

Measuring and Benchmarking Technical Efficiency of Public Hospitals in Tianjin, China: A Bootstrap–Data Envelopment Analysis Approach

Hao Li, PhD¹, and Siping Dong, PhD²

Abstract

China has long been stuck in applying traditional data envelopment analysis (DEA) models to measure technical efficiency of public hospitals without bias correction of efficiency scores. In this article, we have introduced the Bootstrap-DEA approach from the international literature to analyze the technical efficiency of public hospitals in Tianjin (China) and tried to improve the application of this method for benchmarking and inter-organizational learning. It is found that the bias corrected efficiency scores of Bootstrap-DEA differ significantly from those of the traditional Banker, Charnes, and Cooper (BCC) model, which means that Chinese researchers need to update their DEA models for more scientific calculation of hospital efficiency scores. Our research has helped shorten the gap between China and the international world in relative efficiency measurement and improvement of hospitals. It is suggested that Bootstrap-DEA be widely applied into afterward research to measure relative efficiency and productivity of Chinese hospitals so as to better serve for efficiency improvement and related decision making.

Keywords

technical efficiency, health services provision, Bootstrap-DEA, benchmarking, methodology

Introduction

Hospital efficiency and productivity research is one of the priority fields in hospital management and health economics. In our previous research, we have made an extensive literature review about China's efficiency measurement of hospitals and we have found that lots of problems exist in selecting appropriate indicators and data envelopment analysis (DEA) models.¹ Although some articles are relatively standardized compared with the international literature,¹⁻⁴ the authors have yet to address the bias problem of efficiency scores resulted from the application of traditional DEA models such as Charnes, Cooper, and Rhodes (CCR), Banker, Charnes, and Cooper (BCC), Malmquist-DEA, and so forth. It is therefore of special significance to introduce cutting edge international methods to measure the technical efficiency of China's public hospitals so as to not only shorten the research gap between China and the international world but also produce more reliable results for performance improvement activities and decision making.

There have been at least two main popular methods widely recognized and applied worldwide to measure hospital efficiency: parametric and non-parametric methods. The parametric method is represented by stochastic frontier analysis (SFA),

which needs to construct a function and is limited to single output,⁵ though in many cases its application is limited in hospital context, because hospitals have multiple outputs. In comparison, the non-parametric method, represented by DEA, is based on linear programming and can be applied for relative efficiency analysis of hospitals with multiple inputs and outputs. In China, the traditional DEA models such as CCR, BCC, Malmquist-DEA, their derivatives, and so forth have been widely applied for decades to measure hospital efficiency. According to these models, only decision-making units (DMUs) operating in the frontier would be considered as efficient. However, in real context, all the DMUs are subject to environment and random factors, which means that their efficiency scores shall fall into a fluctuating range. To

¹School of Public Health/Global Health Institute, Wuhan University, Wuhan, People's Republic of China.

²National Institute of Hospital Administration, Beijing, People's Republic of China.

Corresponding Author:

Siping Dong, National Institute of Hospital Administration, National Health and Family Planning Commission, Beijing 100191, People's Republic of China.

Email: sipingd@163.com



address this issue, Simar and Wilson, Daraio and Simar, and others have introduced the Bootstrap method for efficiency measurement of DMUs based on DEA to correct the bias of efficiency scores and to calculate their confidential intervals, lower and upper bounds, and so on.⁵⁻⁷ These studies have helped improve the accuracy of DEA efficiency scores. Gradually, the Bootstrap-DEA has been recognized as a milestone in international world in relative efficiency and productivity measurement. To our knowledge, it is yet to be introduced to China to measure the efficiency and productivity of hospitals, and the necessity of doing so has already been recognized in our previous research.^{1,2} Furthermore, most of the Chinese studies have long been stuck in the measurement of hospital efficiency in itself and little work has been done on how to take advantage of the results to conduct benchmarking for inter-organizational learning and continuous efficiency improvement.

Therefore, the purpose of this study is to take the first initiative to introduce the Bootstrap-DEA approach to measure the technical efficiency of public hospitals in the Chinese context and explore the benchmarking mechanism for further efficiency improvement and learning. It can also serve as a preliminary study in the improvement of methodology for the reference of other Chinese researchers to apply in the near future.

