

Establishing a balanced scorecard measurement system for integrated care organizations in China

Xiaobo Wu¹ | Sihan Li¹ | Ning Xu¹ | Dong Wu¹ | Xinli Zhang² 

¹School of Management, Zhejiang University, Hangzhou, China

²Business School, Sichuan University, Chengdu, China

Correspondence

Xinli Zhang, Business School, Sichuan University, Chengdu 610065, China.
Email: zhangxinli@scu.edu.cn

Funding information

Sichuan University, Grant/Award Number: 2018hhs-54; Health and Family Planning Commission of Sichuan Province, Grant/Award Number: 17ZD047; Zhejiang Soft Science Research Project, Grant/Award Number: 2017C35G2010204; National Key Research and Development Program of China, Grant/Award Number: 2017YFB1400601; National Natural Science Foundation of China, Grant/Award Number: 71832013

Summary

To address, among other issues, the regional and international challenges of the heavy health care burden caused by an aging population, integrated care organizations (ICOs) were proposed at the end of the 20th century for health care delivery. However, the implementation of ICOs has not progressed smoothly, and the current results have not eliminated the imbalance of medical service capabilities among hospitals of different levels. To make up for the deficiency in the current evaluation system at ICOs and offer suggestions for improved sustainable health planning and management, this study establishes a balanced scorecard based on a comprehensive measurement system valid for a Chinese ICO by surveying the staff at the West China Hospital ICO. This study collected valid responses from 216 professional staff members at the ICO via questionnaires. K-means clustering and the coefficient of variation method were used to evaluate the weights of the first- and second-level indicators. The results show the importance ranking of the core perspectives of the ICO balanced scorecard in the following order: patient, internal process, learning and growth, and financial. The weight-based analysis identified the importance ranking of all indicators and pointed to the areas that require close attention in future ICO planning and management.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. The International Journal of Health Planning and Management Published by John Wiley & Sons Ltd

KEYWORDS

balanced scorecard, China, evaluation indicator, integrated care organization (ICO), weight-based analysis

1 | INTRODUCTION

1.1 | Integrated health care: a worldwide phenomenon

Integrated health care is a relatively new trend in international health care reform. However, over two decades ago, in 1996, the World Health Organization (WHO) published *Integration of Health Care Delivery* to suggest the establishment of such a system of regional health care that would focus on the integration of primary health care providers, including the full integration of hospitals and their services.¹ Since then, the Institute of Medicine has published *Crossing the Quality Chasm: A New Health System for the 21st Century* and indicated that the integration and coordination of all parts of the medical system could generate more health benefits.²

Globally, many countries face difficulties and pressure delivering effective medical services, especially in developing countries such as China. The severe imbalance (see Figure 1) between the limited supply of medical service resources and the huge demand for them makes health care service delivery a major challenge today.

One of problems is the fact that health care services, which represent a large proportion of health system resources, are often concentrated in large or top-class hospitals and are difficult for patients to access. At the same time, the aging of the population has increased medical demand, and people are expecting better treatment. According to WHO statistics, the global population has been aging rapidly.³ Between 2000 and 2050, the proportion of the population over 60 is expected to double, increasing from 11% to 22%. At the same time, the absolute number of those over 60 is estimated to increase from 0.605 thousand million to 2 thousand million.

1.2 | Characteristics of integrated care organizations

Among the new trends in development and reform in international health care, the establishment of integrated care organizations (ICOs) and the legitimate integration of health resources have been proposed to deal with the mismatch in the supply and demand of medical services. The increasing attention on integrated health care may help build a more reasonable health care system model as determined by policymakers, managers, and practitioners.

ICOs can be found in many different countries but with different aims and features. In China, ICOs are seen as important initiatives to extend health reform.⁴ Integrated delivery networks (IDNs), which prevail in the United

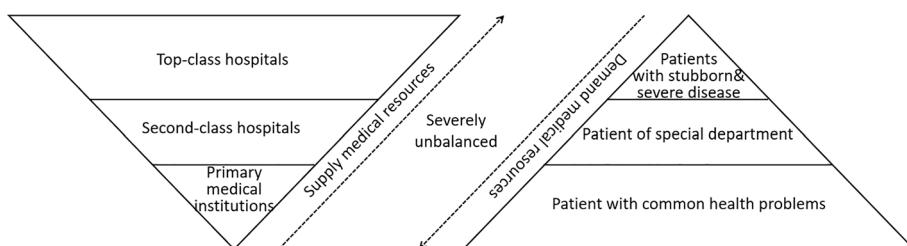


FIGURE 1 Severe unbalance between supply and demand of medical resources

States, link together health care organizations or medical staff at all levels to form medical service networks.^{5,6} These are aimed at helping patients avoid discontinuous medical care or dependency on emergency services as such patients usually encounter the most fragmented forms of medical care and need lower costs.

Integrated care networks (ICNs) used in England are aimed at coping with the financial constraints on National Health Services (NHS) in the past and the increasing pressure created by population growth and aging.^{7,8} ICNs are expected to coordinate general practice, community, and hospital service to satisfy the demand of patients, especially those with severe diseases.

Compared with elderly population problems in the United States and England, China is facing additional problems that include a large number of elderly people in need, a rapid rate of aging, and the fact that aging preceded industrialization. Thus, in China, solving these specific problems through ICOs may be trickier. The medical system involving three main levels of medical institutions and organizations has been perfected to a certain extent in developed countries such as the United States and England. As a result, their ICOs aim to enhance collaboration and improve medical services for patients. However, the capability and medical resources of primary medical and health care institutions in China are far from the top-class public hospitals. Chinese ICOs (Figure 2) are looking to establish networks involving all levels of hospital and medical institutions with the top-class hospitals at their core. Through the establishment of information-sharing platforms and the syncing of high-quality medical resources, the homogeneity of medical examinations throughout the consortium can be achieved. Thus, the overall medical service capacity can be improved. Further, closed-service, business, and value chains can be attained.

1.3 | Balanced scorecard-based performance measurement

Since the end of the 20th century, a variety of ICOs have emerged without any effective performance measurement. In addition, although seemingly integrated, all of these failed due to a lack of long-term strategic thinking.⁹ In this context, a performance measurement is seen as playing a key role in evaluating and uncovering the ICO system's main problems.

