 doi: 10.18502/htaa.v6i1.11132

Review Article

Measuring Health System Efficiency; A Protocol Study

Farhad Hosseinzadeh Lotfi[1], Alireza Olyaeemanesh[2,3*], Efat Mohamadi[2], Reza Majdzadeh[4], Ali Akbari Sari[2,5], Iraj Harirchi[6], Ali Akbar Haghdoost[7], Hamid Sharafi[1], Haniye Sadat Sajadi[8], Zahra Goodarzi[3], Somayeh Noori Hekmat[9], Mohammad Mehdi Kiani[2,5], Leila Freidoony[2,10], Amirhossein Takian[2,5,10**]

[1] Department of Mathematics, Science and Research Branch, Islamic Azad University, Tehran, Iran
[2] Health Equity Research Center (HERC), Tehran University of Medical Sciences (TUMS), Tehran, Iran
[3] National Institute of Health Research (NIHR), Tehran University of Medical Sciences (TUMS), Tehran, Iran
[4] Community Based Participatory Research Centre and Knowledge Utilization Research Centre, Tehran University of Medical (TUMS), Tehran, Iran
[5] Department of Health Management and Economics, School of Public Health, Tehran University of Medical Sciences (TUMS), Tehran, Iran
[6] The Cancer Institute at Imam Khomeini Hospital, Tehran University of Medical Sciences (TUMS), Tehran, Iran
[7] Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran
[8] University Research and Development Center, Tehran University of Medical Sciences (TUMS), Tehran, Iran
[9] Research Center for Health Services Management, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran
[10] Department of Global Health and Public Policy, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran

**Corresponding Author:* Alireza Olyaeemanesh, Health Equity Research Center (HERC), Tehran University of Medical Sciences, Tehran, Iran; National Institute of Health Research, Tehran University of Medical Sciences, Tehran, Iran. Email: arolyaee@gmail.com.

***Corresponding Author:* Amirhossein Takian, Health Equity Research Center (HERC), Tehran University of Medical Sciences, Tehran, Iran; Department of Health Management and Economics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran; Department of Global Health and Public Policy, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran. Email: takian@tums.ac.ir.

Received 2022 January 16; Accepted 2022 January 27.

Abstract

Background: To improve healthcare services' quality, countries should measure their health systems' efficiency and performance by robust methods.

Objectives: We aimed to develop a national study to measure the efficiency of the health system in Iran.

Methods: The literature review identified several methods for measuring efficiency; the most common one was data envelopment analysis (DEA). We adopted DEA, but its findings were simplistic and inaccurate, so we began to modify the method by determining the weight of each indicator. We identified the efficiency measurement indicators, in line with international standards and uniformed units, and then readjusted our input/output indicators according to the study context through four expert panels. We collected data and classified the input/output indicators, followed by determining each indicator's weight and standard limits. Then we rationalized our previous results by applying the revised model. The initial new results of the refined model were valid, accurate, and consistent with previous studies, as approved by experts. We defined proper modeling to achieve the stated objectives. After investigating various DEA models, we finally designed a new model that was consistent with the existing data and conditions, entitled EDEA (extended DEA), to analyze other subprojects.

Conclusions: The conventional DEA methods may not be accurate enough to measure health systems' efficiency. By modifying modeling process, we propose a modified DEA with a very low error rate. We suggest that others interested in measuring health system efficiency adopt our modified approach to increase accuracy and create more meaningful policy-oriented results.

Keywords: Efficiency; Health System; Productivity; Protocol

1. Background

Measuring efficiency has always been a major concern for senior managers across all sectors. Due to resource constraints, policymakers and managers' ultimate goal is to achieve the highest attainable performance using the lowest facilities. To maximize productivity in any organization toward performing its short, medium, or long-term plans, it is necessary to determine the highest possible production capacity while consuming a limited range of resources.

Health systems face spiraling costs due to ever-increasing demand for the use of expensive technology in responding to long-term conditions while decreasing resources and rising financial constraints. Therefore, identifying the current status of efficiency across various sections in any health system is an essential step in planning to improve the quality of healthcare services. Robust evidence about the health system performance is crucial to the maintenance and promotion of community health. In particular, measuring health system efficiency through the use of a valid conceptual framework and sound method, can help policymakers design an effective healthcare system and improve its performance. Appropriate measurement of health system efficiency can help stakeholders understand its strengths and weaknesses, improve performance through targeted measures, and avoid losing valuable health resources by preventing the adoption of non-scientific and inappropriate decisions. Thus, we conducted a national study to measure the efficiency of Iran's health system, and to draw evidence-based options for improvement. We first adopted the conventional data envelopment analysis (DEA) for measuring efficiency and faced some challenges. Now, we have modified our method to apply a modified DEA, developed by our team to measure the efficiency. This paper reports the challenges of traditional DEA and illustrates our way to overcome the methodological bugs.

