An Evaluation Index System for Research Efficiency of Research-Oriented Hospitals in China

Yin Li[1], Jiachang Li[1], Baihong Li[1], Longhao Zhang[1], Zhi Zeng, MD[2] , and Wen Zeng[3]

Abstract

A specific and rational index system is key to scientific research evaluation. According to the characteristics and status of research-oriented hospitals in China, this study aimed to construct a comprehensive and methodical system for scientific research evaluation. Using bibliometric research, we sorted and refined indices for both domestic and international scientific research evaluation systems, established two-dimensional indices of input and output, and constructed the theoretical framework of evaluation after experts. The Delphi method was adopted to determine the evaluation indices at all levels, and the Analytic Hierarchy Process was used to calculate the weights of the indices at all levels. Twenty experts from different medical fields were involved in the 2 rounds study. Altogether, 7 primary, 14 secondary, and 37 tertiary indices were included in the evaluation system. A matrix was built to conduct the maximum eigenvalue, the consistency indices, and the consistency ratio of each expert in the survey. The index weight coefficients of the indices were calculated accordingly. The model exhibited high consistency, and the credibility of the results was verified. The evaluation system for research-oriented hospitals that we established had high specificity, credibility, and rationality. The evaluation system that we established combines some quantitative evaluation indicators, which are subsequently weighted according to their importance in the field of research-oriented hospital. Evaluation index system will provide the practical manner in the future for comparing the potential academic level and impact of research-oriented hospitals. Moreover, further verification, adjustments, and optimization of the system and indicators will be performed in follow-up empirical studies.

Keywords

research-oriented hospital, evaluation index system, scientific research, Delphi method, analytic hierarchy process

Question-And-Answer Highlights

(1) What do we already know about this topic?

A specific and rational index system is key to scientific research evaluation.

(2) How does your research contribute to the field?

This study aimed to construct a comprehensive and methodical system for scientific research evaluation.

(3) What are your research's implications toward theory, practice, or policy?

An evaluation system for research-oriented hospitals.

[1]"Double First-class" Construction Office, West China Hospital, Sichuan University, Chengdu, China
[2]Department of Cardiology, West China Hospital, Sichuan University, Chengdu, China
[3]Graduate Educational Administration, West China Hospital, Sichuan University, Chengdu, China

Corresponding Authors:
Zhi Zeng, Department of Cardiology, West China Hospital, Sichuan University, 37 Guoxue Road, Chengdu 610041, China.
Email: zhizeng_scu@yeah.net

Wen Zeng, Graduate Educational Administration, West China Hospital, Sichuan University, 37 Guoxue Road, Chengdu 610041, China.
Email: 1045147804@qq.com

Introduction

A research-oriented hospital is an institution that integrates medical services, talent training, and scientific research, and has many scientific research resources, such as complete specialties, concentration of talent, and sophisticated equipment and more funds.[1] Compared to other hospitals, and therefore, efficiency evaluation is important for research-oriented hospitals.

As the main undertakers of medical technology innovation, these hospitals undertake the responsibility of researching, disseminating, and applying innovative medical knowledge. In 2020, the COVID-19 pandemic enabled hospitals and the health sector to receive widespread attention from society, and this situation has exposed the problems of the global medical system, the medical industry, and medical research.[2] While the evaluation of hospitals' research performances has the benefit of revealing inefficient aspects of hospitals, it can also create several positive outcomes such as contributing to policy makers' decision-making for enhancing the performance of science research and promote achievement application and transformation within research-oriented hospitals.[3] Thus, evaluation of scientific research is a crucial part of medical research management and hospital management, which primarily require the establishment of a comprehensive evaluation system.[4] In particular, evaluation is valuable for enhancing the performance of scientific research and promoting application of research outcomes and transformation within research-oriented hospitals.[5,6]

Researchers have established an index system for scientific research.[7-11] Li and Hao[12] developed an evaluation index system for determining the academic impact of military medical scholars. Wu[13] created a quantitative medical technology evaluation system through a questionnaire survey within medical institutions to assess medical technologies. At present, China has established Science and Technology Evaluation Metrics (STEM) ranking in China by unofficial third-party organizations.[14] However, in the country, the evaluation of scientific research at hospitals has been done through a single index characterized by untargeted incomplete data, inaccuracy, and a tendency to be based on total output value.[13,15,16] Thus, this study aimed to develop an evaluation index system through literature research based on the characteristics and status of research-oriented hospitals in China.[17] Subsequently, we sought to build a quantitative medical research evaluation system through a questionnaire survey using the Delphi method and set the weight coefficients of the identified indices through the analytical hierarchy process (AHP).[18] Furthermore, to verify the feasibility and effectiveness of the evaluation system, we conducted a case analysis based on technique for order of preference by similarity to ideal solution (TOPSIS) and data envelopment analysis (DEA) to evaluate the research efficiency of specialty in hospital H.