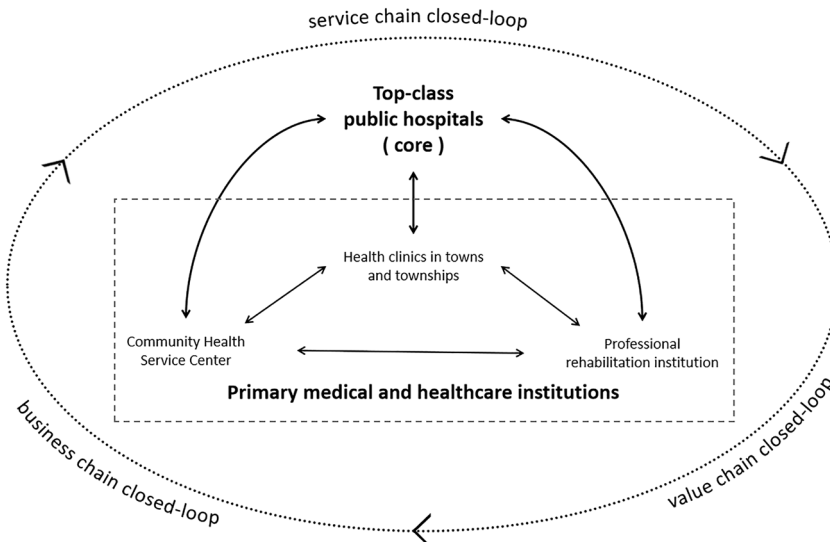


FIGURE 2 The integrated care organization model in China

Since Kaplan and Norton¹⁰ initially introduced the balanced scorecard (BSC), a popular performance management system consisting of four measurement perspectives (financial, customer, internal business process, and learning and growth), numerous articles have appeared on its development and application. It is a holistic method that converts an organization's mission and strategy into comprehensive performance measures that provide the basis for a successful strategic measurement and management system.¹¹ To date, the BSC has been applied successfully worldwide in many organizations including government, private, and nonprofit organizations.¹²

Today, the BSC is increasingly popular among medical organizations and institutions as well. As evidence, in a study assessing the performance and condition of the Tria Dipa Hospital in Indonesia from the four core BSC perspectives (except for learning and growth), the hospital was found to perform well on all other counts. The study also detailed the dissatisfaction of medical staff in the hospital.¹³ In Australia, the development of a performance measurement specific to Australian nonprofit health care organizations was first undertaken (into an application domain) by validating a nonprofit version of the BSC.¹⁴ In another example, researchers used the BSC to evaluate the strategic value of the implementation of cloud computing solutions at a Saudi hospital.¹⁵ As for China, in 2000, the BSC began to be used in health care engendering a wide range of research and applications.¹²

The wide application of the BSC reflects a change in management approach by policymakers and managers. That is, in the information society, the traditional method of performance management is not comprehensive enough and needs more balance. This includes balancing financial and nonfinancial indicators, long- and short-term objectives, outcome and motivational indicators, internal and external organizational groups, and leading and lagging indicators. Similar to the way the BSC is used throughout the world today to support high-level corporate strategy, it can be used by organizations to support specific strategies, such as sustainability, and, thus, to link sustainability objectives with appropriate actions and performance outcomes.¹⁶ Therefore, the application of the BSC to ICO performance measurement offers significant value by enabling balanced and sustainable ICO planning and management.

1.4 | Study objective

The objective of our study is to establish a comprehensive evaluation indicator system valid for Chinese ICOs from the perspective of experienced executives who have insights into ICO operations. Further, the indicator system will reflect the characteristics of the ICOs in China and offer guidance and be of practical significance for their sustainable planning and management.

From the perspective of quantity, Chinese ICOs have attained a certain level of success. However, whether the ICOs bring benefits to their core medical hospitals and institutions remains under discussion. Hence, a comprehensive performance evaluation system is needed to test the operational effectiveness of these ICOs. ICOs face some unique challenges in adapting the BSC to their environment. BSC theory is traditionally applied to both private corporations and nonprofit organizations such as public hospitals. Although the degree of emphasis on the financial, customer, internal process, or learning and growth perspectives depends on the nature of the organization, the organizations are considered single entities with unified strategic targets.

However, due to the coexistence of top-class public hospitals and primary medical and health care institutions comprising the ICO, it is unreasonable to recognize ICOs as simple single organizations. Large gaps exist between top-class public hospitals and primary medical and health care institutions, such as their health service capabilities and the bottleneck faced. Consequently, the ICO can be regarded as a cross-organizational system that is required to balance the conflicts and benefits among multiple organizations. Furthermore, owing to the ICO's specialization, the order of importance among the four core perspectives (financial, customer, internal process, and learning and growth) may differ from the traditional order for profit or nonprofit organizations as well.

The study was structured as shown in Figure 3, and we present each of the study elements in this paper. The rest of the paper is organized as follows. In Section 2, we present our methods including our survey of ICO experts. In Section 3, we show the results of the weighting of the indicators. In Section 4, we discuss the results from theoretical and practical standpoints. Suggestions for ICOs are also given from the financial, patient, internal process, and learning and growth perspectives of the BSC. Finally, we summarize the key findings, strengths, and limitations of our study and point to possible directions for further research.

2 | METHODS

We began our study by searching the China policy database for the key word *ICO* and found that although *ICO* was referred to in the content of some policies, it was rarely referred to in policy titles. However, such policies have dramatically increased in 2017 compared with 2016 (from nine to 51). Through June 2018, there were more than 5000 Chinese ICOs.¹⁷ However, according to the Beijing Report on Social Development (2015-2016), only around one of 3000 outpatient emergency treatments in Beijing was referred¹⁸ through ICOs from 2013 to 2015. The report points out that the “contract rate” of ICOs has been the main assessment by government departments to date. However, first-visit rate and referral rate, which reflect the actual operations of the ICOs, have not been included, which is evidence of the need for other measures.

Subsequently, we identified the strategic targets of the ICOs according to the characteristics of top-class public hospitals and primary medical and health care institutions and the interactions between them. The fieldwork included investigation among experts across the field as well as questionnaires from the staff members of a large China ICO and survey analysis. Finally, by synthesizing the expert opinions, we establish the BSC-based performance evaluation indicator system for ICOs in China.

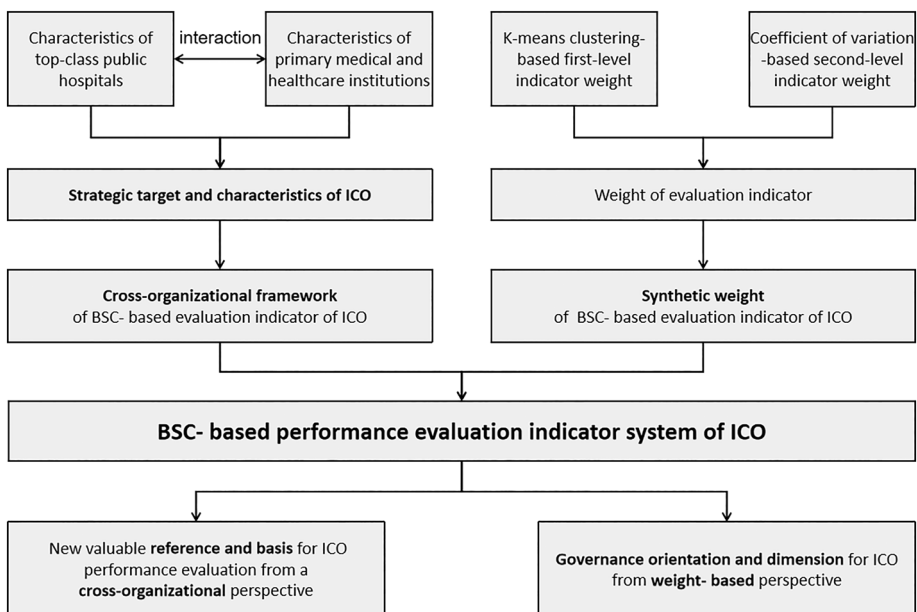


FIGURE 3 Structure of the study. BSC, balanced scorecard; ICO, integrated care organization

2.1 | The Delphi method

In this study, we used the Delphi method, a structured communication technique, originally developed as a systematic, interactive forecasting method, which relies on a panel of experts. We selected 216 experts with significant work experience and knowledge from the ICO of West China Hospital, one of most famous top-class public hospitals in China. We administered self-designed questionnaires to these experts who provided suggestions for modifying the indicator framework and graded the importance of the indicators. We then revised the indicator system according to the feedback. Figure 4 shows the revised version of the BSC-based evaluation indicator framework of an ICO in China.