Our initial review of the health-related literature revealed several methods for measuring efficiency in different parts of the health system (1-3), while conventional data envelopment analysis (DEA) was the most frequent method used (4-7). To meet all objectives of our research, we also reviewed the available performance evaluation methods, including the multiple attribute decision-making methods (MADM) such as TOPSIS, VIKOR, ELECTERE, etc. We noticed that these methods rank options only based on multiple criteria, while they fail to identify the weaknesses and strengths and the degree of progression or regression of the units (8-11), both of which were among the main goals of our research. Based on the literature and seeking the views of experts, we finally decided to use the conventional DEA, based on mathematical planning for evaluation of performance and identification of the strengths and weaknesses of congruent units to measure the health system efficiency in Iran.

Nonetheless, the research team found the first round of findings simplistic and unacceptable. The reason for this was clear. The conventional DEA calculates efficiency with input and output. The units with minimum input can therefore be identified effective, irrespective of the quality of healthcare services that they produce. Our research team, advisors, and experts were aware of the complexities and various components that may affect the health system. To overcome the challenge and study the efficiency in standard conditions, we aimed to readjust our input/output indicators so that the least standard necessary input would be allocated to the health sector to achieve the least outcome in each unit. This led us, in consultation with a number of experts in the field, to reexamine our conventional DEA method used and apply some changes in the model based on the assumptions and principles governing the study context, as well as the input and output indicators of each phase of the research. We began to collaborate with experienced scholars in applied mathematics from outside the health sector who specialized in efficiency measurement methods, and added them to the research team as methodological advisers. We then held four panels of experts in the presence of the entire research team and new advisors, during which we discussed and analyzed the limitations of conventional DEA methodology in measuring the efficiency of the health system and the required configurations to overcome the challenges. Finally, the required modifications were discussed, and the entire research team approved the final revisions of the model by consensus.

In this article, we unfold the journey of modifying the conventional DEA toward the creation and application of the "Modified DEA" in measuring the efficiency of the healthcare system in Iran. Our study can be used as a guide for researchers who aim to design and implement similar studies in Iran and other settings.

1.1. Health System in Iran

Being among few countries in the world with integrated medical education into healthcare services, Iran's health system has unique governance and structure, with the Ministry of Health and Medical Education (MoHME) as the stewardship of the health system (12). The MoHME, through its extensive network of over 60 universities of medical sciences (UMSs) across 32 provinces in the country, is committed to implementing the goal of a healthy community through designing and implementing national health policies. The UMSs are responsible for education research and healthcare provision for people living in their catchment area through a national healthcare network, as presented in Figure 1 (13, 14). During the past decades, several reforms have been conducted in the health system of Iran, e.g., establishment of primary healthcare (PHC) network for provision of basic healthcare, expansion of social health insurance, and the implementation of recent health transformation plan (HTP), all of which have contributed to improving health indicators over the past years in Iran (15).

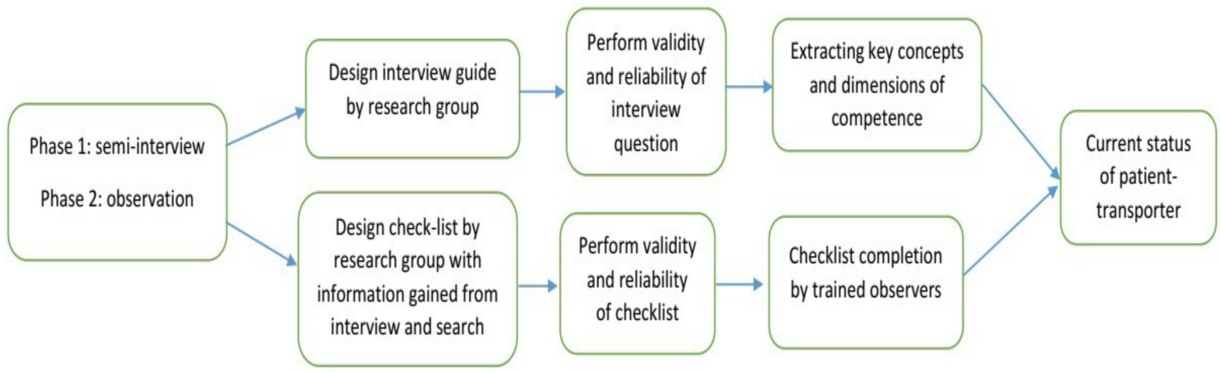

Figure 1. Structure and connection of different parts of the health system in Iran (14).

