PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Technical and scale efficiency of provincial health systems in China: a bootstrapping data envelopment analysis
AUTHORS	Chai, Peipei; Zhang, Yuhui; Zhou, Maigeng; Liu, Shiwei; Kinfu, Yohannes

VERSION 1 – REVIEW

REVIEWER	Abdosaleh Jafari Menzies Institute for Medical Research | University of Tasmania, Australia
REVIEW RETURNED	15-Nov-2018

GENERAL COMMENTS	1- In Table 1, abbreviation needs to be explained as a footnote. 2- In the discussion, please explain your reasons for each result clearly. 3- In the discussion, the results of more studies need to be mentioned. 4- Please do not use references before the year 2000.

REVIEWER	Siping Dong National Institute of Hospital Administration,PRC
REVIEW RETURNED	07-Jan-2019

GENERAL COMMENTS	1. In order to effectively measure relative efficiency in healthcare, it is crucial to select the appropriate input and output variables. But in this study, the authors selected the variables according to their subjective experiences rather than the scientific methods(e.g. A systematic review of international literatures or Delphi method). 2. The input variables of this study (health expenditure, medical personnel and hospital beds) mixed the volume and monetary variables, which resulted in the invalid technical efficiency measurement for the scores of DEA in this study not only reflected the technical efficiency, but also the allocative efficiency. In other words, the monetary variable(health expenditure) should be avoided to use in technical efficiency measurement. Therefore, the results and the conclusions are not reliable. 3. This study ignored a lot of literatures about Chinese Healthcare based on Bootstrap DEA approach for lacking a systematic review. 4. As aforementioned, the limitations of the study didn't discuss adequately. 5. Also, there are several mistakes(in my opinion) in the manuscript, such as "in China and internationally"（the first sentence of the Abstract).

REVIEWER	Li Wang McMaster University, Canada
REVIEW RETURNED	24-Mar-2019

GENERAL COMMENTS	Introduction: 1. The author need to describe the health care system in China, such as the components of public financing or private financing? Does the system apply for all provinces including municipalities or autonomous regions? Any medical program for the child, senior or the specialty group? 2. Line 24: need more detail to describe the overall uneven distribution of resource and health disparities (Not just an example). Do you expect that the inequality of health and resource use would be reduced through the study of the efficiency of health care system? If yes, why and how? 3. Health care system or health system? It's better to keep consistent over the content 4. Paragraph 2: literatures are restrictive with the Chinese studies? OECD countries and USA have developed quite a few studies on the efficiency of health care systems. What do you learn from them and what's the difference of your study? 5. Are there literatures on the system reviewers of the studies of health system efficiency in China? Methodology 1. Province in this study is the DMU. How many? Are the provinces, municipalities, and autonomous are comparable? 2. DEA: what's your study target to? OTE or SE, why? 3. Output orientation: why does your study focus on the output orientation only? Why not input orientation? The policy implication from this study should not be just in the short run? 4. CRS is not suitable for the health sector. 5. Bootstrapping DEA and truncated bootstrapping regression: the main reasons adopting the bootstrapping DEA are serially correlation of the efficiency scores and the random errors in the point inefficiency. 6. Did you use the double bootstrapping process developed in the Simar and Wilson (2008)? Or just the two separated bootstrapping in the efficiency estimation and the 2nd stage of the regression? Simar, L., and P.W. Wilson. 2007. "Estimation and inference in two-stage, semi-parametric models of production processes." Journal of Econometrics 136(1): 31-64. Data and Variables