The importance of the four first-level indicators scored and summed to 100%. The importance of each second-level indicator was categorized according to five levels ranging from 1, very unimportant, to 5, very important.

2.2 | Development of the BSC-based evaluation indicator system via case studies

We constructed an indicator system framework based on BSC theory. Guided by ICOs' strategic targets, the framework was created by consulting ICO experts and summarizing the interactions and problems that ICOs face. The design also includes research¹⁹⁻²³ from China and other countries on hospital performance evaluation.

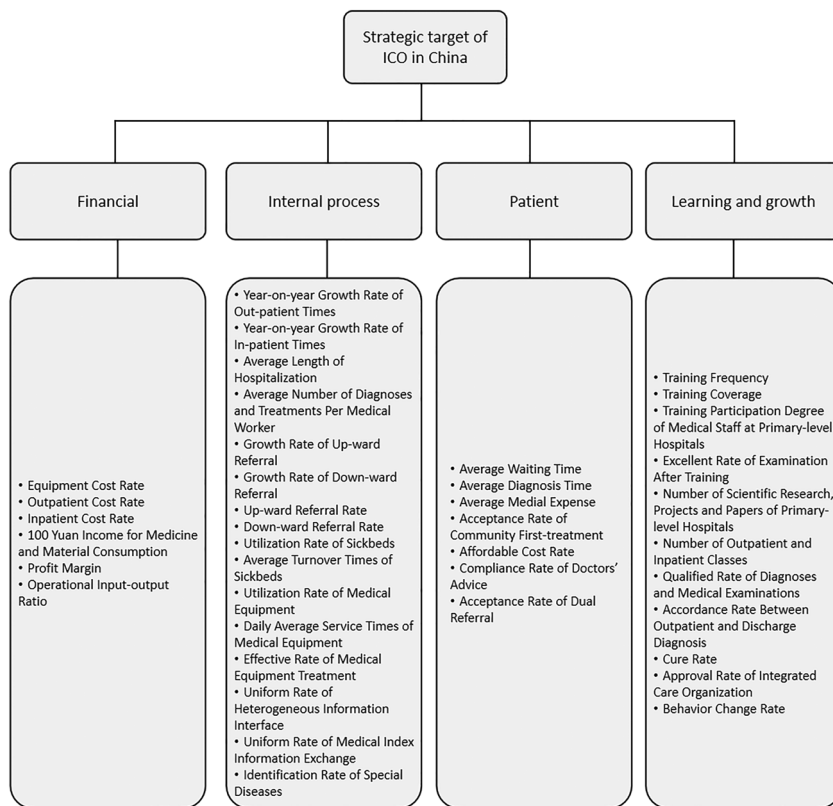


FIGURE 4 Balanced scorecard (BSC)-based evaluation indicator framework of integrated care organizations in China. ICO, integrated care organization

2.2.1 | Strategic target

Considering all the elements of the ICOs, the strategic development target is constructed with the causality of “input-output-purpose-goal” as shown in Figure 5, based on the logical framework application (LFA) method,²⁴ combined with China’s guiding opinions on promoting the establishment and development of ICOs and the problems facing medical providers.

2.2.2 | Principles of the indicator design

Cross-organizational synergies in ICOs require breaking down information barriers between organizations and decomposing the long service flow path at hospitals at all levels into a multinode path. This allows essential productive factors in each node, like registration, consultation, examination, treatment, hospitalization, and rehabilitation, to flow freely as “public objects” in the ICO and medical resources to be dispatched without restrictions. At the same time, for the purpose of realizing the integration and resharing of fragmented medical resources, the ICO reorganizes and re-innovates across these nodes. Thus, in addition to following the general principles of a single organization’s performance evaluation indicators, such as objectivity, science, validity, and operability,²⁵ the ICO evaluation index system should be established from the cross-organizational perspective and based on other special principles as follows.

Pertinence

Although the mechanisms underlying ICOs have begun to be improved, their actual effects lack pertinence verification. Correct analysis and evaluation of the characteristics of an object are critical criteria for evaluation indicator selection. Thus, when selecting indicators, the pain and pressure points during the development of the ICO should be examined. For instance, although the ICO assists in the use of information multilaterally by creating public resources that break through multiorganizational boundaries, the lack of homogeneity in the information from the various organizations hinders its smooth flow and sharing. As a result, information synergy needs to be evaluated through pertinent indicators.

Completeness

ICOs include many stakeholders such as patients and medical staff at hospitals at all levels, involving dual referrals and other aspects of operational processes. Accordingly, indicators should be designed as multidimensional based on a single organization’s evaluation indicators combined with the changes the ICOs create for hospitals at all levels and the synergies in each node across organizations.

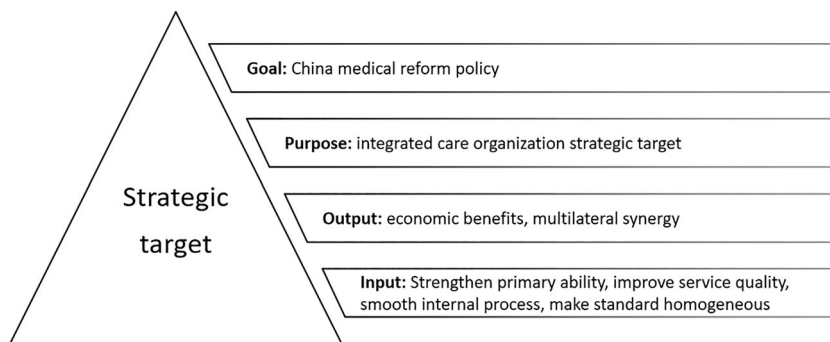


FIGURE 5 Logical framework application (LFA)-based strategy pyramid of integrated care organizations in China

Hierarchy

In the process of the index system establishment, the hierarchy of its structure should be constructed. That is, the strategic target of the ICO is decomposed into specific subtargets. The dimensions and the index layers are set up so that the indicators can be refined and quantized according to the data and guided by the dimensional layer.

Unity

In the analysis of a certain dimension, the indicators should relate closely to the core content of the strategic target and remain coordinated accordingly. Further, the impact of the indicators should be consistent, namely, the greater the index, the greater the impact. At the same time, each index should be standardized.

2.3 | Data collection (questionnaire)

2.3.1 | Sample source

On the basis of universality, representativeness, and relevance, we selected a large hospital and its ICO for our main case study. The questionnaires were completed by medical staff including managers, doctors, and paramedics. The questionnaires consist of two main parts shown in Tables 1 and 2, respectively.

In the study, the participants at the ICO were selected by random sampling. The questions covered the indicators to be graded. These were divided into four BSC dimensions: financial, process, patient, and learning and growth. The respondents were asked to score each dimension and the index based on their knowledge of the dimensions and indicators listed. Figure 6 presents the process used for data collection.