2. Objectives

As countries are planning to reach sustainable health development through various means, such as universal health coverage (UHC), it is necessary to measure the efficiency and performance of their health system through robust methods. We were approached by the MoHME of Iran to run a national project entitled: "Measuring the Efficiency of Iran's Health System". To reach UHC by 2025, since 2014, Iran has started a comprehensive health transformation plan (HTP) (16). The findings of our research are expected to provide policymakers with insight and awareness about how to improve HTP implementation. Considering the extensive nature of the research, we divided our research into a few measurable subprojects.

Our research entailed a comparison of a number of congruent units in each subproject. We, therefore, used comparative techniques to create Decision-Making Units (DMUs), including countries, provinces, and hospitals, for comparison purposes.

3. Methods

To measure the efficiency of each project, we will determine the followings: Efficiency of each DMU, Ranking of each DMU, Identification of the strengths and weaknesses of each DMU, Measuring progression or regression for each DMU.

First of all, we conducted a comprehensive literature review to identify various methodologies used to meet similar objectives both in Iran and globally. Efficiency is a multi-criteria phenomenon. Therefore, a number of indicators and criteria, sometimes contradictory, should be selected from a finite number of choices, and their weaknesses and strengths should also be evaluated. We carried out our research within eight distinct and sequential steps:

3.1. Step 1: Identification and Definition of Efficiency Measurement Indicators

We began with a comprehensive literature review to identify the input and output indicators to measure the efficiency of the health system in each subproject in line with the study objectives. First, we did a comprehensive review for the period of August 2014 to August 2018. We searched international databases, i.e., PubMed/MedLine, Social Sciences Database, and Google Scholar in English, plus the Jihad University Database (SID) and Google Scholar in Persian . We used various MeSH terms, i.e., efficiency in hospitals, indicators of producing health, UHC indicators, efficiency in health system, DEA, and health system productivity. We included the studies that were published in English and in different levels of the health system, i.e., hospitals, healthcare centers, and international comparisons. In terms of efficiency, all studies that assessed various aspects of efficiency and productivity through descriptive analysis, systematic review, as well as analytical studies were included. We excluded protocol studies.

Then, we organized the findings and extracted indicators. We reviewed the data for each indicator and the reliability of the data sources by checking them in electronic databases, hospitals, and workforce information database of the MoHME, global health observatory of World Health Organization (WHO), and World Bank (WB). This led to determining the final measurable indicators with available data.

Finally, we reviewed all indicators through a set of meetings with experts, i.e., researchers and established policymakers in the field of management and health economics. The panel was described for policymakers in terms of the importance and necessity of each indicator based on the study objectives and importance, and the final indicators of the study were determined by consensus. The indicators' identifiers, including definition, the data collection sources, and the calculation method were developed for each indicator.

3.2. Step 2: Data Collection

During this step, according to the sources identified in step one, we began data collection, most of which were secondary data obtained from the WHO, WB international data, and the Health information systems of the MoHME.

3.2.1 Data Collection Tool

In order to collect data, we designed an Excel sheet checklist based on the study indicators and the study years. Then, we extracted and recorded the data for the years in four months. To ensure data accuracy that came from various sources and was missing in some years in some DMUs, we cleaned up the data by checking for each indicator, each DMU, and each year. Irrational data were checked with other sources, and the correct number was replaced after ensuring the integrity of the number. Due to the limited number of input and output indicators in some studies, a DMU, which lacked data for one indicator in a year (or years), was excluded. In the end, when we assured data accuracy, we finalized our data Excel sheet according to the DMUs' sequence and the years studied.

3.3. Step 3: Classification of Input and Output Indicators

We reevaluated the indicators identified in the first step and determined their definitions and methods for analysis. Based on the assumptions of the health system and the world-class standards for each indicator, we attempted to analyze the efficiency in logical conditions according to the reality of health systems. There were two categories of indicators: So-called "input", the first category included the indicators that were essential for the executive process of an organization. The less the input was, the better the performance could be. This category of indicators could be provided to the management team to improve the organization's performance. The second category included criteria that reflected the organization's performance after using inputs. These performance indicators were referred to as the organization outputs, increasing which might improve the organization's performance, i.e. life expectancy, under-five mortality rate, and maternal mortality rate. meaning that the higher they were, the better the health system performance might be. These are called desirable and undesirable outputs, respectively.

3.4. Step 4: Determination of Weight and Standard Limits for Each Indicator

This step aimed to balance and rationalize the results by changing the model. After recognizing and clarifying the conditions of each indicator (step 3), we determined the standard and weight for each indicator. The goal was to measure the true value of each indicator and show the difference in unit spacing. The indicators were also weighed to identify the more important indicators with a higher impact on the ultimate goal of each unit that would determine the performance score ultimately.

