The Most Expensive is Not the Best

Hongxing Hao
Xiangrong Zeng
Boliang Sun

National University of Defense Technology
Changsha, China

Advisor: Ziyang Mao

Abstract

Motivated to evaluate healthcare systems more accurately, we analyze existing evaluation methods. Most methods mainly focus on outcomes and their metrics often ignore internal characteristics of the healthcare systems.

We devise two methods: an improved World Health Organization (WHO) method and a comprehensive evaluation method.

The improved WHO method uses the same metrics as the WHO method, which are determined by the outcomes of the healthcare system. Our improvement is to use a grey comprehensive evaluation and the principle of minimum loss of information to combine the metrics, rather than simply combining them linearly.

In our comprehensive evaluation method, we define five new metrics that concern both outcomes and characteristics of the healthcare system itself, including the effect of the government and the basic situation of a country. Then we use the equal-interval method to get a final score. Compared with other methods, this one does a better job in distinguishing countries and in sensitivity.

After comparing with other four countries that represent the four main modes of healthcare systems, we conclude that the most important reason why the highest cost can't make the U.S. the best is unfairness.

Afterward, we use a neural network algorithm to predict what will happen to the U.S. if some values of the metrics change. We conclude that the U.S. can get the greatest benefit by improving fairness.

We finally consider a policy change, a "medical insurance voucher," as a method to increase insurance coverage and reduce unfairness.

The UMAP Journal 29 (2) (2008) 155–168.

Introduction

Many countries have recently introduced reforms in the health sector with the explicit aim of improving performance [Mathers et al. 2000; 2001]. There is extensive literature on health-sector reform, and recent debates have emerged on how best to measure performance so that the impact of reforms can be assessed [Goldstein 1996]. Measurement of performance requires an explicit framework defining the goals of a healthcare system and a suitable method to make a compelling evaluation.

So our goal is pretty clear:

- Devise metrics to evaluate the effectiveness of a healthcare system.
- Devise a method to combine the metrics.
- Compare several representative countries.
- Restructure the healthcare system of the U.S. and build predictive models to test the changes.

Our approach is:

- Analyze factors that can affect the performance of a healthcare system.
- Search the literature on existing evaluation methods and find their shortcomings.
- Develop a comprehensive evaluation method that asks only for existing data or data easy to measure and collect.
- Collect experimental data that can be used in our method.
- Compare current methods and determine their characteristics.
- Do a sensitivity analysis of variations of our models.
- Compare healthcare systems of several representative countries.
- Restructure the healthcare system of the U.S. and build a model based on neural networks to test changes.
- Do further discussion based on our work.

Four Representative Healthcare Systems

The healthcare system, as an important part of the social security system, is essential to promote the stability of society, and it reflects social justice. Due to the different histories, cultures, and status of human rights protection, healthcare systems vary from country to country.

There are four representative healthcare systems [Ding 2005]:

- **National insurance.** The main countries using this system are the UK, Eastern Europe, and Russia. The government dominates, healthcare is free, with full medical treatment and complete coverage of the population. But it doesn't have high efficiency or make use of the market, and it is a heavy burden to the government.
- **Commercial insurance.** The U.S. is the main country using this system, which makes the market the guideline of the healthcare system. Cost is high, and a large number of people fail to pay.
- **Social insurance.** This system features mandatory coverage and fairness, as in Japan, Germany, and Canada. It has high cost and slow service.
- **Savings insurance.** Singapore is the representative country. The main disadvantage is a low service efficiency. Costs rise rapidly, and it cannot achieve full coverage.

Analysis of the WHO Estimation Method

The WHO's methods focus on the outcomes of a healthcare system without considering any characteristics of the system itself.

Strengths

The metrics that the WHO uses to evaluate a healthcare system aim to measure goal attainment, and they include most of the outcomes that a healthcare system should produce.

Weaknesses

- The weights placed on each dimension are somewhat arbitrary.
- The approach heavily penalizes countries with epidemic disease unrelated to the healthcare system.
- This approach does not look at how the system is organized and managed.
- The WHO 2000 rankings do not look at access, utilization, quality, or cost-effectiveness.

In addition, according to Almeida et al. [2001, 1693]:

- "The measure of health inequalities does not reflect concerns about equity."
- "Important methodological limitations and controversies are not acknowledged."