Method

Sample Selection and Data Source

In this research, all the 14 third-grade public general hospitals in Tianjin, China, were selected for our study purpose. In China, public hospitals have been accredited into 3 grades and the third-grade hospitals are the highest level ones.² Hospitals of each grade differ in capacity, functions, and so forth. Besides, hospitals of different kinds cannot be compared for their operating efficiency. Therefore, in this study we just focus on the efficiency measurement of the third-grade hospitals. The data were collected from National Institute of Hospital Administration (NIHA), National Health and Family Planning Commission of the PRC (NHFPC). NIHA is a research institution directly affiliated to NHFPC, engaged in hospital management, health economics, and policy research.

The Selection of Input and Output Indicators

In the published Chinese literature about the technical efficiency measurement of public hospitals, hospital expenditure, number of beds, number of staff, fixed asset, and so forth have been selected as typical input indicators; number of diagnostic visits, number of discharged inpatients, bed occupancy rate, hospitalization days, hospital revenue, and so forth have been selected as typical output indicators.^{1,2} Such selection method has mixed technical efficiency with allocative efficiency and causes double counting problem.² To avoid this, in this study, the indicator selection is based on our previous research,² in which the actual number of open

beds and the number of staff are selected as the input indicators, whereas the number of diagnostic visits and the number of discharged inpatients are selected as the output indicators. Such selection of indicators is similar to the works of Ng and Yang and Zeng.^{3,4} The difference is that in our study, to reduce the dimensionality,⁸ the number of physicians, the number of nurses, and the number of other staff are merged to the number of hospital staff, because altogether, there are only 14 third-grade general public hospitals in Tianjin.

The Bootstrap-DEA Approach

Bootstrap is a data-based simulation method for statistical inference,⁵ which was first proposed by Bradley Efron in 1979.⁹ The basic idea of bootstrap is to simulate the data-generating process (DGP) with repeated sampling. As the simulated data set is approximately equivalent to the original one, the sampling distributions and standard deviations are therefore close to those of the original ones. Simar and Wilson first introduced the Bootstrap-DEA approach, in which the estimated efficiency scores can be drawn by conducting numerous repeated sampling,^{6,7} thus producing bias corrected efficiency scores and confidential intervals at the α level, making the efficiency scores more accurate. The fundamental calculation theory can be given as follows:

$$Bias(\hat{\theta}_k) = E(\hat{\theta}_k) - \hat{\theta}_k.$$

$$Bias(\hat{\theta}_k) = B^{-1} \sum_{b=1}^B (\hat{\theta}_{kb}^*) - \hat{\theta}_k.$$

The bias corrected efficiency score can be expressed as follows:

$$\tilde{\theta}_k = \hat{\theta}_k - Bias(\hat{\theta}_k) = 2\hat{\theta}_k - B^{-1} \sum_{b=1}^B (\hat{\theta}_{kb}^*).$$

The confidential interval at the α confidence level can be calculated as follows:

$$P_r \left(-\hat{b}_r \leq \hat{\theta}_{kb}^* - \hat{\theta}_k \leq -\hat{a}_\alpha \right) = 1 - \alpha.$$

$$P_r \left(-\hat{b}_\alpha \leq \hat{\theta}_k^* - \hat{\theta}_k \leq -\hat{a}_\alpha \right) = 1 - \alpha.$$

$$\hat{\theta}_k + \hat{a}_\alpha \leq \hat{\theta}_k \leq \hat{\theta}_k + \hat{b}_\alpha.$$

The Benchmarking Method

In our study, the idea of efficiency benchmarking comes from the regional performance evaluation experience in Tuscany (Italy), in which benchmarking is based on a full set

Table 1. Descriptive Characteristics of Sample Indicators.

Indicators	Mean	SD	Minimum	Maximum
Number of employees	1394	676	564	2947
Actual number of open beds	818	517	336	2200
Total number of outpatient and emergency visits	972487	578666	222056	2308032
Number of discharged patients	28070	16458	3277	55249

Table 2. Comparison of Efficiency Scores in 2012 With and Without Bias Correction.