	Input and Output variables: 1. more detail description – what medical personnel include? Physicians? Specialists? Nursing? Pharmacist? Other medical stuffs? If including all types of medical personnel, how do you adjust their various levels of skills? 2. Does the salary of the medial personnel include in the total health expenditure? If yes, it might double count the physician input. Also, it's necessary to describe what's the total health expenditure? What does it include? Hospitals, physicians, drug, dental, vision, long-term care facilities? 3. Does the capital investment include in the health expenditure? If yes, the hospital beds are double counted. 4. Line 46: what's the age cut-off for the premature mortality? 5. Line 51: What are the health worker-related factors and poor-supply factors? Any specific factors include in your 2nd stage of DEA analysis? 6. It's better to provide couple of references that convert the mortality rate to the survival rates. 7. How do you convert the YLD into health life years? 8. Have you considered to use the inverse of the IMR and MMR? 9. What's the OOP in line 11 of page 5? 10. Why do you select these factors in the 2nd stage of DEA.? More description, rational, and references are needed. Data Source Re-organize the order of the data sources. Output, input and factors. Results: Descriptive statistics 1. Number of provinces, municipalities and autonomous regions? 2. The good DEA model requires the input and output variables should be in scales manageable for DEA software and the variables have a similar variation. So Data usually will be rescaled to make them feasible for the DEA models. From the table 1, the scale and variation are quite large. The author may need to rescale the data. 3. What are the two provinces with 11.07 year gaps? 4. Figure 1: correlation tests for the inputs and outputs? I think it's not necessary to look at the association between the input and outcome. More interesting, may look at the efficiency change with the inputs? Efficiency estimates 1. Scale efficiency is not mentioned as the research objectives in the introduction section. Why you include this in the result section?

	2. Line 15 of page 7, what's the geographic location? 3. Table 3: all the descriptive statistics are efficiency scores. Why here regression on the inefficiency? 4. There are only 31 provinces but with 10 factors. The small sample size does create problems that I would think need to be addressed. 5. What's the R2 for the regression? Discussion 1. Line 54 of page 12. Why don't use outpatient-inpatient ratio but two separated factors at the 2nd stage of DEA? 2. Line 44 of page 13: wrong spelling in "system". 3. Need to address the limitation of small number of the DMUs in the regression analysis.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Abdosaleh Jafari
Institution and Country: Menzies Institute for Medical Research | University of Tasmania, Australia

Thank you for your time and expertise to review our paper, and for the observations and comments on the paper. We really appreciate it.

1. In Table 1, abbreviation needs to be explained as a footnote.
Response 1
We have revised and abbreviations have been explained as a footnote for table 1 on page 6, line number 23.
ISR (infant survival rates); MSR (maternal survival rates); HLY_NCDs (NCDs-based healthy life years); OOP (out-of-pocket).
2. In the discussion, please explain your reasons for each result clearly.
Response 2
We have now added the following sentence on page 12, line number 33:
Variables were selected based on evidence from similar previous studies and data availability for the intended analyses.
3. In the discussion, the results of more studies need to be mentioned.
Response 3
We have added the following references (Sun et al. on page 12, line number 41; Ramin et al., Achoko et al., and de Cos et al. on page 13, line number 22; and Hadad et al. and Achoko on page 13, line number 31).
4. Please do not use references before the year 2000.
Response 4
We have removed 3 references that were published before 2000 but retained others because they were either core methodological papers on DEA ad SFA or they were part of classical studies on the concepts of healthy life years and health system inputs.

Reviewer: 2
Reviewer Name: Siping Dong

Institution and Country: National Institute of Hospital Administration,PRC Please state any competing interests or state ‘None declared’: None declared

We appreciate your time and expertise to review our paper. Below are our thoughts on your observations.