- "The multicomponent indices are problematic conceptually and methodologically; they are not useful to guide policy, in part because of the opacity of their component measures."
- "Primary health care is declared a failure without examining adequate evidence, apparently based on the authors' ideological position."
- "These methodological issues are not only matters of technical and scientific concern, but are profoundly political and likely to have major social consequences."

Improved WHO Method

In the WHO methods, the weights in the construction of the composite index are used without considering uncertainty in the values of the different components.

We use a *grey comprehensive evaluation model*[1] to improve the WHO method to make the evaluation more credible.

Methodology

Suppose that c_{ik}, for are the raw data of the metrics $k = 1, \ldots, m$ in country i, for $i = 1, \ldots, n$, giving the $n \times m$ matrix $C = (c_{ik})$. We suppose that c^* is the best value in metric k among all countries. We take $C^* = (c_k^*) = (c_1^*, \ldots, c_m^*)$, a best possible situation, as a reference and compare the value of metric k in country i to this ideal via

$$\xi_i(k) = \frac{\min_i |c_k^* - c_{ik}| + \rho \max_i |c_k^* - c_{ik}|}{|c *_k - c_{ik}| + \rho \max_i |c_k^* - c_{ik}|},$$

where $\rho \in (0, 1)$ is a differentiation coefficient that we generally can take to be 0.5. Using $\xi_i(k)$, we get the evaluation matrix $E = \big(\xi_i(k)\big)_{n \times m}$.

Suppose $W = (w_1, \ldots, w_m)$ is a weight-distribution vector for the m metrics, with w_k the weight of metric k and $\sum w_k = 1$. Based on the discussion above, we get the *grey comprehensive evaluation model*

$$R = W \cdot E^T = (r_1, \ldots, r_n),$$

[1]EDITOR'S NOTE: This method, not known under this name in the U.S., was introduced by Deng Julong in *Tutorial of Grey System Theory* [in Chinese] (1982). It uses ideas of T.L. Saaty's analytic hierarchy process and is well well-known in China (googling `"grey system"` gets 64,100 hits, including *The Journal of Grey System*, edited by Deng). For a numerical example, see Sun, Yan and Zong Sun, The grey comprehensive evaluation model for safety of construction sites, *2007 International Conference on Wireless Communications, Networking, and Mobile Computing*, 5240–5243.

where E^T is the transpose of E and $r_i = \sum_{k=1}^{m} w_k \xi_i(k)$ is the relating degree. The vector $R = (r_1, \ldots, r_n)$ contains the final scores of the countries' healthcare systems. The larger r_i, the better the system.

How to Determine the Weights

We we want to determine the weight vector in a credible way. We use the principle of minimum loss [Wang et al. 2000]. Because our metrics u_j evaluate information from different aspects, combining all the metrics in a linear way would lose a lot of information, according to entropy theory in informatics.

We should maximize conservation of information. So we choose the most classical method: We calculate variance to represent information; the larger the variance, the more information.

In the final score $d = w^T u$, we should choose the best weight w to make the variance of d reach the maximum:

$$D(d) = w^T D(u) w,$$

where $D(d)$ is the variance matrix of d. When $w^T w = 1$, $D(d)$ achieves its maximum.

We use the method of Lagrange multipliers. Suppose that

$$\varphi(w, \lambda) = w^T D(u) w - \lambda(w^T w - 1).$$

Then

$$\begin{aligned} \frac{\partial \varphi}{\partial w} &= 2D(u)w - 2\lambda w = 0, \\ \frac{\partial \varphi}{\partial \lambda} &= w^T w - 1 \qquad\quad = 0, \end{aligned}$$

which reduces to

$$\begin{aligned} D(u)w &= \lambda w, \\ w^T w &= 1. \end{aligned}$$

So λ is an eigenvalue of $D(u)$ with eigenvector w. When $w^T w = 1$, to make $D(d) = w^T D(u) w = \lambda w^T w = \lambda$ reach the maximum, we should take λ as the maximum eigenvalue of $D(u)$.