Hospitals	Efficiency scores (bias not corrected)	Efficiency scores (bias corrected)	Bias	Bootstrap SD	Lower bound	Upper bound
H1	1.0000	0.8051	0.1949	0.0382	0.7139	0.9957
H2	0.9421	0.8646	0.0775	0.0031	0.7952	0.9387
H3	1.0000	0.8977	0.1023	0.0038	0.8438	0.9946
H4	0.8606	0.8026	0.0580	0.0014	0.7500	0.8565
H5	1.0000	0.7771	0.2229	0.0919	0.6400	0.9954
H6	1.0000	0.9032	0.0968	0.0035	0.8413	0.9938
H7	0.6688	0.6050	0.0638	0.0023	0.5510	0.6650
H8	0.6838	0.6290	0.0548	0.0016	0.5816	0.6807
H9	0.9828	0.9112	0.0716	0.0028	0.8333	0.9777
H10	0.4323	0.3970	0.0352	0.0011	0.3502	0.4304
H11	1.0000	0.8172	0.1828	0.0364	0.7163	0.9951
H12	1.0000	0.8060	0.1940	0.0409	0.7071	0.9949
H13	1.0000	0.9183	0.0817	0.0055	0.8171	0.9944
H14	1.0000	0.9389	0.0611	0.0031	0.8501	0.9937

of selected indicators for the performance management of health care institutions. In the Tuscan performance evaluation system, the performance scores are standardized and are further presented in a spider chart, which is composed of 5 bands.¹⁰ As in our study we focus only on efficiency benchmarking, it is unnecessary to use a spider chart. However, we have applied the same color definitions in bar chart to depict different standardized efficiency values. That means, if a standardized efficiency value is regarded as excellent performance, dark green will be applied in the benchmarking bar chart; if it is regarded as good performance, light green will be applied; if it is regarded as average performance, yellow will be applied; if poor performance is regarded, orange will be applied; and if failing performance is regarded, red will be applied. In practice, hospitals colored with yellow would have ample scope for improvement; hospitals colored with orange must improve their performance; hospitals colored with red must improve their performance urgently.

Data Processing and Analysis

R software and FEAR package¹¹ were applied to calculate the output oriented efficiency scores of the 14 hospitals. The efficiency scores before bias correction would return to Farrell scores¹² and the bias corrected ones after Bootstrap-DEA would return to scores based on Shephard's output distance functions, bias, variance, lower bound, and upper bound (2000 times of repeated sampling, with $\alpha = .05$).¹³ Then the bias

corrected efficiency scores are further used in a bar chart for benchmarking.

Results

Descriptive Statistics of the Sample

The descriptive characteristics of the sample in 2012 are depicted in Table 1.

Bootstrap-Data Envelopment Analysis Efficiency Scores

Table 2 is a comparison of the efficiency scores with and without bias corrections. It can be seen that, in the traditional BCC model, 8 hospitals have efficiency scores of 1, which means that they operate efficiently and do not need to improve their technical efficiency. Maintaining current operation would be their best choice. As the volumes of inputs and outputs in each of the hospitals are different and hospital operations are subject to environmental and random factors, it is obvious that there should be some efficiency bias to explain their difference, as is evidenced in the Bootstrap-DEA results in Table 2.

Efficiency Benchmarking

To further make the results visual for performance management, such as inter-organizational learning, other improvement

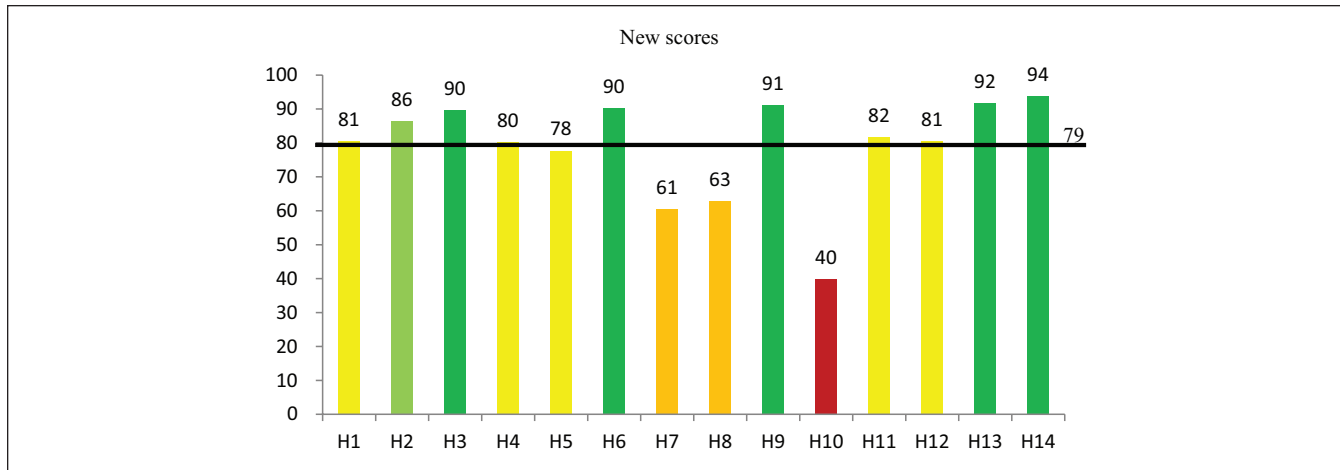


Figure 1. Efficiency benchmarking.