With the establishment of this evaluation system and further verification in research practice, we hope that hospitals would be able to conduct a better assessment of scientific research, promote its output, and provide policy support for the management of research-oriented hospitals.

Materials and Methods

Literature Research

We retrieved the literature on the scientific research evaluation of relevant medical colleges and universities that were available on Web of Science, PubMed, and Scopus, and then determined the frequency of statistical indices.[19] We referred to recognized medical field evaluation ranking indices and selected the most used ones. Thereafter, to establish an index pool, we combined the scientific research status and characteristics of research-oriented hospitals.[20]

Constructing the Initial Framework

The evaluation should be conducted for performance rather than for output alone. A total of 10 experts were interviewed through conference discussions or telephone surveys to form the initial framework. The production theory (ie, the minimum input is used to obtain the maximum output) was used as a guideline. The established index pool was incorporated. In addition to theoretical induction and deduction and other normative research methods, we drafted the initial evaluation system to have the 2 dimensions of input and output.[21]

Constructing an Evaluation Index System Using the Delphi Method

Selection of questionnaire survey experts. In the questionnaire survey, the accuracy and consistency of the results were related to the number of experts participating in the investigation. According to the principle of the Delphi method, to ensure the credibility and authority of results, the optimal number of experts investigated was 15–50.[22] Thus, we selected 20 experts according to the requirements of the Delphi method, and the selected experts were required to be familiar with hospitals or specialty managers. The experts in this study were selected mainly based on the following criteria: (1) from research-oriented hospitals that ranked top 5 in STEM, (2) vice president or head of a department in charge of research management, and (3) clinical specialty director, which, in the Chinese Medical Association and other national societies, are vice chairman and any higher ranks.[23]

Questionnaire survey. We designed an expert consultation questionnaire based on independent expert opinions collected using the Delphi method. In the questionnaire, experts were asked to respond to the necessity, importance, and operability of each index on a five-point Likert scale (5 = *very much*, 4 = *so more*, 3 = *generally*, 2 = *less*, 1 = *not at all*).[24] Theoretical analysis, practical experience, understanding of the provided data, and intuition influenced the experts' judgment, and the

Table 1. Matrix of the Assignment Based on Expert Judgment.

Judgment Basis Items	Ca Basis (Self-Evaluation by Experts)		
	High	Medium	Low
Theoretical analysis	0.3	0.2	0.1
Experience	0.5	0.4	0.3
Learn from peers	0.1	0.1	0.1
Intuition	0.1	0.1	0.1

Table 2. Matrix of Expert Familiarity Quantification.

Familiarity	Very Much	More	Generally	Less	Not at all
Value	1	0.8	0.6	0.4	0.2

degree of influence on their judgment was divided into 3 scales (1 = *small*, 2 = *medium*, 3 = *large influence*). In addition, the experts' suggestions regarding adjustments for the structure of the evaluation system by adding, deleting, or merging indices were noted.

Defining the evaluation system and indices. The positive coefficient for expert consultation was expressed by the questionnaire's recovery rate. We used the formula RR = M/N (where RR is expert positive coefficient, M is experts participating in the evaluation, and N is the total number of experts), with higher positive coefficients indicating better results. A previous study proposed a 70% positive coefficient as a good result.[25] The degree of expert authority was expressed by the authority coefficient Cr, which is dependent on the expert's decision and judgment for each item. Subsequently, we derived the arithmetic mean of the quantified value and the quantified value of familiarity as follows:

Cr = (Ca + Cs)/2 (where Ca is basis for judgment and Cs is familiarity)

This is showed in Tables 1 and 2.

That is, the greater the Cr, the higher the degree of authority. If Cr ≥ 70, then the expert's authority coefficient was considered to be sufficiently high, and, thus, the result of the inquiry was scientific and representative.[26]

The degree of coordination among expert opinions was important. Through calculation, we examined the differences in the evaluation of items by experts, as reflected by 2 indicators: coefficient of variance (CV) and coordination coefficient (Kendall's W, *W*). Here, W ranged from 0 to 1, with larger values indicating better coordination. The calculation and correction formulas for W are expressed as follows

$$W' = \frac{12}{m^2(n^3 - n) - m\sum t_k^3 - t_k} \sum d_i^2 \qquad (1)$$

where t_k represents the number of the Kth same rank, m is the total number of experts, and n is the total number of indices. Meanwhile, CV reflected the degree of fluctuation of the value assigned by the expert group to each index, with smaller values indicating better coordination. In general, the CV was <.3. The formula for calculating the CV is expressed as follows

$$CV_j = S_j / M_j \qquad (2)$$

where CV_j is the coefficient of variation of the expert score of item *j* and S_j is the standard deviation of the expert score of item *j*.