In the real calculation, we do not know $D(u)$, so we use the variance matrix $\hat{D}(u) = (\hat{\sigma}_{lj})$ of the sample $(c_{1j}, \ldots, c_{nj})$ of u_j to represent it, where

$$\hat{\sigma}_{lj} = \frac{1}{n} \sum_{k=1}^{n} (x_{kl} - \bar{x}_l)(x_{kj} - \bar{x}_j), \qquad \bar{x}_j = \frac{1}{n} \sum_{k=1}^{n} x_{kj}.$$

The variance matrix $\hat{D}(u)$ is a nonnegative symmetric real matrix, so all its eigenvalues are real. From the properties of Rayleigh's entropy, we get

$$\lambda_0 = \max_{w \neq 0} \frac{w^T \hat{D}(d) w}{w^T w} = \max_{||w||=1} \frac{w^T \hat{D}(d) w}{w^T w},$$

where λ_0 is the maximum eigenvalue of $\hat{D}(u)$, and the eigenvector w of $\hat{D}(u)$ is the weight vector that we seek.

A Partial Discussion

The improved WHO method does not change the focus on outcomes of the healthcare system. Its improvement is in making the evaluation more credible. This kind of method makes its own sense in that it really can reflect the goals of the healthcare system, but it can't reflect the inside. For example, a country with an epidemic often gets a low score in WHO's evaluation method, but maybe this is not the problem of the healthcare system. So a new method that reflects the inside is needed.

Comprehensive Evaluation Method

We bring up a method to evaluate a healthcare system, mentioned by [Ding 2005], that considers both the outcomes and properties of systems themselves.

Metrics to Evaluate Overall Effectiveness

To make an overall comparison between countries' health care systems more objectively, fairly and quantitatively, the metrics must be made well. The World Bank has specified the goals of a healthcare system [Schieber and Maeda 1997, 2]:

- "Improving a population's health status and promoting social well-being"
- "Ensuring equity and access to care"
- "Ensuring microeconomic and macroeconomic efficiency in the use of resources"
- "Enhancing clinical effectiveness"
- "Improving the quality of care and consumer satisfaction"
- "Assuring the system's long-run financial sustainability"

Pursuant to this definition, we make five metrics for the overall healthcare system:

- **Efficiency,** the proportionality between inputs and outcomes. It can be divided into technical efficiency, economic efficiency, and allocative efficiency. For our purposes, we choose technical efficiency.
- **Fairness,** both in medical treatment and in contributing to the costs.
- **Responsiveness** "refers to the non-health improving dimensions of the interactions of the populace with the health system, and reflects respect of persons and client orientation in the delivery of health services, among other factors" [Tandon et al. 2000, 2–3].
- **The effect of the government.**
- **The basic situation of a country.** This means a composite index of sectors, which include economy, education, scientific research, and population.

The Model to Deal with the Index and Data

Accordingly, we make five new indexes, one for each metric above.

Choose the Operation Model

We use the method of equal intervals to combine the indexes, which is also used in the Human Development Index by the United Nations to compare countries. We also solve the problem of how to determine the weights.

The Equal Interval Method

The Operating Process

- Divide the subindexes into positive indexes and negative indexes.
- Use different algorithms to make the standardization to the two kinds of indexes.
- According to the subindexes, we can get the five main indexes' composite values.
- Calculate the final score of different countries based on the five metrics' values.

Classification of the Indexes

- **Classification.** Positive index: the higher the value, the better the healthcare system; for example, availability of safe drinking water. Negative index: the higher the value, the worse the healthcare system; for example, the proportion of smokers.

- **Standardization.** The indexes have different units, so we should standardize before calculating the final score. After the classification, we can deal with the two kinds of indexes differently.
  - Positive index: $F_{ilj} = \dfrac{R_{ilj} - R_{il\,\min}}{R_{il\,\max} - R_{il\,\min}} \times 100,$
  - Negative index: $F_{ilj} = \dfrac{R_{il\,\max} - R_{ilj}}{R_{il\,\max} - R_{il\,\min}} \times 100,$

  where
  - i is the one of five metrics,
  - l is the subindex of the metric i,
  - j is the one of the countries,
  - $R_{il\,\min}$ is the minimum value of the l subindex of the metric i in the statistical data, and
  - $R_{il\,\max}$ is the maximum value of the l subindex of the metric i in the statistical data, and F_{ilj} is the value of the l subindex of the metric i after standardization.

Determine the weights. We can get the value of every metric using the function

$$F_{ij} = \left(\sum_{i=1}^{n} \frac{(F_{ilj})^{\alpha}}{n} \right)^{1/\alpha},$$

where n is the number of the subindex in metric i and α is a weight of the metric i.

Get the final score of the evaluated country. Based on the discussion above, we get the function

$$S = \left(\sum_{i=1}^{n} \frac{(F_{ilj})^{\alpha}}{k} \right)^{1/\alpha},$$

where k is the number of metrics (in our case, $k = 5$).