activities, and so forth, they are further benchmarked in Figure 1, in which the new scores are derived from efficiency scores multiplied by 100. The average score is 79. Altogether, 4 hospitals have scores lower than 79 and 10 hospitals have scores greater than 79. The hospitals can be categorized into 5 groups, where 5 hospitals fall into the first group (dark green), representing excellent performance; 1 hospital falls into the second group (light green), representing good performance; 5 hospitals fall into the third group (yellow), representing average performance, which have ample scope for improvement; 2 hospitals fall into the fourth group (orange), which must improve their performance; and 1 hospital falls into the fifth group (red), which must make urgent improvements.

Discussion

From Table 2, it is easy to find that all the bias corrected efficiency scores are lower than those before correction, indicating that the Bootstrap-DEA approach has improved the accuracy of the estimated efficiency scores. The model is therefore more precise than traditional DEA models. Furthermore, in the case of the small sample, the application of traditional DEA models is limited, because one of the basic conditions to apply DEA requires that the number of DMUs should be 3 times more than the total number of input and output indicators.¹⁴ However, the Bootstrap-DEA approach can help break the bottleneck by repeated sampling (normally 2000 times) to amplify the number of DMUs, so as to make the estimated efficiency scores much closer to their real scores. As in many Chinese studies, the number of DMUs did not meet the minimum requirement, and because few studies have applied Bootstrap-DEA approach,¹ in future study of Chinese hospitals, Bootstrap-DEA can be widely applied to provide more reliable results for efficiency improvement and decision making.

In Chinese studies, efficiency benchmarking has seldom been used for further improvements, though it can effectively

help hospital managers to identify best practices for other peer hospitals to learn from. According to Brown et al,¹⁵ the benchmarking of health care performance results can serve as a basis to further construct an inter-organizational mechanism for the health care institutions to learn and make continuous and sustainable improvements. Moreover, in some context, for example, when the environment factors do not change substantially, researchers can consider the DMUs in different periods as different DMUs in one period. In this way, the efficiency scores resulting from Bootstrap-DEA can be used for both horizontal and longitudinal benchmarkings. However, according to our previous study,¹⁶ China has yet to learn more from international experience to build its information systems, performance evaluation systems, performance rewarding systems, and so forth to enable benchmarking among hospitals with the support of a performance evaluation agency.

In China, the Bootstrap-DEA approach can be applied not only to general hospitals but also to health care institutions with homogeneous service provision such as township hospitals, community health services centers, and so forth. In the current literature, some studies have applied 2-stage DEA and 3-stage DEA to reduce the impact of environmental factors on efficiency scores.¹ In the 2-stage DEA, the efficiency scores are first estimated based on BCC and CCR models at the first stage. Then the inefficiency scores can be calculated and Tobit regression can be applied to calculate the statistical significance of their impact on inefficiency scores. This model can be further improved by applying Bootstrap-DEA at the first stage and then Tobit regression at the second stage to further improve the reliability of results. In contrast, the 3-stage DEA model is based on the work of Fried et al,¹⁷ where at the first stage, traditional BCC model can be applied to estimate the efficiency scores; then at the second stage, SFA approach can be applied to adjust the input volumes to generate same outputs; at the third stage, the traditional BCC model can be applied again to re-estimate the efficiency scores. The

Bootstrap-DEA can be applied at the first and third stages to help improve the reliability of efficiency scores.

This research has some limitations. First, to simplify the situation, all the environmental factors have been considered as random factors, though in future research, the factors can be further addressed as environmental and random factors, respectively, in 2- and 3-stage DEA models. Second, compared with most western countries, hospitals in China provide both outpatient and inpatient services instead of focusing on complicated disease treatment, plus the situation that few studies have selected the same input and output indicators, an international comparison is currently impossible. However, we will try in future research when we apply the same method and indicator selection on condition that the domestic and international hospitals have homogeneous nature. Third, the perspective we propose for benchmarking and inter-organizational learning here needs to be piloted in some hospitals so that the international experience can be tailored and become more suitable for the Chinese context.