We performed 2 sets of comparison of the modified indices and subsequently constructed a comparison judgment matrix. Finally, according to the results of statistical analysis, if the mean values of the necessity indices and CV were ≥4 and ≤.15, respectively, then expert recognition of the indices was deemed high, and the index was retained. The corresponding values of ≤ 3 or ≥.15 points imply that the indices required further revision to be set as the final indices.[27]

Allocation of index weights using analytic hierarchy process (AHP). We used R3.5.1 software to complete the AHP to determine the index weights. In constructing the AHP model, we referred to the suggestion of Saaty et al,[28] which used a scale of 1–9 to evaluate the relative importance of the 2 indices. Thereafter, we calculated the weights of each level and checked for consistency. In the comparison matrix, λ_{max} is the maximum eigenvalue, and n is the matrix of the order of comparison. The eigenvector value corresponding to λ_{max} of the judgment matrix represents the weight *Wi* of the importance of each factor of this level to a factor of the previous level. Finally, we calculated the normalized weight coefficient *W* using the following formula

$$W_i = w_i' / \sum w_i' \qquad (3)$$

We calculated the combined weight coefficient C_i of each index by continuous multiplication and the comprehensive score index GI using the following formula

$$GI = \sum C_i \cdot P_i \qquad (4)$$

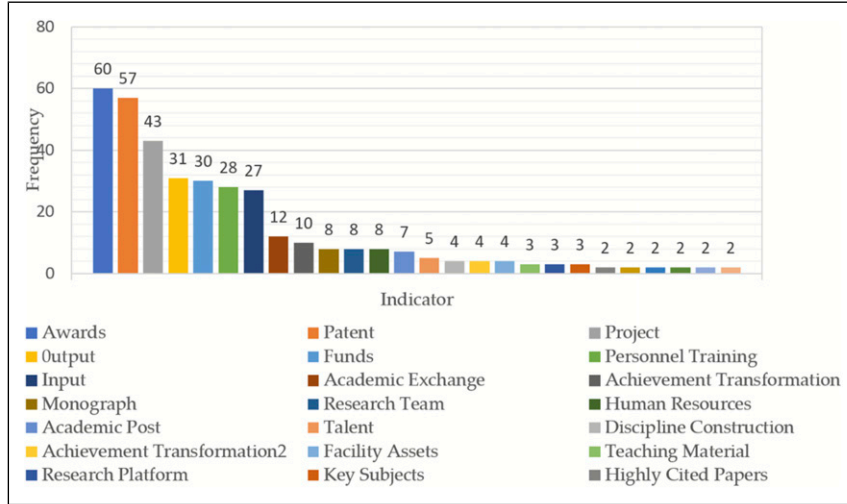where P_i is the measured value of the ith indicator.

After normalizing the weight coefficients, we calculated the consistency indices *(CIs)*. The calculation formula for *CI* is expressed as follows

$$CI = \frac{\lambda_{max} - m}{m - 1}$$

$$\lambda_{max} = \sum_i^m \lambda_i / m \qquad (5)$$

Table 3. Corresponding RI Values to the 1–9 Orders of the Matrix.

Matrix Order	1	2	3	4	5	6	7	8	9
RI	0	0	.58	.90	1.12	1.24	1.32	1.41	1.45

Figure 1. Description of indicator frequency of scientific research evaluation index in China.

Table 4. Basic Information Description of Consulting Experts (n = 20).

Categories	Frequency (%)
Years of occupations	
20~ years	11 (55)
>30 years	9 (45)
Professional levels	
Junior level	0
Medium level	0
Sub senior level	3 (15)
Senior level	17 (85)
Education levels	
Bachelor	1 (5)
Master	5 (25)
Doctor	14 (70)
Research areas	
Health administrative management	2 (1)
Clinical medical and management	15 (75)
Public health	2 (1)
Hospital management	7 (25)
Years of management	
5~	0
10~	11 (55)
≥20	9 (45)

where m is the number of sub-targets of the tested level, λ_{max} is the maximum characteristic root, and λ_i is the characteristic root of the optimal matrix for pairwise comparison and judgment of the sub-targets of this layer.

The calculation formula for the consistency ratio (CR) is expressed as follows

$$CR = RI/_{CI} \qquad (6)$$

The value of CI reflects the consistency of the matrix; that is, the smaller the value, the higher the consistency. The CR test items were relatively limited to determine whether there was any logical confusion. CR<.1 can be regarded as no logical confusion and as having acceptable weights.[29]

The random index (RI) refers to the average random consistency index, and the RI values of orders 1–9 had a fixed corresponding value, as presented in Table 3.