In order to determine the standard of current health expenditure (CHE) per capita and General government health expenditure (GGHE), countries with out-of-pocket payment (OOP) of less than 20% in 2015 (goal set by UHC) were first listed. We then computed the rate of CHE and GGHE for these countries and examined the data dispersion (to decide whether the mean or median of numbers is better for determining the standard). According to the data dispersion, the median of available numbers was used as the standard for CHE and GGHE indicators (the median of countries with OOP of less than 20%). This number was calculated to be at least 1,636 for CHE. Therefore, we corrected this indicator for each country using the following equation:

$$CHE\ included\ in\ the\ model\ = \begin{cases} CHE, CHE \geq 1636 \\ 1636 + (1636 - CHE), CHE < 1636 \end{cases}$$

In fact, a penalty would be imposed on any country below this standard. Given this limitation in the model, the standard CHE for each country was considered to be > 1,636.

Similarly, the minimum necessary GGHE in each country was calculated to be at least 1,106, and hence, the following changes were made in data related to this criterion in the modeling process:

$$GGHE\ included\ in\ the\ model\ = \begin{cases} GGHE, GGHE \geq 1636 \\ 1106 + (1106 - GGHE), GGHE < 1636 \end{cases}$$

In fact, a penalty would be imposed on any country that expended less than this standard. Given this limitation in the model, the following limitation was added to provide this condition to the newly proposed weight.

$$Standard\ GGHE\ for\ each\ country\ >\ 1106$$

In order to determine the standard of other study indicators, the best performance among the study units was considered. For example, the standard of 83.7 was considered for Japan's life expectancy indicator (17), which was the highest among the countries studied. According to experts, the life expectancy difference of each country and 83.7 was considered the output. In this case, the obtained difference is an undesirable output that was applied in the model with the following changes:

$$\overline{y_j} = 83.7 - y_j, j = 1, \ldots, n,$$
$$K = max\{\overline{y_j}: j = 1, \ldots, n\} + \varepsilon$$
$$New, y_j = \overline{y_j} + K, j = 1, \ldots, n.$$

Where, ε is a small positive number.

As previously stated, the under-five mortality rate has an undesirable nature. Therefore, countries try to reduce this indicator; the larger this figure is, the poorer the performance of countries will be. According to the standard of this indicator (2.8, as the best performance among countries) (9), the new standard was derived from the following equation:

$$\overline{y_j} = Mortality\ rate\ of\ children\ under\ age\ five - 2.8, j = 1, \ldots, n$$
$$K = \max\{\overline{y_j}: j = 1, \ldots, n\} + \varepsilon$$
$$New\ indicator\ of\ the\ jth\ country\ = y_j = K - \overline{y_j}, j = 1, \ldots, n.$$

In fact, a standard threshold of 2.8 was considered for each country with this action. If the under-five mortality rate was more than 2.8, we would consider a penalty for each country. In the new indicator, each country with a higher value scored higher in terms of performance. The equations of other indicators were corrected in the same way based on the standard and weight assigned to them. It should be noted that the decision about the weight of the indicators was made during the panel of experts (including the research team members and team advisors).

3.5. Step 5: Modeling

As stated in the introduction, we initially decided to use the traditional and common method of DEA to calculate the efficiency of this research. We supposed that there were M congruent decision makers that the jth unit uses the input vector to generate the output vector. Intuitively, each unit could be represented as follows:

$$\begin{pmatrix} x_{1j} \\ M \\ x_{mj} \end{pmatrix} \rightarrow \boxed{DMU_j} \rightarrow \begin{pmatrix} y_{1j} \\ M \\ y_{sj} \end{pmatrix}$$

Charnes et al. suggested model 1 to calculate the relative efficiency of DMUp (18) (Box 1):

CCR (Input and Output Axis Model) Envelopment Form	CCR Multiplier Form
$Max\ \theta$	$Max\ UY_p$
$s.t. \sum_{j=1}^{n} \lambda_j X_j \leq \theta X_p$	$s.t. UY_j - VX_j \leq 0, \forall j$
$\sum_{j=1}^{n} \lambda_j y_j \geq Y_p$	$VX_p = 1$
$\lambda \geq 0$	$U \geq 0, V \geq 0$

Box 1.

If (λ^*, θ^*) is the best envelopment form of Table 1, then $0 < \theta^* \leq 1$ its value is called the relative efficiency of DMUp. If $\theta^* = 1$, then DMUp is relative efficiency. Otherwise, if $\theta^* < 1$, DMUp is ineffective. Vector$(\sum_{j=1}^{n} \lambda_j^* X_j, \sum_{j=1}^{n} \lambda_j^* Y_j)$ is called the template point of DMUp.