Conclusion

In this study, we have first introduced the Bootstrap-DEA approach to measure and benchmark the technical efficiency of public hospitals in Tianjin for efficiency improvement. More researches need to be conducted in hospitals of different grades in whole China. In our literature searching scope, no similar research has been found in the hospital context in China, though this approach has been widely applied in international world and in other sectors of China. Therefore, this research has helped fill the long-existing gap in China's efficiency measurement of hospitals. Moreover, in our research we have proposed further benchmarking for further inter-organizational learning in Chinese hospitals. In further studies, the Bootstrap-DEA approach can be further embedded into 2-stage and 3-stage DEA models to improve the reliability of efficiency scores. Besides, the Bootstrap-DEA approach we have applied here can also be applied to estimate efficiency scores of other types of health care institutions in China such as community health services centers, township hospitals, and so forth.

Acknowledgments

The writing of this article was inspired by the comments of Professor Cinzia Daraio in Sapienza University of Rome to a preliminary study paper of the authors. It is right after the discussion with her that we try to fix the identified efficiency bias problem existing in China for decades. The authors are very grateful to Professor Daraio for her insightful advice and encouragement. Besides, they also thank Professors Sabina Nuti and Lino Cinquini for their support during Dr Hao Li's doctoral study in Scuola Superiore Sant'Anna, where he had the opportunity to get in-depth understanding of the Tuscan regional health care performance evaluation experience for benchmarking, inter-organizational learning, and continuous improvement, which were very helpful to generate some perspectives in this article.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The work was sponsored by the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry, People's Republic of China, and by the Fundamental Research Funds for the Central Universities.

References

1. Dong SP, Zuo YL, Tao HB, et al. Study on DEA-based Chinese hospital efficiency and applied indicators. *Chinese Health Policy Res.* 2014;7(12):40-45. (In Chinese)
2. Li H, Dong SP, Liu TF. Relative efficiency and productivity: a preliminary exploration of public hospitals in Beijing, China. *BMC Health Serv Res.* 2014;14:158. doi:10.1186/1472-6963-14-158.
3. Ng YC. The productive efficiency of Chinese hospitals. *China Econ Rev.* 2011;22(3):428-439.
4. Yang J, Zeng W. The trade-offs between efficiency and quality in the hospital production: some evidence from Shenzhen, China. *China Econ Rev.* 2014;31(4):166-184.
5. Daraio C, Simar L. *Advanced Robust and Nonparametric Methods in Efficiency Analysis: Methodology and Applications.* New York: Springer; 2007.
6. Simar L, Wilson PW. Sensitivity analysis of efficiency scores: how to bootstrap in nonparametric frontier models. *Manage Sci.* 1998;44(1):49-61.
7. Simar L, Wilson PW. A general methodology for bootstrapping in non-parametric frontier models. *J Appl Stat.* 2000;27(6):779-802.
8. Nuti S, Daraio C, Speroni C, Vainieri M. Relationships between technical efficiency and the quality and costs of health care in Italy. *Int J Qual Health C.* 2011;23:324-330.
9. Efron B. Bootstrap: another look at the Jackknife. *Ann Stat.* 1979;7(1):1-26.
10. Nuti S, Bonini A, Murante AM, Vainieri M. Performance assessment in the maternity pathway in Tuscany region. *Health Serv Manage Res.* 2009;22(3):115-121.
11. Wilson PW. FEAR 1.0: a software package for frontier efficiency analysis with R. *Socio Econ Plan Sci.* 2008;42(4):247-254.
12. Farrell MJ. The measurement of productive efficiency. *J Roy Stat Soc.* 1957;120(3):253-281.
13. Shephard RW. *The Theory of Cost and Production Functions.* Princeton: Princeton University Press; 1970.
14. O'Neill L, Rauner M, Heidenberger K, et al. A cross-national comparison and taxonomy of DEA-based hospital efficiency studies. *Socioecon Plann Sci.* 2008;42(3):158-189.
15. Brown P, Vainieri M, Bonini A, Nuti S, Calnan M. What might the English NHS learn about quality from Tuscany? Moving from financial and bureaucratic incentives towards "social" drivers. *Soc Public Policy Rev.* 2012;6(2):130-146.
16. Li H, Barsanti S, Bonini A. Building China's municipal health-care performance evaluation system: a Tuscan perspective. *Int J Qual Health C.* 2012;24(04):403-410.
17. Fried HO, Lovell CAK, Schmidt SS, Yaisawarng S. Accounting for environmental effects and statistical noise in data envelopment analysis. *J Prod Anal.* 2002;17(1-2):157-174.