1. In order to effectively measure relative efficiency in healthcare, it is crucial to select the appropriate input and output variables. But in this study, the authors selected the variables according to their subjective experiences rather than the scientific methods(e.g. A systematic review of international literatures or Delphi method).
Response 1
We agree that a comprehensive review of available literature is an important element of a research activity, and we believe we have done so in our paper. However, what we have not attempted is a “systematic review” or the other proposed approach because each of the proposed exercise can be formulated as a paper on their own, and that was not our objective. What we did, as with many other studies which we have referenced in our paper, is use evidence from past research and data availability to guide our study and select the variables appropriate for the intended analysis. To make this clearer, we have now inserted the following sentence in the data and methods section of the paper on the page 4 and line number 33.
“Variables for analysis were selected based on availability of data and evidence on relevance of variables in other similar studies.”
2. The input variables of this study (health expenditure, medical personnel and hospital beds) mixed the volume and monetary variables, which resulted in the invalid technical efficiency measurement for the scores of DEA in this study not only reflected the technical efficiency, but also the allocative efficiency. In other words, the monetary variable(health expenditure) should be avoided to use in technical efficiency measurement. Therefore, the results and the conclusions are not reliable.
Response 2
The inclusion of health expenditure in the model is well grounded both from a theoretical as well as methodological perspective. First, our approach, as with other underlying models for efficiency analysis in the DEA or SFA tradition, is based on the well-known Cobb-Douglass production function, which views an output as a function of labour and capital, meaning that capturing monetary values involved in the production process is an established tradition in such analysis. Second, all major and influential efficiency analyses work which we are aware of (for example, see Evans DB, Tandon A, Murray CJL, et al. Comparative efficiency of national health systems: cross national econometric analysis. BMJ 2001;323(7308):307-10. doi: 10.1136/bmj.323.7308.307, Hadad S, Hadad Y, Simon-Tuval T. Determinants of healthcare system’s efficiency in OECD countries. The European Journal of Health Economics 2013;14(2):253-65. doi: 10.1007/s10198-011-0366-3, de Cos PH, Moral-Benito E. Determinants of health-system efficiency: evidence from OECD countries. International journal of health care finance and economics 2014;14(1):69-93. Kinfu Y, Sawhney M. Inefficiency, heterogeneity and spillover effects in maternal care in India: a spatial stochastic frontier analysis. BMC Health Services Research 2015;15:118.) use health expenditure as part of their modelling exercise. Third, as a matter fact excluding such a crucial variable from the model would also make the production process incomplete. For these reasons, we think our approach is justified.
3. This study ignored a lot of literatures about Chinese Healthcare based on Bootstrap DEA approach for lacking a systematic review.
Response 3
The second paragraph on the first page describes what is already known and the gap in the literature on health system efficiency analysis in China. As we pointed out, most existing studies are ‘activity’ (as opposed to outcome) based and focus on health institutions (such as hospitals) rather than the whole health system which we are trying to address in the present work. In this exercise, we have used all available articles that we found to be relevant for our paper and published in the English language media, but we are open to include any missing work if they are specifically indicated to us.

4. As aforementioned, the limitations of the study didn't discuss adequately.
Response 4
The discussion section extends to almost to two pages (which we feel is fit for the purpose), and we also have a separate section on limitations as well. As we indicated above, we believe the review of available literature is exhaustive and doesn't need to be reflected as a limitation.
5. Also, there are several mistakes (in my opinion) in the manuscript, such as "in China and internationally"（the first sentence of the Abstract).
Response 5
We feel the paper is well composed and written in an intelligible language. We also feel that the statement referenced above is grammatically correct. However, we have corrected a spelling error identified on page 13, line 44 ("ssystems" as "system").

Reviewer: 3
Reviewer Name: Li Wang
Institution and Country: McMaster University, Canada Please state any competing interests or state 'None declared': None declared
Thank you for your time and expertise to review our paper, and for the observations and comments on the paper. We really appreciate it.

Introduction:
1. The author need to describe the health care system in China, such as the components of public financing or private financing? Does the system apply for all provinces including municipalities or autonomous regions? Any medical program for the child, senior or the specialty group?
Response 1
We have now included the following paragraph as our first paragraph in the section on 'input and output variables on the page 4, line number 33.
The Chinese health system is made up of a hybrid system, where by both private and public delivery of health services and financing schemes co-exist. Regarding health financing, there are mainly two health insurance systems in China: the medical insurance scheme designed for urban employees and the system available for rural residents and those who are not in the labor force. The second scheme includes children and the elderly in urban areas as well as rural resident who are covered through the new rural cooperative medical schemes. A similar scheme is applied throughout the country, and the provincial health system.