Results

Constructing the Initial Evaluation Index System

Using "hospital," "Medical College," "Medical University," "scientific research," "evaluation," and "assessment" as keywords, we searched CNKI, Wanfang Data, VIP, CBM, WoS, PubMed, Scopus, and other databases. After removing the repetitive literature, 996 studies were obtained, and after excluding irrelevant literature, 141 studies were included. We sorted out and extracted the indicators of research performance by domestic scholars in China according to the literature frequency of the indicators. Subsequently, we counted the most commonly used indicators, as shown in Figure 1. Based on available studies, practical experiences in medical research, and results of the expert

Table 5. Distribution of expert authority coefficients.

Authority Coefficient	Frequency of First round	Frequency of second round
.75~	7	6
.85~	4	4
.95~1	9	10
Total	20	20

Table 6. Overall Authority Coefficient of Experts.

Rounds of Questionnaire	Ca	Cs	Cr
First	.90	.90	.90
Second	.92	.97	.94

Table 7. Coordination Coefficient and Chi-Square Value for the two Rounds of Expert Consultation.

Categories	Number	W	Chi-Square Value	P
First round				
Primary-level indices	7	.083	7.984	.239
Secondary-level indices	14	.126	24.594	.026
Tertiary-level indices	36	.189	99.223	<.001
	57	.183	143.15	<.001
Second round				
Primary-level indices	7	.154	14.753	.022
Secondary-level indices	16	.306	64.235	<.001
Tertiary-level indices	41	.210	109.273	<.001
	64	.262	197.715	<.001

consultation, we drafted the initial evaluation system that contained 4 primary-, 14 secondary-, and 46 tertiary-level indices.

Characteristics of Experts and Distribution of Questionnaires

The Delphi method highlighted the importance of how experts are selected. The 20 experts selected in this study were famous specialty directors and hospital managers. All of them had worked for more than 20 years, with an average working experience of 28.88 years. All experts ranked above the associate senior level, and 85% held senior titles. In terms of education, 95% had completed higher than a bachelor's degree, and 70% were doctors; the data are presented in Table 4.

Establishment of the Evaluation System and Indices

Expert positive coefficient. The 2 rounds of questionnaires were answered by the 20 experts surveyed. The RR was found to be 100%, indicating extremely high levels of involvement and attention by the experts in the research project.

Expert authority coefficient. Table 5 shows the authority coefficients of all experts were above .75, whereas the authority coefficients of experts in the first and second rounds were mostly above .90.

The number of rounds of experts' Cr, Ca, and Cs exceeded .90. The average value of the second round was higher than that of the first round, indicating a high degree of authority and relatively reliable research results. The overall authority coefficients of the 2 rounds of consultation with experts were .9 for the first round and .94 for the second round. The experts selected in this study had a high level of authority, good representativeness, and credible prediction accuracy for the index system, as presented in Table 6.

Consistency of expert consultation. Table 7 shows the coordination coefficients of the 2 rounds of Delphi expert consultations in this study. Based on the W values in the first round, the experts did not have a high level of agreement regarding the indices at all levels; thus, the second round of opinion consultation was necessary. In the second round, the W values of all indices exceeded those of the first round. The Ws values of the primary and secondary levels of indices were .154 and .306, respectively, ($P < .05$), and the W value of the tertiary

Table 8. Two Rounds of Expert Consultation on Assigning the Mean of Necessity and CV of Indices at All Levels.