Given that most indicators and criteria affecting the performance had their own constraints, the envelopment and multiplier models (Box 1) could not answer the research questions. We, therefore, began to transform the model after identifying the constraints corresponding to the criteria that conform to existing rules and standards (Box 1), as follows. To identify all inefficiency factors, including technical and combination, models will be proposed for evaluation based on auxiliary variables:

$$Max\ Z_p = \frac{1 - \frac{1}{m}\sum_{i=1}^{m}\frac{S_i^-}{R_i^-}}{1 + \frac{1}{S}\sum_{r=1}^{S}\frac{S_r^+}{R_r^+}}$$

$$\sum_{j=1}^{n} \lambda_j X_{ij} + \sum_{j=1}^{n} \mu_j a_{ij} + s_i^- = X_{ip}, i = 1, \ldots, m$$

$$\sum_{j=1}^{n} \lambda_j y_j + \sum_{j=1}^{n} \eta_j b_{rj} - s_r^+ = y_{rp}, r = 1, \ldots, s$$

$$\sum_{j=1}^{n} \lambda_j x_{ij} + \sum_{j=1}^{n} \mu_j a_{ij} + s_i^- \epsilon A_i, i = 1, \ldots, m$$

$$\sum_{j=1}^{n} \lambda_j y_j + \sum_{j=1}^{n} \eta_j b_{rj} - s_r^+ \epsilon B_r, r = 1, \ldots, s$$

$$\lambda \geq 0, \mu \geq 0, \eta \geq 0, s^- \geq 0, s^+ \geq 0$$

Box 2.

Where µj is the corresponding weight limits for inputs, and is the corresponding weight limits for outputs. Ai refers to the conditions that the ith input should have, and the constraint is added because the input of the pattern point should also be applied to the input conditions of the observed units.

Similarly, Br is the order of conditions that the rth output must have, and the corresponding environment is added because the pattern point output must be applied in the output conditions of the observed units. The optimal value of the target function (Box 2) that is represented with Z_p^* shows the relative efficiency of the DMUp. If $Z_p^* = 1$, then DMUp is referred to as relative efficiency; otherwise, if $Z_p^* < 1$, then DMUp is called relative inefficiency. For all inefficient units, the improvement rate for each of the inputs and outputs will be obtained by solving Box 2. If (λ*, η*, μ*, s-*, s+*) is the optimal answer of Box 2, then the coordinates of the template point for DMUp, i.e., the coordinates that DMUp must reach to become efficient, are shown in Box 3:

$$\text{Input coordinates of the ith pattern of DMUp} = \sum_{j=1}^{n} \lambda_j^* X_{ij} + \sum_{j=1}^{n} \mu_j^* a_{ij}, i = 1, \ldots, m.$$

$$\text{Input coordinates of the ith pattern of DMUp} = \sum_{j=1}^{n} \lambda_j^* y_j + \sum_{j=1}^{n} \eta_j^* b_{rj}, r = 1, \ldots, s.$$

Box 3.

It is obvious that the coordinates of the pattern point rely on the principle that the inputs should be reduced if possible, and the outputs should be increased if possible. The efficiency obtained from solving Box 2 can also be used for ranking. Obviously, each unit with higher efficiency will rank better.

One of the research objectives was to determine the level of progression or regression of decision-making units during the study period. The Malmquist Productivity Index (MPI) was used to answer this question. This index results from the comparison of changes in the evaluated unit performance with technological changes, meaning that if the improvement of DMUp efficiency was more than the improvement in the production technology (efficiency boundary) in two time points of t and t+1, DMUp progressed. Similarly, if the DMUp efficiency improvement was less than the improvement in the production technology (efficiency boundary) in these two time points, DMUp regressed. If the improvement was equal, there was no change in the productivity rate. This scale can be obtained from following equation:

$$MPI_p = \frac{\Delta E_p}{\Delta T}$$

This calculation is obtained by implementing Box 2 with appropriate changes. As mentioned earlier, the final models are determined to answer the research questions.

3.6. Step 6: Implementation of the Model

In this step, the models designed in step 5 were first coded in General Algebraic Modeling System (GAMS) 23.4 (19) and the R Project for Statistical Computing 3.5.1, and the preliminary results were obtained. Obviously, a corresponding model was written and implemented to calculate the efficiency of each category of DMU. The efficiency table, pattern point coordinates, rank, and MPI were calculated for DMU sets. The units compared were different in each study.

3.7. Step 7: Validation of the Model and Results

During this step, we investigated the preliminary results obtained from solving the designed models. These results were subjected to a slight error that was consistent with the available results and judgments. If the model led to an optimization problem during the modeling, statistical methods could not test and judge the model as it is assumed that all variables, effective factors, and constraints were identified during modeling.

In this study, Box 1 was first used to calculate the efficiency. The results of the Box 1 regarding the calculations of efficiency and ranking did not match the facts; this was confirmed with the previous results (i.e., the results are inaccurate). Therefore, the process was reconstructed and reexamined. This time, Box 2 and the conditions governing the data in the modeling were designed. The preliminary results of the designed models in step 5, with a very small error, were approved by the experts.