2.3.2 | Sample size estimation

The sample size was first calculated with an estimated error of not more than 6% at $P = 0.5$.

$$n = \frac{Z_{\alpha/2}^2 P(1-P)}{\Delta_p^2} = \frac{1.65^2 \times 0.5 \times (1-0.5)}{0.06^2} \approx 189, \quad (1)$$

where n is the sample size, $Z_{\alpha/2}^2$ is the critical value, and Δ_p^2 is the limit error.

TABLE 1 The first main part of the questionnaires

Total Score: 100%				
First-level Indicator	1st Respondent	2nd Respondent	3rd Respondent	...
Financial	__%	__%	__%	__%
Internal process	__%	__%	__%	__%
Patient	__%	__%	__%	__%

TABLE 2 The second main part of the questionnaires

Score of Each Indicator: Five Levels Ranging From 1, Very Unimportant, to 5, Very Important.				
Second-level Indicator	1st Respondent	2nd Respondent	3rd Respondent	...
Equipment cost rate	__ (1 to 5)	__ (1 to 5)	__ (1 to 5)	__ (1 to 5)
Outpatient cost rate	__ (1 to 5)	__ (1 to 5)	__ (1 to 5)	__ (1 to 5)
Inpatient cost rate	__ (1 to 5)	__ (1 to 5)	__ (1 to 5)	__ (1 to 5)
...	__ (1 to 5)	__ (1 to 5)	__ (1 to 5)	__ (1 to 5)

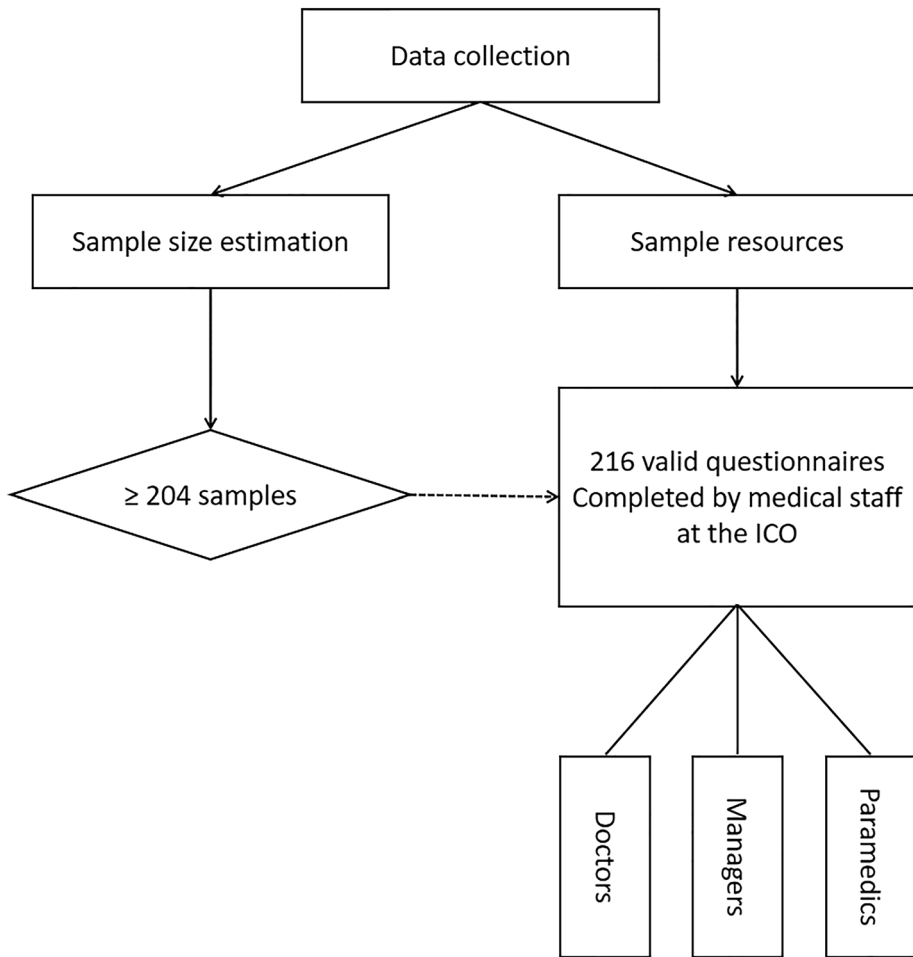


FIGURE 6 Process of data collection. ICO, integrated care organization

Considering the invalidity of some questionnaires, we needed to attain the number as calculated in formula (2):

$$N = n \times (1 + \omega) = 189 \times (1 + 8\%) \approx 204. \quad (2)$$

We collected 216 valid questionnaires, including 109 from the large hospital and 107 from the primary-level hospitals in its ICO.

2.4 | Screening and weight calculation of indicators

Figure 7 presents the process used for data analysis.

2.4.1 | Agreement among the expert suggestions

Due to the diversity of the 216 individuals we interviewed, we needed to test the agreement among these, as high consistency represents the high reliability of the results. The consistency of the scores of each indicator is defined by its coefficient of variation, which is used to explain the convergence among the internal sample data. The coefficient