2. Line 24: need more detail to describe the overall uneven distribution of resource and health disparities (Not just an example). Do you expect that the inequality of health and resource use would be reduced through the study of the efficiency of health care system? If yes, why and how?
Response 2
We understand the feedback but a detailed discussion on disparities in outcome and health inputs is beyond the scope of the present study, as there are also other studies that are fully devoted on this matter. We are also mindful of the word limit available to us. Within these boundaries, the examples in the paper are primarily meant to show that differences in health outcomes are not necessarily a result of low inputs and that in some instances improving health system performance requires a combination of increased efficiency of existing resources in low outcome but high resource settings, while infusing additional resources in others that are already functioning efficiently.

3. Health care system or health system? It's better to keep consistent over the content
Response 3
We agree with the suggestion. We have now used the phrase "health system" consistently throughout the work.

4. Paragraph 2: literatures are restrictive with the Chinese studies? OECD countries and USA have developed quite a few studies on the efficiency of health care systems. What do you learn from them and what's the difference of your study?
Response 4
It is correct that the literature on health system efficiency in China is limited, that is why we have actually proposed the present study. Where we see the similarities of our work with others is in the methodological approach and the framework for analysis, but beyond that the implications of the results need to be treated with care, because in our case the outcome variable is a composite indicator generated out of three elements, maternal mortality, child mortality and premature mortality and morbidity form non-communicable conditions. There is no other study in any part of the world that has captured health outcomes in the way we did. Ours analysis captures all major population segments (children, mothers and adults). In contrast, most existing studies are based on a single outcome or population group.

5. Are there literatures on the system reviewers of the studies of health system efficiency in China?
Response 5
Yes, we have reviewed these studies in China and cited some in the paper. For further details, please look at references 10, 14, 15-16.

Methodology
1. Province in this study is the DMU. How many? Are the provinces, municipalities, and autonomous are comparable?
Response 6
Our analysis is based on provinces as DMUs. We have now clearly indicated the number of DMUs used in the study as shown on page 4, line number 33. "To examine health system efficiency at the provincial level, we have included all the 31 administrative areas in mainland China. These provinces form the DMUs in our analysis."

2. DEA: what's your study target to? OTE or SE, why?
Response 7
Our study is focused on overall technical efficiency as well as its pure technical efficiency and scale efficiency components. Specifically, given the importance of global technology in the DEA model and the absence of a definitive conclusion of RTS, a two-part returns-to-scale test as proposed by Simar and Wilson 22 was performed. The test result showed the null hypothesis of CRS was rejected at a 99% level of significance. Therefore, a VRS technology was implemented and accordingly obtained both OTE and SE components.

3. Output orientation: why does your study focus on the output orientation only? Why not input orientation? The policy implication from this study should not be just in the short run?
Response 8
The main reason for implanting an output orientation model is linked to the strategic goals of Health China 2030, which is focused on improving population health. Second, health systems are often organized on the principles of the desirability of maximizing health itself (Evans et al. 2001; Smith and Street, 2006). Hence, any findings about how to improve health level given input will be more directly useful for this strategic action. That said, we agree an input orientation strategy also generate useful results, but this is not considered in this study.

4. CRS is not suitable for the health sector.
Response 9
As indicated in our response (Response 7) the VRS approach was adopted based on the returns-to-scale test as proposed by Simar and Wilson.

5. Bootstrapping DEA and truncated bootstrapping regression: the main reasons adopting the bootstrapping DEA are serially correlation of the efficiency scores and the random errors in the point inefficiency.
Response 10
We agree with you, and that's partly why we used it in our study.

6. Did you use the double bootstrapping process developed in the Simar and Wilson (2008)? Or just the two separated bootstrapping in the efficiency estimation and the 2nd stage of the regression?
Response 11
Yes, that is correct. The model we used was fitted jointly.

Data and Variables
Input and Output variables:
1. more detail description – what medical personnel include? Physicians? Specialists? Nursing? Pharmacist? Other medical stuffs? If including all types of medical personnel, how do you adjust their various levels of skills?
Response 12
Medical professional category includes physicians (namely general practitioners and specialists), but do not include nurse, pharmacist or other medical staff. We have added this explanation in the draft paper now, page 4, line number 35.