3.8 Step 8: Results analysis

Following the implementation of step 6, which included the implementation of the designed models to achieve efficiency goals, patterns, rank, progression, and regression, the results were presented in tables and charts. Two main works were carried out in this step: First, we analyzed the efficiency equation results, the efficiency ranking, progression, and regression of each unit with their initial data, assessing whether a member with a better ranking had better initial indicators than the other members, followed by the opinions of experts regarding these results. Second, we analyzed the relationship between the efficiency score of each unit with the contextual and uncontrollable factors that did not interfere with efficiency. If it was possible to identify contextual factors that had a direct or indirect relationship with efficiency, ranking, progression, and regression, a medium-term and long-term solution could be developed to improve the situation considering these variables. We summarize all these eight steps as a flowchart in Figure 2.

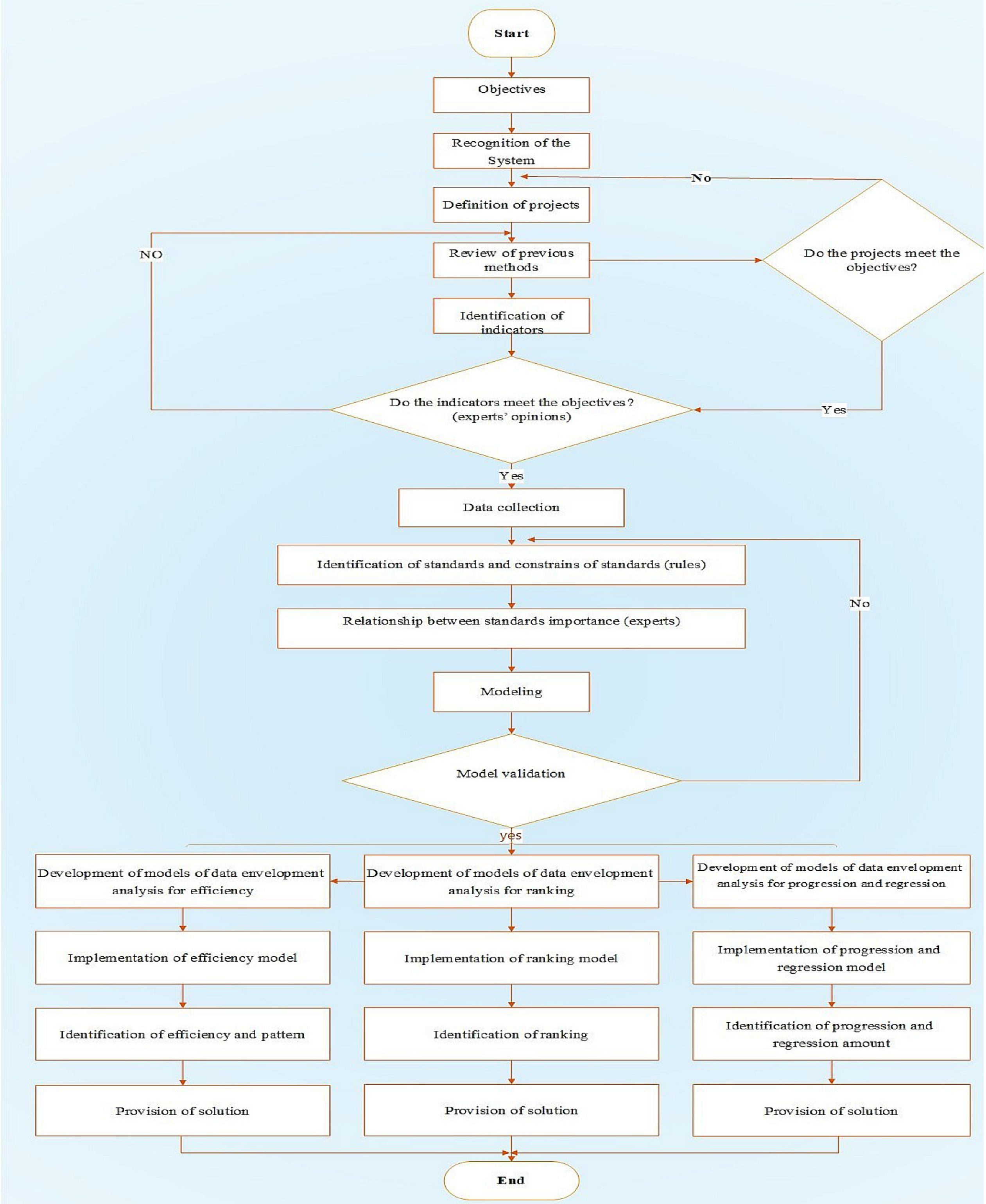

Figure 2. Implementation process of measurement of the health system efficiency of Iran.

