*  $p < 0.05$

Table G4. DID estimates for wave 1 TCM hospitals, year 1

		ln(NCMS-eligible exp)	ln(Non-NCMS-eligible exp)	ln(OOP)	LOS
Model 1	DD	-0.003 (0.099)	0.318 (0.344)	-0.222 (0.279)	-1.120 (1.194)
Model 2	DD	0.012 (0.099)	0.344 (0.416)	-0.272 (0.288)	-0.388 (1.189)
	DD x staff (in 100)	0.017 (0.040)	0.665 (0.599)	0.423 (0.201)	-0.873 (0.462)
	DD x high NCMS share	0.061 (0.043)	-0.451 (0.524)	-0.431* (0.151)	1.118 (0.574)
	DD x WMS score	-0.200 (0.114)	-1.032 (0.628)	-0.410 (0.307)	-2.805* (0.801)
<i>N</i>		281959	260124	281547	258978

NCMS = New Rural Cooperative Medical Scheme; WMS = World Management Survey; LOS = length of stay; OOP = out-of-pocket payment

Notes: This admission level analysis estimates effect for wave 1 TCM hospitals in year 1 using wave 1 control counties' TCM hospitals in 2016 and all wave 2 counties' TCM hospitals in 2016 (both eventual treatment and control) as a counterfactual. DD refers to the interaction term of treatment dummy and post dummy, "staff" refers to the demean number of medical staff in 100, "high NCMS share" refers to being in the top 30% of the distribution of revenue from NCMS. All models control for pair fixed effects and baseline hospital characteristics, including number of NCMS enrollees, total revenue, revenue from inpatient services, number of beds, number of medical equipment worth more than 10,000 RMB, dummies for management style (decentralization to departments, some decentralization, no decentralization), and dummies for self-perceived degree of competition (fierce, some, none). Standard errors are clustered at the county level and shown in parentheses.

\*  $p < 0.05$

### 3. Falsification test for wave 2 using pre-intervention data

Table G5. Falsification test for wave 2 general and TCM hospitals

		ln(NCMS-eligible exp)	ln(Non-NCMS-eligible exp)	ln(OOP)	LOS
General hospital	DD	0.029 (0.037)	0.149 (0.268)	0.001 (0.081)	-0.298 (0.583)
<i>N</i>		384261	366415	383832	385257
TCM hospital	DD	0.016 (0.051)	0.108 (0.137)	-0.02 (0.116)	-0.142 (0.244)
<i>N</i>		125409	123074	125407	125491

NCMS = New Rural Cooperative Medical Scheme; WMS = World Management Survey; LOS = length of stay; OOP =

Notes: This admission level analysis is a falsification test. It estimates effect of wave 2 general and TCM hospitals using data prior to intervention. DD refers to the interaction term of treatment dummy and post dummy, "staff" refers to the demean number of medical staff in 100, "high NCMS share" refers to being in the top 30% of the distribution of revenue from NCMS. All models control for pair fixed effects and baseline hospital characteristics, including number of NCMS enrollees, total revenue, revenue from inpatient services, number of beds, number of medical equipment worth more than 10,000 RMB, dummies for management style (decentralization to departments, some decentralization, no decentralization), and dummies for self-perceived degree of competition (fierce, some, none). Standard errors are clustered at the county level and shown in parentheses.

\*  $p < 0.05$