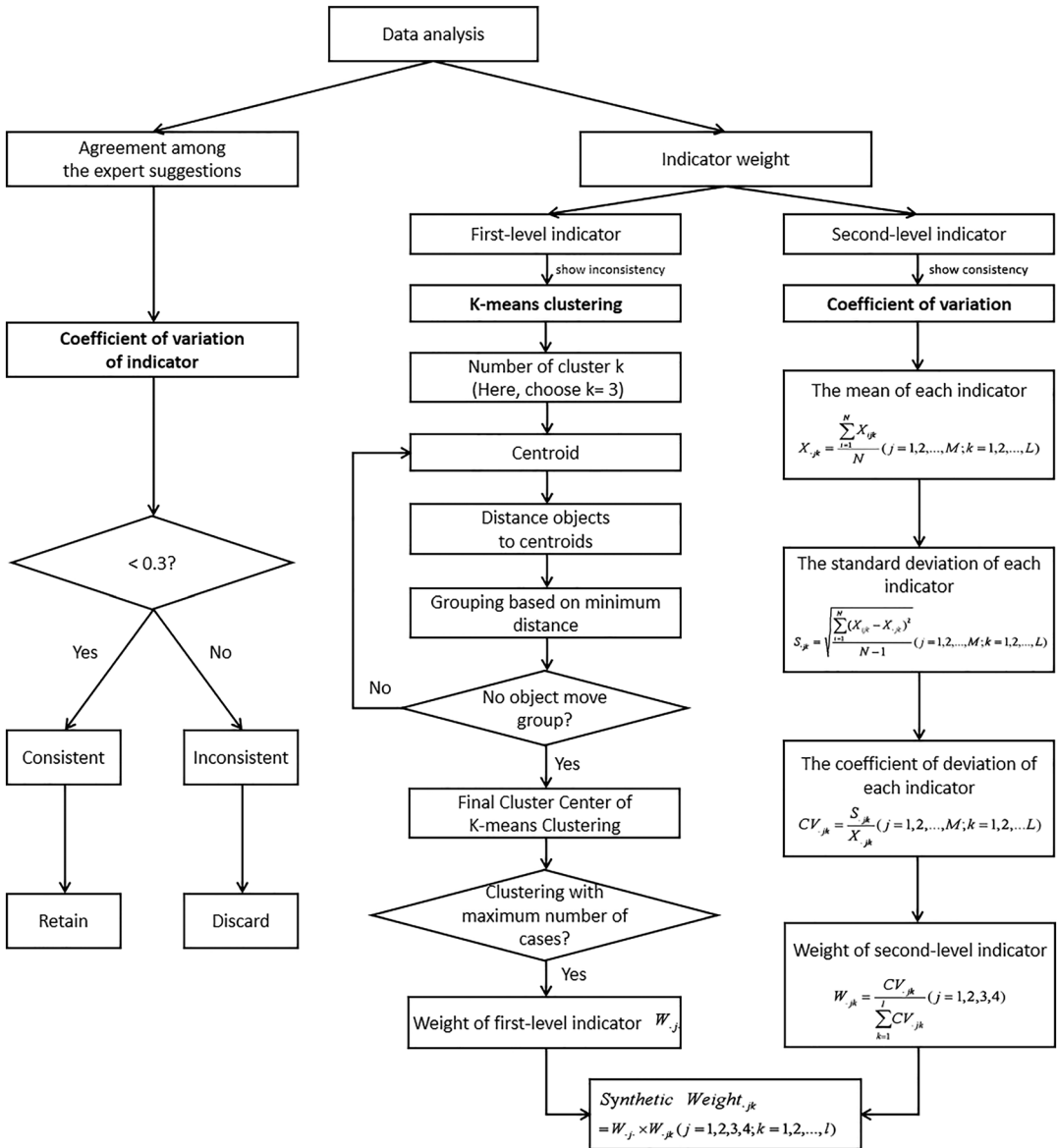


FIGURE 7 Screening and weight calculation of indicators

of variation is equal to the standard deviation of the sample divided by the mean. The smaller the coefficient of variation, the higher the consistency of the indicator, that is, the more the scores of the internal sample converge. Thus, mean, standard deviation, and coefficient of variation of the first-level and second-level indicators were calculated to assess consistency. At the same time, considering the overall sample, indicators can be regarded as consistent when the coefficient of variation is less than 0.3.

First-level indicators

To analyze the average level and agreement of the scores by the first-level indicators, the four BSC areas of emphasis, the mean, standard deviation, and coefficient of variation were calculated as shown in Table 3. As all the first-level

TABLE 3 Agreement on the first-level indicators

First-level Indicator	Financial	Process	Patient	Learning and Growth
Mean	23.52	24.44	27.19	24.85
Standard deviation	11.30	9.45	10.65	16.51
Coefficient of variation	0.48	0.39	0.39	0.66

indicators displayed high inconsistency, we adjusted the weight calculation method. That is, we chose the clustering method to ensure the consistency.

Second-level indicators

The mean, standard deviation, and coefficient of variation were calculated, as shown in Tables 4–7, to analyze the average level and concordance of the scores by the second-level indicators by perspective, such as equipment cost and outpatient cost (financial), length of hospital stay and utilization rate of sickbeds (process), average waiting time and diagnosis time (patient), and cure rate and training participation degree (learning and growth). Although some of the second-level indicators show a relatively high inconsistency, no coefficient of variation is higher than 0.3, which means there is no agreement among the suggestions for the second-level indicators.

2.4.2 | K-means clustering-based first-level indicator weight

The main objects of the weight analysis of the four perspectives are the scores provided by the 216 respondents on the basis of their degree of importance. However, due to the large coefficient of variation for the four first-level indicators, that is, the difference in the score distribution was significant, *K*-means clustering analysis²⁶ was used additionally to refine the weighting further. We used SPSS 20.0 to explore the respondents' common knowledge of each indicator implied by the data. In this case, we chose *K* as three, which means the data would be divided into three clusters and we would receive a new set of results, that is, three new cluster centers including a different number of cases in every operation.

Step 1:. Determination of the initial set of cluster centers.

SPSS 20.0 gives the initial cluster center after fast clustering analysis.

Step 2:. Determination of the final set of cluster centers.

The *K*-means clustering algorithm calculates the distance from each point to the cluster's center, divides each point into the nearest category, and determines a new cluster center by updating the coordinate mean of each point in each category. The above process is repeated until the cluster center does not move significantly.

TABLE 4 Agreement on the financial perspective

Second-level Indicators of the Financial Perspective	Equipment Cost Rate	Outpatient Cost Rate	Inpatient Cost Rate	100 Yuan Income for Medicine and Material Consumption	Profit Margin	Operational Input-Output Ratio
Mean	3.8241	3.8935	3.8981	3.8843	4.0926	4.0185
Standard deviation	1.1357	1.1300	1.0907	1.1455	0.9933	1.0868
Coefficient of variation	0.2970	0.2902	0.2798	0.2949	0.2427	0.2704

TABLE 5 Agreement on the internal process perspective

Second-Level Indicators of the Internal Process Perspective	Year-on-year Growth Rate of Out-patient Times	Year-on-year Growth Rate of In-patient Times	Average Length of Hospitalization	Average Number of Diagnosis and Treatment Per Medical Worker	Growth Rate of Upward Referral	Growth Rate of Downward Referral	Upward Referral Rate	Downward Referral Rate
Mean	4.0926	3.9722	3.8287	3.9398	3.9074	3.9213	3.8657	3.8935
Standard deviation	1.0743	1.1203	1.1425	1.0701	1.1251	1.1364	1.1475	1.1504
Coefficient of variation	0.2625	0.2820	0.2984	0.2716	0.2879	0.2898	0.2968	0.2955

Step 3:. Determination of the final cluster center.

By comparing the number of cases in the final three clusters, the cluster with the most cases is chosen as the basis of the weight calculation for the first-level indicator.

Step 4:. Determination of the weights of the first-level indicators.

The weight of each first-level indicator is the ratio of the final cluster center's coordinate value to the sum of the coordinate values of all cluster centers. In the study, the initial cluster number was three. After five iterations, the cluster center no longer moved. The final weight of the four first-level indicators was W_j ($j = 1, 2, 3, 4$).

2.4.3 | Coefficient of variation-based second-level indicator weight

Although the mean of the scores can evaluate the overall situation of the indicator, the coefficient of variation can compensate better for the defects of equalization. When dealing with weighting the second-level indicators, those under the same perspective were regarded as a whole. The means and standard deviations of the scores were calculated, and the coefficient of variation (CV) was calculated by the ratio of the standard deviation to the mean. Using the CV method, the index weight was determined by the ratio of the CV of a certain index to the sum of the CV of all the indicators; the BSC with the index weight was established as well. The specific calculation process is as follows.

Step 1:. Definition of variables.

The k th index score of the j th dimension of the i th respondent is defined as in formula (3).

$$X_{ijk} (i = 1, 2, \dots, N; j = 1, 2, \dots, M; k = 1, 2, \dots, L), \quad (3)$$

where N is the number of respondents, M is the number of first-level indicators (perspectives), and L is the number of second-level indicators for each perspective.

Step 2:. Calculation of mean, standard deviation, and coefficient of deviation.