5. Discussion

The main objective of this study is to measure the health system efficiency in Iran. We defined a proper modeling to achieve the stated objectives. All calculations presented earlier are used to measure the efficiency of Iran's health system in comparison with other selected countries. Since one of the important goals was to identify the strengths and weaknesses of each unit, DEA was the only technique that could achieve this goal. After investigating various DEA models, we finally designed a new model that is consistent with the existing data and conditions.

A systematic review and meta-analysis of OECD countries showed that various methods were used to calculate country-level efficiency. Studies that used SFA and DEA were stronger than others, while DEA was the most frequent method for efficiency measurement. Most studies used life expectancy as the most important indicator to measure health production efficiency in many countries. The authors argued that the methodological challenges of inter-country comparisons of the health system efficiency could not help policymakers identify appropriate strategies (2).

We also used the methodology firstly employed in a global research by WHO on comparative efficiency analysis of national health systems in 191 countries. The inputs for this study were per capita health expenditure, under-five mortality rate, neonatal mortality rate, and life expectancy. In addition, the mean income and mean years of education were used as controlled variables outside the health sector which can influence the health outcomes, as well (7).

A recent study in the WHO Eastern Mediterranean region used similar indicators to our study and also utilized the DEA method and MPI to compare and analyze the data, while their outputs were neonatal mortality rate and life expectancy. The advantage of our study is to use the GGHE to CHE ratio to take into account the fair financial participation in the health sector (3).

The results of a systematic review of hospital efficiency (20) showed that eight methods, i.e., Pabon Lasso, BSC, EFQM model, Baldrige model, DEA, accreditation, ratio analysis, and models hybrid, were generally used to measure hospitals efficiency, among which the DEA method was the most widely used. Investigating the input and output variables in a systematic review study (21) showed that the number of beds, the length of stay, the number of visits, and the number of surgical procedures were the most widely used input indicators. The DEA method was similarly used in this study, and these variables were considered a part of indicators for measurement. In addition, we included all public hospitals in the country and categorized them into various specialties. Although most similar studies measured efficiency in a cross-sectional study design, we are measuring the trend of efficiency in all categories of public hospitals over a period of five years (2012 - 2016). Based on this research's findings and methodological observations, to measure the efficiency of any part of the health system, we propose that the selection of input and output indicators should be in line with the international standards, the indicators unit (monetary, volumetric, relation) should be uniformed, the number of DMUs should be at least three times more than the input and output indicators, and the current two-stage and three-stage DEA models should be combined with the bootstrap-DEA method to give more accurate efficiency scores (21) and reach more tangible results.

6. conclusion

Based on our observations, the conventional DEA methods might not be accurate enough to measure the efficiency of the health systems. Through modification of the modeling process, we proposed a modified DEA that its results were confirmed by experts with a very low error rate. We propose others interested in measuring health system efficiency may adopt our modified approach for increasing accuracy and creating more meaningful policy-oriented results. We should also note that special conditions of an indicator might have a considerable impact on the final model. The exact recognition of the conditions and realities of the indicators and rules governing them can lead to modeling in which the constructed model would correspond to the desired objectives.

Acknowledgments

The authors acknowledge the MoHME's Iran for providing the hospital data.

Authors' Contributions:

AT, AO, and RM conceived the study and designed its method. EM performed the computations and applied the model, with help from FHL, for the revision of the analytical method. All authors discussed the results and contributed to the final manuscript. EM HSH and HY carried out the analytical experiment. AO and AHT wrote the manuscript. All authors contributed to the development and approved the final manuscript. AT is the guarantor.

Author Inclusion Criteria:

Based on the standards of the International Committee of Medical Journal Editors (ICMJE)

Conflict of Interests:

The authors declare that they have no competing interests.

Data Reproducibility:

All relevant data are within the supporting information files and in the following international repositories:

- http://apps.who.int/nha/database/Select/Indicators/en
- http://apps.who.int/gho/data/node.imr#ndx-T
- https://datacatalog.worldbank.org/dataset/world-development-indicators

Ethical Approval:

This study was approved by the Ethics Committee of Tehran University of Medical science, under the license no: IR.TUMS.VCR.REC.1396.4018.