Indicator Number	First round			Indicator Number	Second round		
	Mean	Standard Deviation	CV		Mean	Standard Deviation	CV
1	4.938	.250	.051	1	4.991	.027	.005
2	4.875	.342	.070	2	4.969	.101	.020
3	4.813	.403	.084	3	4.919	.251	.051
4	4.625	.806	.174	4	4.794	.384	.080
5	4.813	.403	.084	5	4.859	.339	.070
6	4.688	.479	.102	6	4.838	.344	.071
7	4.500	.730	.162	7	4.644	.472	.102
1.1	4.813	.403	.084	1.1	4.981	.054	.011
1.2	4.813	.403	.084	1.2	4.919	.251	.051
2.1	4.813	.403	.084	2.1	4.988	.050	.010
2.2	4.625	.619	.134	2.2	4.653	.592	.127
3.1	4.875	.342	.070	3.1	4.922	.251	.051
4.1	4.563	.727	.159	4.1	4.697	.444	.094
4.2	4.563	.629	.138	4.2	4.572	.482	.106
5.1	4.875	.342	.070	5.1	4.866	.339	.070
5.2	4.563	.629	.138	5.2	4.688	.574	.122
6.1	4.688	.602	.128	6.1	4.740	.414	.087
6.2	4.688	.479	.102	6.2	4.667	.450	.096
7.1	4.500	.730	.162	6.3	4.567	.458	.100
7.2	4.188	.981	.234	7.1	4.500	.483	.107
7.3	4.250	.931	.219	7.2	4.306	.470	.109
1.1.1	4.438	.814	.183	7.3	4.369	.492	.113
1.1.2	4.375	.806	.184	7.4	4.260	.671	.157
1.1.3	4.563	.727	.159	1.1.1	4.625	.466	.101
1.1.4	4.938	.250	.051	1.1.2	4.500	.606	.135
1.1.5	4.438	.892	.201	1.1.3	4.725	.444	.094
1.2.1	4.750	.577	.122	1.1.4	4.727	.454	.096
1.2.2	4.625	.719	.155	1.1.5	4.600	.455	.099
1.2.3	4.375	.806	.184	1.2.1	4.850	.339	.070
1.2.4	4.625	.619	.134	1.2.2	4.719	.570	.121
2.1.1	4.625	.719	.155	1.2.3	4.513	.616	.137
2.1.2	4.625	.619	.134	1.2.4	4.525	.611	.135
2.1.3	4.375	.885	.202	2.1.1	4.588	.482	.105
2.2.1	4.625	.619	.134	2.1.2	4.347	.695	.160
2.2.2	4.563	.727	.159	2.1.3	4.466	.610	.137
3.1.1	4.875	.342	.070	2.1.4	4.671	.448	.096
3.1.2	4.313	.793	.184	2.1.5	4.300	.678	.158
3.1.3	4.375	.957	.219	2.2.1	4.569	.602	.132
4.1.1	4.500	.894	.199	2.2.2	4.588	.608	.132
4.1.2	4.688	.873	.186	3.1.1	4.863	.338	.070
4.1.3	4.063	1.124	.277	3.1.2	4.494	.511	.114
4.2.1	4.688	.873	.186	3.1.3	4.344	.598	.138
4.2.2	4.125	1.025	.248	4.1.1	4.713	.443	.094
5.1.1	4.688	.704	.150	4.1.2	4.794	.397	.083
5.1.2	4.688	.602	.128	4.1.3	4.231	.553	.131
5.1.3	4.688	.602	.128	4.2.1	4.731	.439	.093
5.1.4	4.313	1.078	.250	5.1.1	4.794	.397	.083
5.2.1	4.250	.775	.182	5.1.2	4.731	.439	.093
6.1.1	4.813	.544	.113	5.1.3	4.775	.404	.085
6.1.2	4.563	.629	.138	5.1.4	4.425	.480	.108

(continued)

Table 8. (continued)

Indicator Number	First round			Indicator Number	Second round		
	Mean	Standard Deviation	CV		Mean	Standard Deviation	CV
6.1.3	3.563	1.263	.355	5.2.1	4.481	.489	.109
6.2.1	4.625	.719	.155	6.1.1	4.669	.469	.100
6.2.2	4.438	.727	.164	6.1.2	4.569	.602	.132
7.1.1	4.750	.577	.122	6.2.1	4.719	.437	.093
7.1.2	4.533	.743	.164	6.2.2	4.625	.466	.101
7.2.1	4.375	.885	.202	6.3.1	4.300	.592	.138
7.3.1	4.375	.885	.202	6.3.2	4.367	.481	.110
				7.1.1	4.775	.399	.084
				7.1.2	4.631	.473	.102
				7.2.1	4.372	.496	.113
				7.3.1	4.363	.486	.111
				7.4.1	4.387	.797	.182
				7.4.2	4.033	.667	.165
				7.4.3	4.700	.414	.088
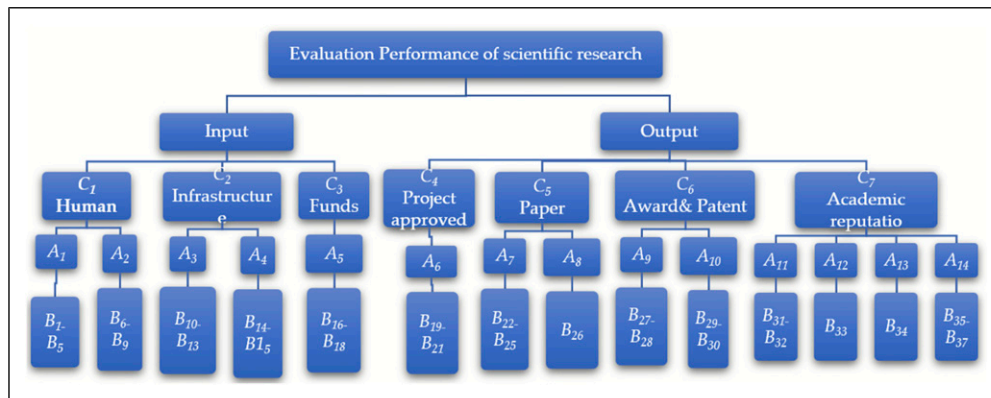
Figure 2. Hierarchical structure of the research evaluation index system for hospitals.