Comparisons between Methods

Before the comparison, each component measure was rescaled on a 0 to 100 scale:

- for healthy life expectancy, $H = \dfrac{\text{Health} - 20}{80 - 20} \times 100;$
- for health inequality, $\text{HI} = (1 - \text{HealthInequality}) \times 100;$

- for responsiveness level, $R = \dfrac{\text{Responsiveness}}{10} \times 100$;
- for responsiveness inequality, $\text{RI} = 100(1 - \text{ResponsivenessInequality})$;
- for fairness in financing, $\text{FF} = \text{FairnessofFinancingContribution} \times 100$.

The overall composite was, therefore, a number from 0 to 100.

Dipartite Degree Analysis

As we know, a good metric should distinguish. But the WHO's method can't; for example, its method gives 36 countries the same value in its metric of responsiveness. We evaluate the degree of distinction via

$$\text{DD} = \sqrt{n_1^2 + \cdots + n_i^2},$$

where $N - i$ is the number of countries that can't be distinguished in criterion i. The smaller DD, the better the degree of distinction.

Monte Carlo Simulation

To test the dipartite degree (degree of distinction) of every method, we use Monte Carlo simulation to make a small change to every data value, since the value must contain some error. The process is as below.

First, we use the beta distribution to determine the change in each value. Because the beta distribution is restricted to the interval $[0, 1]$, a linear function of a beta-distributed random variable can be used to scale the sampling interval appropriately.

The beta distribution can be described by the probability density function

$$\text{Beta}(\alpha, \beta)(x) = \begin{cases} \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\, x^{\alpha-1}(1-x)^{\beta-1}, & 0 < x < 1; \\ 0, & \text{else.} \end{cases}$$

It has expected value $E[X] = \alpha/(\alpha+\beta)$.

Suppose that x_{ij}, with $1 \leq i \leq 191$ and $1 \leq j \leq 10$, is the unknown true mean of the random variable X_{ij} representing the jth metric in country i. We let

$$X_{ij} = (x_{ij} - 1) + 2\text{Beta}(2, 2)(X),$$

which takes values in $[x_{ij} - 1, x_{ij} + 1]$ and has expected value

$$E[(x_{ij} - 1) + 2\text{Beta}(2, 2)(X)] = x_{ij} - 1 + 2[2/(2+2)] = x_{ij}.$$

We use Monte Carlo simulation to create 1,000 numbers randomly in the interval $[x_{ij} - 1, x_{ij} + 1]$ and calculate a 95% confidence interval for x_{ij}.

Sensitivity Analysis

About the Values of the Metrics

In this part, we change the values but keep the weights to see how can this change affect the evaluation result. Then we can arrive at the most important metric, the one that can affect the final score acutely.

Suppose that G_p and G_q are the final scores of countries p and q. Let U_{qr} be the the value of metric r in country q. Change it to make $G_p = G_q$; then we can get the marginal value U_{qr}^B:

$$U_{qr}^B = U_{qr} + \frac{G_p - G_q}{w_r}.$$

We can do sensitivity analysis to the values of the metrics following the process below:

- If U_{qr}^B is outside of the allowable interval, whatever it changes, it won't change the order of the two countries; so r is a value-insensitive metric.
- When U_{qr} is close to U_{qr}^B, changing the value will change the order of the two countries; so r is a value-sensitive metric.

About the Weights

In this part, we change the weights but keep the values of the metrics to see how doing so affects the evaluation result. Then we can get the most important weight, the one that can affect the final score acutely.

When a weight changes, it affects others, since the weights sum to 1. To make a simple analysis, when a weight changes, let only one another change at the same time, and keep the others fixed.

Suppose that the weights' values before they change are $\bar{w}_j$, $\bar{U}_{ij}$, $\bar{G}_j$ and after changing they are w_j, U_{ij}, G_j. Suppose that the changing weights are r and s, so that

$$w_r + w_s = \bar{w}_r + \bar{w}_s.$$

The changing interval of w_r and w_s is $[0, \bar{w}_r + \bar{w}_s]$. When they change, maybe the final score of one country will equal that of another. Let the two countries be p and q. Then we can get the marginal weights

$$w_r^B = \frac{\bar{B}_p - \bar{G}_q}{\left(\bar{U}_{pr} - \bar{U}_{qr}\right) - \left(\bar{U}_{ps} - \bar{U}_{qs}\right)},$$
$$w_s^B = (\bar{w}_r + \bar{w}_s) - w_r^B.$$

When the two countries have the same score, we can get r and w_s as

$$w_r = \bar{w}_r - w_r^B, \qquad w_s = \bar{w}_s - w_s^B.$$

We can do the sensitivity analysis to the weights following the process below. Because the changing interval of w_r and w_s is $[0, \bar{w}_r + \bar{w}_s]$, if w_r and w_s are outside the interval, the change won't affect the final order of the two countries; the metrics r and s are insensitive. If not, this change may affect the final order of the two countries.