References

1. Hollingsworth B. The measurement of efficiency and productivity of health care delivery. *Health Econ.* 2008;**17**(10):1107-28. doi:10.1002/hec.1391. [PubMed:18702091].
2. Varabyova Y, Müller JM. The efficiency of health care production in OECD countries: A systematic review and meta-analysis of cross-country comparisons. *Health Policy.* 2016;**120**(3):252-63. doi:10.1016/j.healthpol.2015.12.005. [PubMed:26819140].
3. Masri MD, Asbu EZ. Productivity change of national health systems in the WHO Eastern Mediterranean region: application of DEA-based Malmquist productivity index. *Glob Health Res Policy.* 2018;**3**:22. doi:10.1186/s41256-018-0077-8. [PubMed:30083615]. [PubMed Central:PMCPMC6069987].
4. Hunt DJ, Link CR. Better outcomes at lower costs? The effect of public health expenditures on hospital efficiency. *Appl Econ.* 2019;**52**(4):400-14. doi:10.1080/00036846.2019.1646405.
5. Nayar P, Ozcan YA. Data envelopment analysis comparison of hospital efficiency and quality. *J Med Syst.* 2008;**32**(3):193-9. doi:10.1007/s10916-007-9122-8. [PubMed:18444355].
6. Hafidz F, Ensor T, Tubeuf S. Efficiency Measurement in Health Facilities: A Systematic Review in Low- and Middle-Income Countries. *Appl Health Econ Health Policy.* 2018;**16**(4):465-80. doi:10.1007/s40258-018-0385-7. [PubMed:29679237].
7. Evans DB, Tandon A, Murray CJ, Lauer JA. The Comparative efficiency of national health systems in producing health: An analysis of 191 countries. *World Health Organization.* 2000;**29**(29).
8. Hsu C-H, Chang A-Y, Luo W. Identifying key performance factors for sustainability development of SMEs – integrating QFD and fuzzy MADM methods. *J Clean Prod.* 2017;**161**:629-45. doi:10.1016/j.jclepro.2017.05.063.
9. Yazdani M, Zavadskas EK, Ignatius J, Doval Abad M. Sensitivity Analysis in MADM Methods: Application of Material Selection. *Eng Econ.* 2016;**27**(4). doi:10.5755/j01.ee.27.4.14005.
10. Trestian R, Ormond O, Muntean G-M. Performance evaluation of MADM-based methods for network selection in a multimedia wireless environment. *Wirel Netw.* 2014;**21**(5):1745-63. doi:10.1007/s11276-014-0882-z.
11. Mobinizadeh M, Raeissi P, Nasiripour AA, Olyaeemanesh A, Tabibi SJ. A model for priority setting of health technology assessment: the experience of AHP-TOPSIS combination approach. *Daru.* 2016;**24**:10. doi:10.1186/s40199-016-0148-7. [PubMed:27068692]. [PubMed Central:PMCPMC4827190].
12. Danaei G, Farzadfar F, Kelishadi R, Rashidian A, Rouhani OM, Ahmadnia S, et al. Iran in transition. *Lancet.* 2019;**393**(10184):1984-2005. doi:10.1016/s0140-6736(18)33197-0. [PubMed:31043324].
13. Mehrdad R. Health system in Iran. *Japan Med Assoc J.* 2009;**52**(1):69-73.
14. Naderimagham S, Jamshidi H, Khajavi A, Pishgar F, Ardam A, Larijani B, et al. Impact of rural family physician program on child mortality rates in Iran: a time-series study. *Popul Health Metr.* 2017;**15**(1):21. doi:10.1186/s12963-017-0138-0. [PubMed:28576122]. [PubMed Central:PMCPMC5455106].
15. Takian A, Rashidian A, Kabir MJ. Expediency and coincidence in re-engineering a health system: an interpretive approach to formation of family medicine in Iran. *Health Policy Plan.* 2011;**26**(2):163-73. doi:10.1093/heapol/czq036. [PubMed:20732859].
16. Piroozi B, Takian A, Moradi G, Amerzadeh M, Safari H, Faraji O. The effect of Iran's health transformation plan on utilization of specialized outpatient visit services: An interrupted time series. *Med J Islam Repub Iran.* 2018;**32**:121. doi:10.14196/mjiri.32.121. [PubMed:30815416]. [PubMed Central:PMCPMC6387799].
17. World Health Organization. Monitoring health for the SDGs sustainable development goals. Geneva, Switzerland: World Health Organization 2016.
18. Charnes A, Cooper WW, Rhodes E. Measuring the efficiency of decision making units. *Eur J Oper Res.* 1978;**2**(6):429-44. doi:10.1016/0377-2217(78)90138-8.
19. Bussieck MR, Meeraus A. General Algebraic Modeling System (GAMS). In: Kallrath J, editor. *Modeling Languages in Mathematical Optimization.* New York, USA: Springer; 2004. p. 137-57.
20. Bahadori M, Ghardashi F, Izadi AR, Ravangard R, Mirhashemi S, Hosseini SM. Pre-Hospital Emergency in Iran: A Systematic Review. *Trauma Mon.* 2016;**21**(2):e31382. doi:10.5812/traumamon.31382. [PubMed:27626016]. [PubMed Central:PMCPMC5003496].
21. Dong S, Zuo Y, Guo S, Li M, Liu X, Li H. Data Envelopment Analysis for Relative Efficiency Measurement of Chinese Hospitals: A Systematic Review. *Res Health Sci.* 2017;**2**(2):79. doi:10.22158/rhs.v2n2p79.