2. Does the salary of the medial personnel include in the total health expenditure? If yes, it might double count the physician input. Also, it's necessary to describe what's the total health expenditure? What does it include? Hospitals, physicians, drug, dental, vision, long-term care facilities?
Response 13
We have now added a sentence explaining the components of health expenditure captured on our analysis on page 4, line number 38.

3. Does the capital investment include in the health expenditure? If yes, the hospital beds are double counted.
Response 14
The data on hospital beds refers to density per thousand population and not expressed in monetary terms, hence we don't think there would be measurement bias affecting the analysis. That said it is true that expenditure on hospital beds forms part of capital investment but there are no separate data on expenditure on hospital beds to sperate that out. However, given that our analysis is focused on the entire health system (not just hospitals) the investment attributable to hospital bed is expected to be a small proportion of capital investment, hence the effect of 'double counting' can be assumed to be minimal.

4. Line 46: what's the age cut-off for the premature mortality?
Response 15
We believe this question is related to the NCDs-based healthy life years we used in our analysis, which is based on the original data from the China CDC and Global Burden of Disease study from the Institute of Health Metrics and Evaluation. According to IHME, premature mortality is defined based on the lowest mortality level observed at each age in the global population and corresponds to a life expectancy over 85 years. In the case of China, the age cut-off for premature mortality from NCDs proposed in Health China 2030 is 70. More information can be found in "Healthy China 2030" Planning Outline Tutoring Reader ("健康中国"2030规划纲要辅导读本) authored by National Health and Family Planning Commission.

5. Line 51: What are the health worker-related factors and poor-supply factors? Any specific factors include in your 2nd stage of DEA analysis?
Response 16
The purpose of second stage analysis is to examine the potential impact of external factors (i.e. factors that are beyond the direct control of the health system) on efficiency level, which was measured in the first stage of our analysis. The specific variables captured in the second stage analysis are described on the page 5, line number 9, and also can be found in Table 1.

6. It's better to provide couple of references that convert the mortality rate to the survival rates.
Response 17
References have been added that convert the mortality rate to the survival rates on the page 5, line number 3.

7. How do you convert the YLD into health life years?
Response 18
NCDs mortality and YLD was converted into NCDs-based healthy life years (HLY_NCDs), whereas only mortality and morbidity caused by NCDs were considered when calculating healthy life years. The concept of healthy life years has emerged as one of the more commonly used health status measures that incorporates both mortality and morbidity. A reference has been added in the paper on the page 5, line number 7.

8. Have you considered to use the inverse of the IMR and MMR?
Response 19
We are not sure what is meant by 'inverse'. But what we have used is "survival probabilities", which are a direct complement of "death probabilities" and are preferred in our analysis given their intuitive meaning and wider use in relevant literature.

9. What's the OOP in line 11 of page 5?
Response 20
OOP refers to out-of-pocket health expenditures. We have now added the full name on page 5, line number 11.

10. Why do you select these factors in the 2nd stage of DEA.? More description, rational, and references are needed.
Response 21
Variables for analysis were selected based on evidence from similar previous studies and availability of data. We have now added the following sentence on page 5, line number 14.
Data Source
Re-organize the order of the data sources. Output, input and factors.
Response 22
It would be clearer if list data source in the order of output, input and factors. However, given the condition that one data source usually involves output, input and environmental factors, take China Health and Family Planning Statistical Yearbook 2016 for example, MMR, medical personnel density, hospital bed density, and percentage of high-risk pregnancy in the analysis are all from it. To control the number of words, list data source by data sources would be helpful.

Results:
Descriptive statistics
1. Number of provinces, municipalities and autonomous regions?

Response 23

We have revised and added description in the section of descriptive statistics on the page 5, line number 39.

2. The good DEA model requires the input and output variables should be in scales manageable for DEA software and the variables have a similar variation. So Data usually will be rescaled to make them feasible for the DEA models. From the table 1, the scale and variation are quite large. The author may need to rescale the data.
Response 24
We tried various forms of transformation but in our case the analyses with the original variables were more feasible.