Table 9. Primary-Level Index Judgment Matrix and Weight G.

	C_1	C_2	C_3	C_4	C_5	C_6	C_7
C_1	1	2	2	2	2	2	3
C_2	1/2	1	2	2	2	2	3
C_3	1/2	1/2	1	2	2	2	2
C_4	1/2	1/2	1/2	1	1/2	1/2	2
C_5	1/2	1/2	1/2	2	1	1/2	2
C_6	1/2	1/2	1/2	2	2	1	2
C_7	1/3	1/3	1/2	1/2	1/2	1/2	1

level was .21 ($P < .001$), all of which met the statistical requirements. Thus, the results were deemed acceptable.

Concentration of expert opinions and coefficient of variation. In the first round of expert consultation, the maximum mean value of the necessity for all indices was 4.938, with a minimum of 3.563, whereas the maximum coefficient of variation was .355,

with a minimum of .051. In the second round, the maximum mean value of the necessity of all indices was 4.99, with a minimum of 4.033 (ie, all greater than 4), whereas the maximum CV was .182, with a minimum of .005, and the mean value was .101. Except for 5 indices, the CV for the other indices was less than .15, as presented in Table 8, indicating a better degree of coordination, more concentrated opinions, and a higher degree of recognition among the experts.

The coordination coefficient from the second round of expert evaluation was higher than that in the first round. For the second round, the average value of necessity for all indices exceeded 4, whereas the CV was less than .15, except for five of the indices. These values indicate that the experts agreed on the index system. Therefore, the index system has good rationality and practicality.

Modifications to Some Indices

The revision of the indices was based on statistical results, combined with expert opinions. We deleted 2 indices with

Table 10. Normalized Weights of Indices at All Levels.

Primary-Level Index	Wa	Secondary-Level Index	Wb	Normalized-Wb	Tertiary-Level Index	Wc	Normalized-Wc
C1	.249	A1	.667	.166	B1	.136	.023
					B2	.087	.014
					B3	.257	.043
					B4	.339	.056
					B5	.180	.030
		A2	.333	.083	B6	.410	.034
					B7	.321	.027
					B8	.118	.010
					B9	.151	.013
C2	.205	A3	.667	.137	B10	.269	.037
					B11	.190	.026
					B12	.420	.057
					B13	.121	.017
		A4	.333	.068	B14	.333	.023
					B15	.667	.046
C3	.158	A5	1	.158	B16	.626	.099
					B17	.238	.038
					B18	.137	.022
C4	.087	A6	1	.087	B19	.403	.035
					B20	.444	.039
					B21	.153	.013
C5	.107	A7	.751	.080	B22	.292	.023
					B23	.187	.015
					B24	.413	.033
					B25	.107	.009
		A8	.249	.027	B26	1	.027
C6	.130	A9	.667	.087	B27	.667	.058
					B28	.333	.029
		A10	.333	.043	B29	.667	.029
					B30	.333	.014
C7	.064	A11	.391	.025	B31	.667	.017
					B32	.333	.008
		A12	.138	.009	B33	1	.009
		A13	.276	.018	B34	1	.018
		A14	.195	.012	B35	.320	.004
					B36	.122	.002
					B37	.559	.007

W: Index weight.

relatively low research quality: the number of municipal science and technology awards and that of newly approved funds by enterprises. In the definition of indices, advanced specifications were recognized at the national level. To highlight the research transformation, we revised "output" to "awards and patents." Considering the significance of international exchange and cooperation on academic reputation, we added the indices of the following: "number of international academic conferences held," "number of participants in international academic conferences," and "number of international cooperation projects and funding" under the academic reputation index.

Establish hierarchical structure. To establish the goals, rules, and scheme layer relationships, a hierarchical structure, as shown in Figure 2, was developed and divided into 4 layers. The first layer was the target layer: the evaluation performance of scientific research. The second layer was the primary-level index layer (C_i) with 7 evaluation dimensions. The third layer includes several aspects corresponding to each dimension and belongs to the secondary index layer (A_i). The fourth layer included specific measurement indices for each aspect that belonged to the tertiary-level index layer (B_i).

Constructed the judgment matrix of all levels. According to the average value assigned by the experts to the importance of the index, a judgment matrix was constructed for all levels of indices using a scale of 1–9, as shown in Table 9.