The mean, standard deviation, and coefficient of deviation of each indicator are calculated, respectively, as in formulas (4) to (6).

$$X_{.jk} = \frac{\sum_{i=1}^N X_{ijk}}{N} (j = 1, 2, \dots, M; k = 1, 2, \dots, L), \quad (4)$$

TABLE 5 (Continued)

Second-Level Indicators of the Internal Process Perspective	Utilization Rate of Sickbeds	Average Turnover Times of Sickbeds	Utilization Rate of Medical Equipment	Daily Average Service Times of Medical Equipment	Effective Rate of Medical Equipment Treatment	Uniform Rate of Heterogeneous Information Interface	Uniform Rate of Medical Index Information Exchange	Identification Rate of Special Diseases
Mean	4.1157	4.0880	4.1898	4.1111	4.1157	4.1713	4.1528	4.1944
Standard deviation	1.0300	1.0146	0.9865	1.0146	1.0164	0.9661	0.9740	0.9592
Coefficient of variation	0.2503	0.2482	0.2354	0.2468	0.2470	0.2316	0.2345	0.2287

TABLE 6 Agreement on the patient perspective

Second-level Indicators of the Patient Perspective	Average Waiting Time	Average Diagnosis Time	Average Medial Expense	Acceptance Rate of Community First-treatment	Affordable Cost Rate	Compliance Rate of Doctors' Advice	Acceptance Rate of Dual Referral
Mean	4.0463	4.0417	4.0926	4.2593	4.0972	4.2870	4.1944
Standard deviation	1.1113	1.0797	1.0569	0.9824	1.0229	0.9001	1.0112
Coefficient of variation	0.2747	0.2671	0.2582	0.2306	0.2497	0.2100	0.2411

$$S_{\cdot jk} = \sqrt{\frac{\sum_{i=1}^N (X_{ijk} - X_{\cdot jk})^2}{N-1}} (j = 1, 2, \dots, M; k = 1, 2, \dots, L), \quad (5)$$

$$CV_{\cdot jk} = \frac{S_{\cdot jk}}{X_{\cdot jk}} (j = 1, 2, \dots, M; k = 1, 2, \dots, L). \quad (6)$$

Step 3:. Determination of the weights of second-level indicators.

According to the CV method, the indicator weight is achieved from the weighted mean as shown in formula (7).

$$W_{\cdot jk} = \frac{CV_{\cdot jk}}{\sum_{k=1}^L CV_{\cdot jk}} (j = 1, 2, 3, 4). \quad (7)$$

Step 4:. Determination of the synthetic weight.

The synthetic weight takes both first-level and second-level weights into account, calculated as in formula (8).

$$\text{Synthetic Weight}_{\cdot jk} = W_{\cdot j} \times W_{\cdot jk} (j = 1, 2, 3, 4; k = 1, 2, \dots, L). \quad (8)$$

3 | RESULTS

The case study had 216 respondents and 40 second-level indicators from four perspectives forming 8640 sample points, as shown as formula (9).

TABLE 7 Agreement on the learning and growth perspective

Second-level Indicators of Learning and Growth Perspective	Training Frequency	Training Coverage	Training Participation of Medical Staff at Primary-level Hospitals	Excellent Rate of Examination After Training	Number of Scientific Research, Project and Papers of Primary-level Hospitals	Number of Outpatient and Inpatient Teaching	Qualified Rate of Diagnosis and Medical Examination	Accordance Rate Between Outpatient and Discharge Diagnosis	Cure Rate	Approval Rate of integrated care organization	Behavior Change Rate
Mean	4.2593	4.3102	4.3750	4.2500	3.8287	4.1574	4.3194	4.2315	4.2778	4.3889	4.2778
Standard deviation	0.9288	0.8845	0.8425	0.8640	1.1303	1.0176	0.8865	0.9411	0.9483	0.8554	0.9483
Coefficient of variation	0.2181	0.2052	0.1926	0.2033	0.2952	0.2448	0.2052	0.2224	0.2217	0.1949	0.2217

$$x_{ijk} (i = 1, 2, \dots, 216; j = 1, 2, 3, 4; k = 1, 2, \dots, l). \quad (9)$$

The number of indicators in each dimension is shown in formula (10).

$$l = \begin{cases} 6, & j = 1 \\ 16, & j = 2 \\ 7, & j = 3 \\ 11, & j = 4 \end{cases} \quad (10)$$

3.1 | Weight of first-level indicators

In the study, the initial clustering number was three ($k = 3$). After five iterations, the maximum absolute coordinates of any center changed to 0.000, which means that the cluster center no longer moved. The process is shown in Table 8. The final set of cluster centers is shown in Table 9.

According to the clustering results shown in Table 10, the 157 cases in cluster 2 represent 73% of all effective cases, which indicates that over 70% of ICO work participants agreed on the degree of importance of the four perspectives. Thus, the final center of cluster 2 was taken as the basis for the weight determinations of the first-level indicators. After transforming the center values of each dimension of cluster 2 (shown in Table 10) into decimals, the weights of the first-level indicators are as shown Table 11.

According to the order of importance, the final result of the four perspectives (first-level indicator) can be logistically described as in Figure 8.

TABLE 8 Iterative record of *K*-means clustering^a

Iteration	Change in Cluster Center		
	1	2	3
1	37.326	40.089	27.845
2	3.307	1.145	3.133
3	1.734	0.554	0.000
4	1.198	0.353	0.000
5	0.000	0.000	0.000

^aConvergence occurs as there are no changes or minor changes in the cluster center. The maximum absolute coordinates of any center are changed to 0.000. The current iteration is five. The minimum distance between the initial centers is 83.857. Bold figures emphasize that this is the key result of this step and the basis for allowing the next step.

TABLE 9 Final cluster center of *K*-means clustering

Dimension	Cluster Center		
	1	2	3
Financial	39.26	19.99	6.55
Process	23.49	25.99	5.82
Patient	22.60	29.85	9.18
Learning and growth	14.66	24.17	78.45

Bold figures emphasize that this is the key result of this step and the basis for allowing the next step.

TABLE 10 Number of cases in each cluster

Cluster classification	1	47.000
	2	157.000
	3	11.000
Effective		215.000
Missing		0.000

Bold figures emphasize that this is the key result of this step and the basis for allowing the next step.

TABLE 11 Weight of each perspective

Perspective	Final Cluster Center	Weight W_j
Financial	19.99	0.1999
Process	25.99	0.2599
Patient	29.85	0.2985
Learning and growth	24.17	0.2417

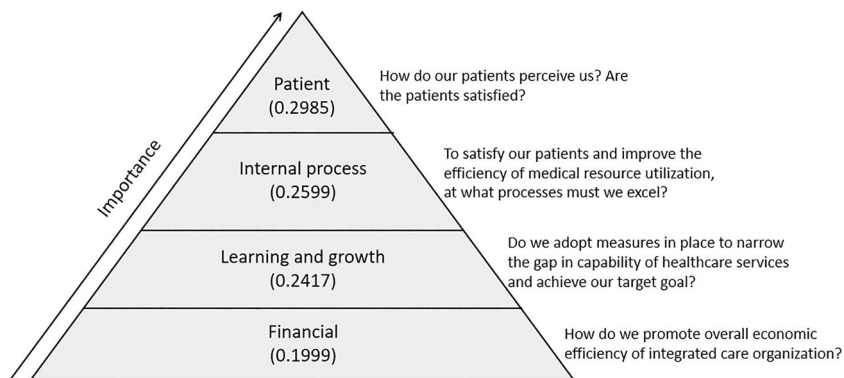


FIGURE 8 Balanced scorecard (BSC)-based importance hierarchy of integrated care organizations in China

3.2 | Weights of second-level indicators

The final BSC-based performance evaluation indicator system for the ICOs, with the index weights, is shown in Table 12.