- If $w_r > \bar{w}_r$, so that the weight of metric r is bigger than w_r, the final order of the two countries will be changed; then r is a weight-insensitive metric for the country with the lower score.
- If $w_r < \bar{w}_r$, when the weight of metric r is smaller than w_r, the final order of the two countries will be changed too; then s is a weight-insensitive metric for the country with the lower score.

Analysis of American Healthcare System Based on Neural Networks

The Design of the Back Propagation Network

Because of the difficulty of data collection, we choose just satisfaction and seven other indexes as the inputs to a back propagation (BP) neural network:

- health expenditure per capita,
- number of doctors per thousand people,
- number of sickbeds per thousand people,
- anticipated lifespan,
- infant mortality,
- proportion of the healthcare cost in GDP, and
- extent of healthcare coverage.

So, the network should have 7 nerve cells in input layer, and 15 (= 2×7+1) nerve cells in middle layer.

We choose satisfaction to be the target of the network, so there is just one nerve cell in the output layer. According the principles for designing a BP network, the passing function to the middle layer is a sigmoid function. We created the neural network in Matlab.

Application of the BP Network

To check the effect on satisfaction when an index changes, we make one of them rise by 20% once and keep the others unchanged. Doing that, we can get the satisfaction for each year as shown in **Figure 1**.

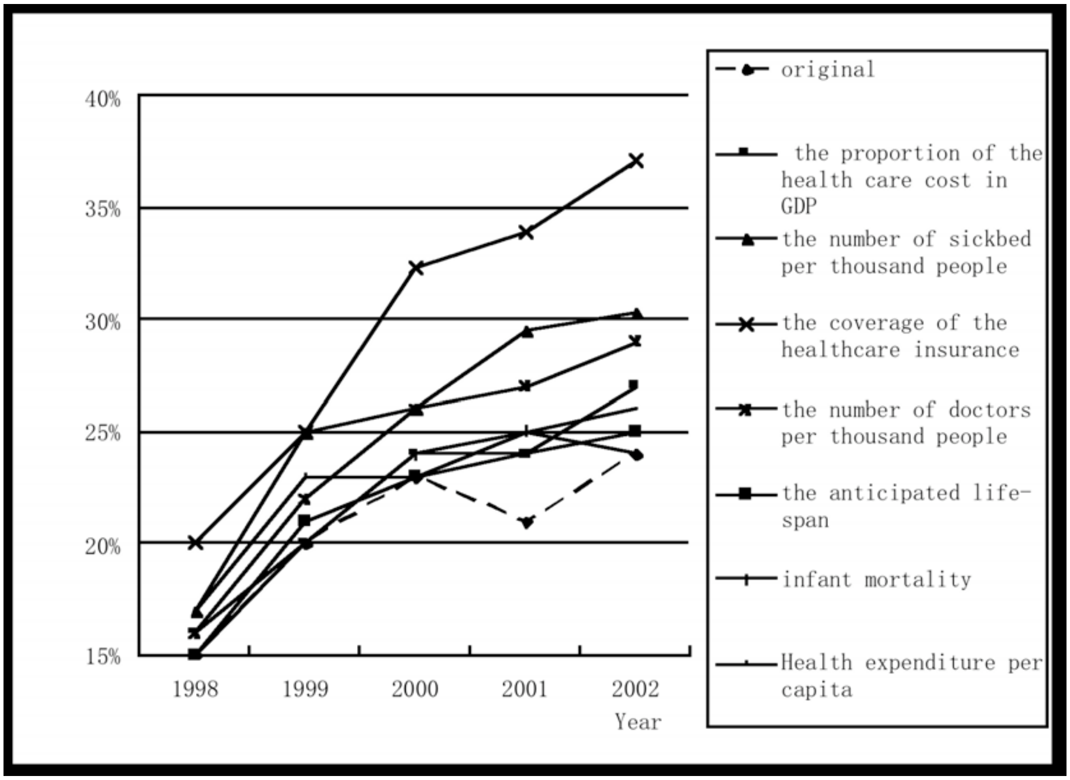

Figure 1. The satisfaction curve after adjustment.