Weight value of indices. Table 10 shows the normalized weights of the indices at three levels. The consistency test results showed that all of the *CR* values were below .1. This finding suggests that the opinions of the experts were relatively consistent, which can be regarded as no logical confusion, and each weight was acceptable (W: Index weight).

Discussion

According to production theory, we constructed an evaluation index system with 2 dimensions— research input and output. To establish a quantitative evaluation index system, we conducted a literature review, a two-round questionnaire survey of experts, statistical analyses, and credibility verification of the results through a consistency evaluation test. The final index included 7 primary-, 14 secondary-, and 37 tertiary-level indices relevant to research-oriented hospitals.

The statistical results showed that among the primary-level indices, the weight of human input was the highest, followed by infrastructural input. Meanwhile, the weight of the number of outstanding talents, graduate students, and post-doctorate staff of the tertiary-level indices was high, in line with strategic deployment by the Chinese government, such as the formulation of plans for training outstanding talent, employment of graduate students as the main research force, and establishment of post-doctoral mechanisms in universities. The scientific research platform was among the infrastructural inputs in the secondary-level indices with the highest weight. The highest-weighted tertiary-level indices relevant to infrastructural input were laboratory and sample banks, which implies laboratory support in research. Moreover, there is increasing importance given to sample banks for clinical research. At the project level, national projects ranked better than provincial projects based on the calculated weights. The total number of citations and the H index were given the highest weights, which reflected the trend of emphasizing "quantity" rather than "quality." In summary, the higher the level of indicators, the higher the weight. Hospital health policy makers should realize that the evaluation of scientific research pays more attention to the level and quality of output. In addition, talent input is the key indicator, and talent level as the most important factor determines the performance of scientific research.

This system conformed to the status and development trend of scientific research in research-oriented hospitals, which is consistent with related theories and practices. Therefore, it is scientific, practical, and reasonable. Subsequently, we will carry out case study, adopt the optimized EDM model to evaluate the performance of multiple hospitals or hospital specialties by the index system, and find out the existing problems and improvement methods for scientific research performance.[30,31]

This study had some limitations. For the construction of the index system, the Delphi method was used. This method has a subjective problem primarily based on the opinions of experts. In the selection of indices, we referred to foreign evaluation indices as there is no clear domestic definition for some excluded indices. Therefore, the method of constructing the evaluation index system and the selection of indices need to be further investigated. Subsequently, we will conduct follow-up research, conduct annual evaluations, and perform a retrospective study in 3–5 years to explore the correlation analysis of measures and improve scientific research performance. These additional methods would verify the scientific and rational nature of the evaluation model.

The constructed evaluation index system of the scientific research performance of research-oriented hospitals is aligned with the development trend of science and technology in China and with relevant theories and practices. As a special institution, research-oriented hospitals in developing countries have experienced rapid growth in recent years. Thus, the results of this study can be applied to the scientific research evaluation of other hospitals and guide scientific research investment and resource allocation to improve the quality of scientific research.

ORCID iD

Zhi Zeng https://orcid.org/0000-0002-1517-7648

References

1. İlgün G, Konca M. Assessment of efficiency levels of training and research hospitals in Turkey and the factors affecting their efficiencies. *Health Policy and Technology*. 2019;8:343-348.
2. Dowd B, McKenney M, Elkbuli A. The impact of COVID-19 pandemic on medical school admissions: challenges and solutions. *J Surg Res*. 2021;258:213-215.
3. Evaluating the technical efficiency of hospitals providing tertiary health care in turkey: an application based on data envelopment analysis. *Hosp Top*. 2021;99(2):49-63.
4. Schapper CC, Dwyer T, Tregear GW, Aitken M, Clay MA. Research performance evaluation: the experience of an independent medical research institute. *Aust Health Rev*. 2012;36:218-223.
5. Herrmann-Lingen CE, Brunner E, Hildenbrand S, Loew TH, Raupach T, Spies C, et al. Evaluation of medical research performance–position paper of the association of the Scientific medical societies in Germany (AWMF). *German med sci : GMS e-journal*. 2014;12:Doc11.
6. Alawi S, Bohr S, Stromps J, Alharbi Z, Pallua N. Eine analyse der forschungsaktivität universitärer kliniken für plastische

chirurgie im nationalen vergleich: eine systematische 5-jahres evaluation der forschungsaktivität in Deutschland. *Handchir Mikrochir Plast Chir.* 2016;48:73-77.

7. Praus P. Statistical evaluation of research performance of young university scholars: a case study. *Transinformação.* 2018;30: 167-177.