4 | DISCUSSION

4.1 | Analysis of BSC-based performance evaluation indicator system

The weighted order of the four first-level indicators in the BSC-based performance evaluation system was as follows: patient, internal process, learning and growth, and financial. Each indicator had a different weight at different levels.

In the analysis of the agreement of the first-level indicators, the score distribution of learning and growth was the most discrete with the biggest deviation. Compared with the other three perspectives, this requires a long time to verify its validity, which reflects the ICOs' future development direction and upward trend. Learning and growth

TABLE 12 Weight of indicators

First-level Indicators	W_j	Second-level Indicators	W_{jk}	Synthetic weight _{ij}
Financial	0.1999	Equipment cost rate	0.1837	0.0367
		Outpatient cost rate	0.1710	0.0342
		Inpatient cost rate	0.1649	0.0330
		100 yuan income for medicine and material consumption	0.1781	0.0356
		Profit margin	0.1430	0.0286
		Operational input-output ratio	0.1594	0.0319
Process	0.2599	Year-on-year growth rate of out-patient times	0.0623	0.0162
		Year-on-year growth rate of in-patient times	0.0669	0.0174
		Average length of hospitalization	0.0726	0.0189
		Average number of diagnosis and treatment per medical worker	0.0644	0.0167
		Growth rate of upward referral	0.0683	0.0178
		Growth rate of downward referral	0.0688	0.0179
		Upward referral rate	0.0704	0.0183
		Downward referral rate	0.0701	0.0182
		Utilization rate of sickbeds	0.0594	0.0154
		Average turnover times of sickbeds	0.0589	0.0153
		Utilization rate of medical equipment	0.0559	0.0145
		Daily average service times of medical equipment	0.0586	0.0152
		Effective rate of medical equipment treatment	0.0586	0.0152
		Uniform rate of heterogeneous information Interface	0.0550	0.0143
		Uniform rate of medical index information exchange	0.0556	0.0145
		Identification rate of special diseases	0.0543	0.0141
Patient	0.2985	Average waiting time	0.1586	0.0473
		Average diagnosis time	0.1543	0.0461
		Average medial expense	0.1492	0.0445
		Acceptance rate of community first-treatment	0.1332	0.0398
		Affordable cost rate	0.1442	0.0430
		Compliance rate of Doctors' advice	0.1213	0.0362
		Acceptance rate of duel referral	0.1392	0.0416
Learning and growth	0.2417	Training frequency	0.0894	0.0216
		Training coverage	0.0841	0.0203
		Training participation degree of medical staff at primary-level hospitals	0.0789	0.0191
		Excellent rate of examination after training	0.0833	0.0201
		Number of scientific research, project and papers of primary-level hospitals	0.1272	0.0307
		Number of outpatient and inpatient teaching	0.1003	0.0242
		Qualified rate of diagnosis and medical examination	0.0841	0.0203
		Accordance rate between outpatient and discharge diagnosis	0.0911	0.0220
		Cure rate	0.0908	0.0219
		Approval rate of integrated care organization	0.0799	0.0193
		Behavior change rate	0.0908	0.0219

reflect a holistic view of development. However, except for ICO managers, most of the staff in the ICO tends to be confined to their daily work, thereby neglecting ICO development from a global perspective. Therefore, the different positions in the ICOs reflect a diverse understanding of learning and growth, which engenders the discrete score situation. This implies a lack of synergy in the understanding among the ICO work participants, which to some extent will influence the further development of the ICOs.

In the weight determination of first-level indicators, the weights of the patient, internal process, learning and growth, and financial perspectives decreased in turn. The weight of the patient perspective was the largest and that of the financial perspective the smallest, which represents the characteristics of the Chinese ICOs. That is, the weights of the first-level indicators emphasize "human-orientation." China is committed to

solving the problem of difficult and expensive access to medical care caused by the uneven distribution of medical resources from a social welfare perspective, with medical economic benefits relatively less of a consideration.

4.2 | Implications for further planning and management of the ICO

According to our study results, attention should be paid to the following aspects of ICO performance improvement.

4.2.1 | Implications from the financial perspective

In the agreement analysis, from the financial perspective, the mean of the indicators based on the strategic target of increasing economic benefits was relatively large, and the CV of that was small, while the mean of the indicators based on the target of controlling medical costs showed the opposite results. This illustrates that the ICO staff held divergent views on controlling medical costs and were more attentive to improving the profitability of the ICO.

4.2.2 | Implications from the internal process perspective

From the internal process perspective, the mean of indicators based on the strategic target of achieving homogeneity and mutual recognition was relatively large, and the CV of that was small, which points to the significant need for standardization to improve ICO internal processes, whereas the mean of the indicators based on the target of increasing referral efficiency showed the opposite results. This may be because ICO participants work in different quality hospitals, which means that participants from superior hospitals care more about indicators representing downward referrals (transfers from top-class hospitals to primary medical institutions or lower-level hospitals) and those from primary-level hospitals care more about indicators representing upward referrals (transfers from the primary medical institutions or low-level hospitals to top-class hospitals).

4.2.3 | Implications from the patient perspective

From the patient perspective, the mean of the indicators based on the target of enhancing patient acceptance of community first-treatment and dual referral was relatively large, and the CV of that was small; however, enhancing patients' satisfaction showed the reverse results. This may be because ICO workers focus more on directly related patient behaviors rather than patients' personal experience with the ICO medical services.

4.2.4 | Implications from the learning and growth perspective

From the learning and growth perspective, except for the index, based on the target of the amount of scientific research, projects, and papers at primary-level hospitals, the mean of other indicators was large, and the CV of that was small, which may be because ICO workers at primary-level hospitals place more emphasis on improvement in practical abilities rather than medical research right now.

4.3 | Strengths, limitations, and suggestions for future research

We believe this is a pioneering study of ICO BSC measurements and evaluations of collaborative performance of cross-organizational medical services. The ICO's performance evaluation index system is based on the actual situation among China's ICOs and a comprehensive consideration of the actual principles, methods, and operations. Based on the comprehensiveness of the BSC, it can reflect all aspects of the ICO's performance measurements, and the method is feasible as well as beneficial in the acquisition of practical information. The analysis of the indicator

weighting shows that when the BSC is applied to actual performance measurements, data-driven index weight calculation is necessary, and clustering analysis and the CV method can help in the BSC design.

However, this study has some limitations that point to areas of further study:

1. The BSC here is designed based on the practice at a large hospital in China, and the indicators are considered from the perspective of its characteristics. The specific indicators of the BSC require further scope of application.
2. The weight analysis is divided into two parts: first-level and second-level indicator layers. The process of determining the weights of the two levels is independent. The relationship between cross-dimensional indicators and the influence between the weighting in the two parts is not explained and is worthy of further study.
3. The *K*-means clustering algorithm and the CV method are used to determine the weights. Although the *K*-means clustering algorithm and the CV method have relatively high efficiency and flexibility, they are determined by the samples and driven by the sample data, which will change with different sample data. Thus, a problem still remains as to how to determine the reasonable initial clustering center during the *K*-means cluster calculation and how to improve the CV method to be more reliable. In follow-up research, because the weight is driven by the concrete data, the data processing method still needs improvement. However, the normal form of the index importance ranking given here has a large reference value.
4. Although the evaluation system reflects the general experience of internal experts in a larger hospital in China and has significant reference value, the index system has not yet been implemented in the hospital. At the same time, China ICOs are still in a semi-mature stage of development. Thus, there are some subjective elements and uncertain factors in the index system, and its validity still needs further testing. Based on the results of the system's practical application and the deviation between them and the actual quality improvements at ICOs, a follow-up study could integrate the results of the measurement evaluation of the indicators into the BSC to further improve the indicators.