The satisfaction has a rising trend when an index rises. The coverage of healthcare insurance improves the result to the greatest degree. So increasing the coverage of healthcare insurance is a good way to improve the performance of the U.S. healthcare system.

Advice to the Healthcare System of the U.S.

According to the analysis above, the most important problem in the healthcare system of the U.S. is coverage. Though the government has established insurance for the elderly and for children, a lot of people still fail to buy insurance because it is expensive. Universal healthcare coverage will not only lead to fairness in healthcare but also encourage insurers to give better service.

Based on this, we bring up a plan of a "medical insurance voucher" to make the U.S. reach universal healthcare coverage rapidly. We suggest that the government run an insurance institution itself, while at the same time encouraging commercial healthcare insurance institutions. The government should put out the same "medical insurance voucher" to all residents, who can choose a healthcare insurance institution in which to participate.

To fund this program, we would tax smoking and alcohol consumption.

The differences between public and private insurance are in service and cost. The government should provide basic medical care—the lowest level of service. Commercial insurance should offer more service and better conditions, at a slightly higher cost. A resident who participates in commercial insurance should thus pay a little more in addition to using the medical insurance voucher. When healthcare coverage becomes universal, people will pay only a small part of their income to get the healthcare. Advantages of this plan are:

- The plan designs a competitive relationship among insurance institutions, to make them to do their best to reduce cost and improve quality of healthcare—thus improving the effectiveness of the healthcare system.
- In particular, there is a competitive relationship between government (social) insurance and commercial insurance. In some countries where social insurance dominates, needs can't be satisfied and effectiveness is low. Besides, setting social insurance at a minimal level can not only make commercial insurance institutions improve themselves, but adjust the national view.
- Collect the funding for the medical insurance vouchers by taxes, which solves the problem of fairness. Fairness asks the healthcare system not to provide the medical care by income but by need. The tax system has a target of reallocating incomes, and it also can be used to solve the problem of fairness.
- This plan protects the right of choice of residents. It combines competition and human rights, making for a balance between two important problems.

References

Ding, Chun. 2005. An empirical comparison of four models of healthcare systems in the world. *World Economic Papers* 2005 (4) #36. `http://scholar.ilib.cn/A-sjjjwh200504036.html`.

Goldstein, H., and D.J. Spiegelhalter. 1996. League tables and their limitations: Statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society Series A* 59 (3): 385–443. `http://www.cmm.bristol.ac.uk/team/HG_Personal/limitations-of-league-tables.pdf`.

Mathers, Colin D., Ritu Sadana, Joshua A. Salomon, Christopher J.L. Murray, and Alan D. Lopez. 2001. Healthy life expectancy in 191 countries, 1999. *Lancet* 357: 1685–1691. `http://jasalomon.googlepages.com/Mathersetal_HealthyLifeExpectancy_Lancet2001.pdf`.

________. 2000. Estimates of DALE for 191 countries: methods and results. GPE Discussion Paper No. 16. Geneva, Switzerland: World Health Organization. https://www.who.int/entity/healthinfo/paper16.pdf .

Schieber, George, and Akiko Maeda. 1997. A curmudgeon's guide to financing health care in developing countries. In *Innovations in Health Care Financing* , edited by George J. Schieber, 1–40. World Bank Discussion Paper No. 365. Washington, DC: World Bank.

Tandon, Ajay, Christopher J.L. Murray, Jeremy A. Lauer, and David B. Evans. 2000. Measuring overall health system performance for 191 countries. GPE Discussion Paper No. 30. Geneva, Switzerland: World Health Organization. https://www.who.int/entity/healthinfo/paper30.pdf .

Zheng, Xiaowei, Gong Zhao Hui, and Wang Xue. 2000. Linear synthetical evaluation function and the determination of its weighting coefficients [in Chinese]. *System Engineering—Theory and Practice* 2000 (10). http://scholar.ilib.cn/A-xtgcllysj200010016.html .

From left to right: Advisor Ziyang Mao and team members Hongxing Hao, Boliang Sun, and Xiangrong Zeng.