8. Bornmann L, Leydesdorff L. Count highly-cited papers instead of papers with h citations: use normalized citation counts and compare "like with like"! *Scientometrics.* 2018;115:1119-1123.

9. Weiss AP. Measuring the impact of medical research: moving from outputs to outcomes. *Am J Psychiatr.* 2007;164:206-214.

10. De Witte K, Rogge N. To publish or not to publish? On the aggregation and drivers of research performance. *Scientometrics.* 2010;85:657-680.

11. Caminiti C, Iezzi E, Ghetti C, De' Angelis G, Ferrari C. A method for measuring individual research productivity in hospitals: development and feasibility. *BMC Health Services Research.* 2015;15:468.

12. Li X, Hao JY. Construction of an evaluation index system for determining the academic impact of military medical scholars. *J Roy Army Med Corps.* 2018;164:164-169.

13. Wu YQ, Liang GW, Li Y, Wang DL, Liu HX, Al e. Construction of evaluation index system of hospital scientific research management information system. *Chinese J Med Res Manag.* 2017;30:217-219+26.

14. Liang J, Zhang Z, Fan L, et al. A comparison of the development of medical informatics in china and that in western countries from 2008 to 2018: a bibliometric analysis of official journal publications. *J Healthc Eng.* 2020;2020: 8822311.

15. Pan LM, Zuo JX, XI B, et al. Evaluation of hospital scientific research and analysis of influencing factors. *Chinese J Med Res Manag.* 2014;27:253-256+66.

16. Ye Q, CX Z. Investigation on the output of scientific research in a comprehensive tertiary hospital. *PLA Hospital Management Journal.* 2016;23:431-434.

17. Li J, Wu D, Li J, Li M. A comparison of 17 article-level bibliometric indicators of institutional research productivity: evidence from the information management literature of China. *Inf Process Manag.* 2017;53:1156-1170.

18. Muhammad A, Shaikh A, Naveed QN, Qureshi MRN. Factors affecting academic integrity in e-learning of saudi arabian universities. an investigation using delphi and AHP. *Ieee Access.* 2020;8:16259-16268.

19. Zakaria R, Ahmi A, Ahmad AH, Othman Z. Worldwide melatonin research: a bibliometric analysis of the published literature between 2015 and 2019. *Chronobiol Int* 2020;38:27-37.

20. von Thenen M, Frederiksen P, Hansen HS, Schiele KS. A structured indicator pool to operationalize expert-based ecosystem service assessments for marine spatial planning. *Ocean Coast Manag.* 2020;187:105071.

21. Shengbin Y, Hongyun Z. Construct a credit evaluation framework of E-commerce. *Adv Mater Res.* 2013;765-767 13704.

22. Soong JT, Poots AJ, Bell D. Finding consensus on frailty assessment in acute care through Delphi method. *Bmj Open.* 2016;6:e012904.

23. Skroumpelos A, Zavras D, Pavi E, Kyriopoulos J. Recommending organized screening programs for adults in Greece: A Delphi consensus study. *Health Pol.* 2013;109:38-45.

24. Khajehaminian MR, Ardalan A, Hosseini Boroujeni SM, Nejati A, Ebadati OM, Aghabagheri M. Prioritized criteria for casualty distribution following trauma-related mass incidents; a modified delphi study. *Archives of academic emergency medicine.* 2020;8:e47.

25. Torrecilla-Salinas CJ, De Troyer O, Escalona MJ, Mejías M. A delphi-based expert judgment method applied to the validation of a mature agile framework for web development projects. *Inf Technol Manag.* 2019;20:9-40.

26. Evans M, Farrell P, Mashali A, Zewein W. Critical success factors for adopting building information modelling (BIM) and lean construction practices on construction mega-projects: a delphi survey. *J Eng Des Technol* 2020;19:537-556.

27. Zhang WQ, Xi ZL. Application of delphi method in screening of indexes for measuring soil pollution value evaluation. *Environ Sci Pollut Res.* 2020;28:6561-6571.

28. Saaty TL. Decision-making with the AHP: why is the principal eigenvector necessary. *Eur J Oper Res.* 2003;145:85-91.

29. Franek J, Kresta A. Competitive strategy decision making based on the five forces analysis with ahp/anp approach. In: Proceedings of the 11th International Conference on Liberec Economic Forum, 16th–17th September 2013, Sychrov Czech republic, EU, 2013:135-145.

30. Li Y, Li J, Li B, et al. Factors associated with the research efficiency of clinical specialties in a research-oriented hospital in China. *PLoS One.* 2021;16:e0250577-e.

31. Rouyendegh BD, Oztekin A, Ekong J, Dag A. Measuring the efficiency of hospitals: a fully-ranking DEA-FAHP approach. *Ann Oper Res.* 2019;278:361-378.