3. What are the two provinces with 11.07 year gaps?
Response 25
Those were Tibet and Shanghai. We have now added a description to make this clearer to readers. See page number 5 and line number 45.

4. Figure 1: correlation tests for the inputs and outputs? I think it's not necessary to look at the association between the input and outcome. More interesting, may look at the efficiency change with the inputs?
Response 26
We have now removed Figure 1 and related description as suggested.

Efficiency estimates
1. Scale efficiency is not mentioned as the research objectives in the introduction section. Why you include this in the result section?
Response 27
It is a known fact that technical efficiency is decomposable into pure technical efficiency and scale efficiency, especially if the system is found to operate under variable return to scale process. As we indicated earlier, our scale test has confirmed that this is indeed the case for the study area, hence the reason for decomposing overall technical efficiency and including the scale efficiency component in the results and discussion sections.

2. Line 15 of page 7, what's the geographic location?
Response 28
The geographic location refers to the 31 provinces in the country.

3. Table 3: all the descriptive statistics are efficiency scores. Why here regression on the inefficiency?
Response 29
We have now converted the regression analysis to efficiency score to make them consistent with the results of the descriptive part.

4. There are only 31 provinces but with 10 factors. The small sample size does create problems that I would think need to be addressed.
Response 30
To address the issue, we have used bootstrapping technique and 2000 repetitions had been used for the regression analysis. We believe this practice can help overcome the concerns. That said, the DEA approach is less affected by number of observations and degrees of freedom than a standard regression-based analysis such as the SFA.

5. What's the R2 for the regression?

Response 31
The application used for the analysis does not provide R2, but we believe the overall significance of the model as shown by Wald Chi is rather more relevant. And these are now included in the output. See Table 3 on page number 11.

Discussion
1. Line 54 of page 12. Why don't use outpatient-inpatient ratio but two separated factors at the 2nd stage of DEA?
Response 32
Both approaches are valid. The two separate variables used in our analysis can help answer the amount of outpatient and inpatient impact on efficiency of health system. We the ratio indicator can only reflect the relation of outpatient use and inpatient use impact on efficiency.

2. Line 44 of page 13: wrong spelling in "system".
Response 33
We have addressed the spelling mistake.

3. Need to address the limitation of small number of the DMUs in the regression analysis.

Response 34
We are aware of work in the literature with even fewer DMUs than ours, and in fact the main advantage of the DEA over its SFA equivalent is its ability to handle fewer DMUS. That said, as we indicated in our response to 30, we have also used bootstrapping technique and 2000 repetitions for the regression analysis. We believe this practice can help overcome the concerns.

VERSION 2 – REVIEW

REVIEWER	Li Wang McMaster University
REVIEW RETURNED	10-May-2019

GENERAL COMMENTS	The statistical power in the 2nd stage of the linear regression is problematic. By giving only 31 DMUs, only 2-3 variables are allowed to include in the model. The accuracy of the coefficients and standard errors are need to be concerned.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 3

Reviewer Name: Li Wang

Institution: McMaster University

Thank you for your time and expertise to review our paper and for the following comment on the paper.

“The statistical power in the 2nd stage of the linear regression is problematic. By giving only 31 DMUs, only 2-3 variables are allowed to include in the model. The accuracy of the coefficients and standard errors need to be concerned.”

Response

To address the concerns on degrees of freedom, we re-estimated the model by reducing the number of independent variables that went into the model. Specifically, in the current analysis, we constructed a composite index of socioeconomic status derived from rate of urbanisation, education attainment, and disposable income per capita, all of which had been proven to be associated with health in previous studies. In addition, we have excluded several variables (such as access to improved water source (i.e. tap water), emission rates, percentage of high-risk pregnancy, and outpatient care) that were not significantly associated with the dependent variables at the bivariate level. These revisions allowed us to reduce the number of independent variables to three.

VERSION 3 – REVIEW

REVIEWER	Li Wang McMaster University
REVIEW RETURNED	02-Jul-2019

GENERAL COMMENTS	I recommend to accept the manuscript.