AUTHOR CONTRIBUTION

The study was designed by X.W. in collaboration with all coauthors. The data were collected by S.L. and N.X. The first and final drafts were written by N.X. The paper was critiqued by S.L. The results were analyzed by X.W. and D.W. The research and key elements of the models were reviewed by X.W. The writing of the corresponding parts and the major revisions of this paper were completed by X.Z.

ETHICS STATEMENT

This article does not require ethics approval. There is no human involvement in this study. Consent from the patient is not required as this study reports an assessment of the integrated care organization situation overall.

ACKNOWLEDGEMENTS

The authors are grateful for the support from the West China Hospital and funds from the National Natural Science Foundation of China, the National Key Research and Development Program of China, Zhejiang Soft Science Research Project, Sichuan Health and Family Planning Commission, and Sichuan University.

The research is supported by the National Natural Science Foundation of China (Reference: 71832013), the National Key Research and Development Program of China (Reference: 2017YFB1400601), the Zhejiang Soft Science Research Project (Reference: 2017C35G2010204), the Health and Family Planning Commission of Sichuan Province (Reference: 17ZD047), and the Sichuan University (Reference: 2018hhs-54).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Xinli Zhang  <https://orcid.org/0000-0002-0954-793X>

REFERENCES

1. World Health Organization. *Integration of Health Care Delivery: Report of a WHO Study Group*. Geneva: World Health Organization; 1996 <http://www.who.int/iris/handle/10665/38408>.
2. Institute of Medicine. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: The National Academies Press; 2001.
3. World Health Organization. Ageing and life-course. <http://www.who.int/ageing/about/facts/en/>. Accessed September 30, 2014
4. Qiu GQ, Xu LZ. Construction of health alliance in China based on the perspective of British experience. *Modern Hosp*. 2016;16(12):1801-1803.
5. Schardt G, Nolan K. Expansion of an integrated delivery network (IDN) specialty pharmacy through targeted clinic engagement and focused administrative support. *Res Social Adm Pharm*. 2018;14(5):e10. <https://doi.org/10.1016/j.sapharm.2018.03.042>
6. Park PW, Dryer RD, Hegemangle R, et al. Cost burden of chronic pain patients in a large integrated delivery system in the United States. *Value Health*. 2016;16(8):1001-1011.
7. Mur-Veeman I, Hardy B, Steenbergen M, Wistow G. Development of integrated care in England and the Netherlands: managing across public-private boundaries. *Health Policy*. 2003;65(3):227-241.
8. Nurjono M, Shrestha P, Lee A, et al. Realist evaluation of a complex integrated care programme: protocol for a mixed methods study. *BMJ Open*. 2018;8(3):e017111. <https://doi.org/10.1136/bmjopen-2017-017111>
9. Chen GY, Feng ZY. Some cold thoughts on the new upsurge of the group of public hospitals. *Chongqing Med*. 2014;22:2960-2961.
10. Kaplan RS, Norton DP. The balanced scorecard—measures that drive performance. *Harvard Business Review*, January–February 1992:71.
11. Kaplan RS, Norton DP. Strategic learning & the balanced scorecard. *Strateg Leadersh*. 1996;24(5):18-24.
12. Gao H, Chen H, Feng J, et al. Balanced scorecard-based performance evaluation of Chinese county hospitals in under-developed areas. *J Int Med Res*. 2018;46(5):1947-1962.
13. Dewi NF, Santoso RK. The performance analysis of inpatient installation at Tria Dipa Hospital with balanced scorecard, 2013–2015. *KnE Soc Sci*. 2018;3(11):1566-1583. <https://doi.org/10.18502/kss.v3i11.2869>
14. Zahoor A, Sahaf MA. Investigating causal linkages in the balanced scorecard: an Indian perspective. *Int J Bank Mark*. 2018;36(1):184-207. <https://doi.org/10.1108/IJBM-09-2016-0128>
15. Alharbi F, Atkins A, Stanier C, Al-Buti HA. Strategic value of cloud computing in healthcare organisations using the balanced scorecard approach: a case study from a Saudi hospital. *Procedia Comput Sci*. 2016;98:332-339.
16. Kalender ZT, Vayvay Ö. The fifth pillar of the balanced scorecard: sustainability. *Procedia-Soc Behav Sci*. 2016;235:76-83.
17. Xia YH. Reconstructing the relationship of medical production and bringing about new business opportunities—integrated care organization business value report. <https://vcbeat.net/ODFiMTMyNWJlYmZlMDNjZTJhMWNjNGQ3YmQ4MmNkNWI=>. Accessed June 9, 2018.
18. Li WD. Annual report on social development of Beijing 2015-2016. <http://www.bass.gov.cn/info/cn/lpsxl/bjshfzbg/View.aspx?nid=10123>. Accessed May 3, 2016.
19. Hsiao B, Chen L-H, Wu H-T. Assessing performance of Taiwan hospitals using data envelopment analysis: in view of ownership. *Int J Health Plann Mgmt* 2018; 1–15. <https://doi.org/10.1002/hpm.2676>
20. Lloyd GF, Bajorek B, Barclay P, Goh S. Narrative review: status of key performance indicators in contemporary hospital pharmacy practice. *J Pharm Pract Res*. 2015;45(4):396-403. <https://doi.org/10.1002/jppr.1124>.
21. Chang CS, Bonhoure P, Alam S, Hansen P, Burnham G. Use of the balanced scorecard to assess provincial hospital performance in Afghanistan. *World Med Health Policy*. 2010;2(1):83-106. <https://doi.org/10.2202/1948-4682.1045>.
22. Castañeda-Méndez K, Mangan K, Lavery AM. The role and application of the balanced scorecard in healthcare quality management. *J Healthc Qual*. 1998;20(1):10-13. <https://doi.org/10.1111/j.1945-1474.1998.tb00243.x>.

23. Cefalu MS, Elliott MN, Setodji CM, Cleary PD, Hays RD. Hospital quality indicators are not unidimensional: a reanalysis of Lieberthal and Comer. *Health Serv Res* 2018; 1–7. <https://doi.org/10.1111/1475-6773.13056>
24. Barra AO, Silva SAPS, Vitorino MR. Public sports policy: a formative experience for sports administrators based on the logical framework approach. *J Phys Ed Sport*. 2016;16(3):841-849.
25. Lin T. The 14 editorial boards of the journal of editors rank “the list of the most influential scholars in Chinese philosophy and social sciences.”. *Acta Editologica*. 2017;499(5):499.
26. Boutsidis C, Zouzias A, Mahoney MW, Drineas P. Randomized dimensionality reduction for K-means clustering. *IEEE Trans Inf Theory*. 2015;61(2):1045-1062.

How to cite this article: Wu X, Li S, Xu N, Wu D, Zhang X. Establishing a balanced scorecard measurement system for integrated care organizations in China. *Int J Health Plann Mgmt*. 2019;34:672–692. <https://doi.org/10.1002/hpm.2726>