Mongolian Journal of Economic Review-MJ **Volume 30/19 (2026)**
ARTICLE
https://doi.org/10.69588/mjer.v30i19.5062

Systematic review

Input–Output Specification in Hospital Data Envelopment Analysis: What Choices Drive Efficiency Scores?

Ganzorig Batbaatar [1,2§], Buyantogtokh Batsukh [2§], Amarzaya Batchimed [2], Byambakhorloo Sukhbaatar [2], Byambatseren Boldbaatar [3], Nursabi Khizatkhan [3], Dariimaa Ganbat [3*]

[1] *National Center for Pathology of Mongolia, Ulaanbaatar, Mongolia*
[2] *Graduate School, University of Finance and Economics, Ulaanbaatar, Mongolia*
[3] *Graduate School, Mongolian National University of Medical Sciences, Ulaanbaatar, Mongolia*
[§] *Equally contributed*

Corresponding author:* **Dariimaa Ganbat E-mail: **dariimaa@mnums.edu.mn**

Received: 28 February 2026 | Accepted: 21 May 2026 | Published online: 29 June 2026

Kye words: *Data Envelopment Analysis; Hospital Efficiency; Technical Efficiency; Inputs and Outputs; Model Orientation; Returns to Scale; Case-Mix Adjustment; Quality of Care*

Abstract: *Frontier efficiency methods are widely used to measure healthcare productivity, but results vary greatly because of methodological heterogeneity. There is limited consensus on input-output variable selection, and differences in DEA orientation, returns-to-scale assumptions, case-mix adjustment, and quality incorporation complicate cross-study interpretation. Methods: Following PRISMA guidance, we screened 509 records on public hospital efficiency and included 60 empirical hospital DEA studies. Extracted data covered inputs, outputs, model orientation, returns to scale, DEA variants, case-mix adjustment, quality measures, and sensitivity analyses. Results: DEA was the dominant frontier method. Reported efficiency scores were highly conditional on model specification, commonly ranging from 73% to 97% across studies. In the authors' extraction, most studies used input-oriented models (36/60, 60.0%) and either VRS or CRS/VRS comparisons (42/60, 70.0%). Core inputs were staff (56/60, 93.3%) and beds (50/60, 83.3%), while core outputs were inpatient volume (54/60, 90.0%) and outpatient volume (48/60, 80.0%). Case-mix adjustment was present in 21 studies (35.0%). Quality was directly incorporated into the DEA model in only 5 studies (8.3%) and considered either directly or in parallel in 15 studies (25.0%). Conclusion: Frontier efficiency analysis is useful for benchmarking but is highly sensitive to specification choices. Robust input-output selection, justified orientation and scale assumptions, and explicit case-mix and quality safeguards are essential for valid comparisons and responsible policy use.*

https://doi.org/10.69588/mjer.v30i19.5062

I. INTRODUCTION

Hospitals represent some of the most resource-demanding elements within healthcare systems, and enhancing their operational efficacy continues to be a central focus of policy as nations grapple with financial constraints, increasing demand, and heightened expectations regarding quality and safety. In contrast to industries where outputs can be distilled to a singular product, hospitals provide a diverse array of services utilizing a variety of inputs—including labor, beds, capital, and operational expenses—to yield multiple outputs, such as inpatient treatments, outpatient consultations, procedures, and health outcomes. Within this context, metrics that examine single-output ratios or isolated cost indicators prove inadequate for assessing performance since they overlook joint production dynamics, input substitution opportunities, and trade-offs between service volume and quality (Hollingsworth, 2008; Aigner et al., 1977).

To navigate this intricacy, frontier-based efficiency methodologies have gained considerable traction, with Data Envelopment Analysis (DEA) emerging as one of the most commonly utilized instruments for evaluating hospital technical efficiency (Hollingsworth, 2008). DEA constitutes a non-parametric linear programming technique that appraises the relative efficiency of similar decision-making units by establishing an empirical best-practice frontier derived from observed input-output configurations (Charnes et al., 1978). The original Charnes–Cooper–Rhodes (CCR) model presupposes constant returns to scale (CRS), whereas the Banker–Charnes–Cooper (BCC) adaptation accommodates variable returns to scale (VRS), thus facilitating the distinction between pure technical efficiency and scale efficiency (Banker et al., 1984). Furthermore, DEA models necessitate a choice of orientation: input-oriented models concentrate on possible reductions in inputs for a specific output level, while output-oriented models investigate the potential expansion of outputs with current resources. These modeling decisions reflect distinct managerial and policy viewpoints rather than solely technical considerations.

A persistent methodological issue in hospital DEA research is the pronounced sensitivity of efficiency estimates to model specifications, particularly regarding the selection and quantification of inputs and outputs. Investigations that ostensibly employ identical DEA methodologies may, in reality, encapsulate fundamentally disparate hospital production processes. Labor inputs might be quantified through headcounts or full-time equivalents; capital may be represented by beds, floor area, or fixed assets; and operating resources can be captured through total expenditure or detailed cost categories (Bannick & Ozcan, 1995; Almiman, 2018). On the output side, indicators vary from basic activity metrics such as discharges or outpatient visits to case-mix-adjusted outputs, including diagnosis-related group (DRG) weighted discharges, or advanced models that integrate quality and outcome indicators (Carey & Lin, 2016). Given that DEA constructs the efficiency frontier directly from the chosen variables, modifications in the input-output framework can significantly influence efficiency scores, rankings, and policy implications.

https://doi.org/10.69588/mjer.v30i19.5062

Recent systematic reviews substantiate significant variability in the application of DEA within hospital settings and emphasize that the lack of standardization impedes the comparability and synthesis of findings across various studies (Alatawi et al., 2019; Zubir et al., 2024; Pai et al., 2024). Zubir et al. (2024) primarily concentrated on the methodologies for selecting inputs and outputs in hospital DEA, with their findings aligning with the current investigation by identifying beds, staffing, inpatient services, and outpatient visits as frequently occurring variables. However, this review diverges in three critical aspects. First, it relies on a distinct extraction of empirical hospital DEA research published from 2000 to 2025. Second, it scrutinizes input-output specifications in conjunction with orientation, returns-to-scale assumptions, case-mix adjustments, quality integration, and advanced variants of DEA, rather than merely presenting variable selection as a disjointed list. Third, it articulates these methodological insights within a practical interpretative framework that warns against interpreting high DEA scores as indicative of superior performance when essential quality measures and patient complexity are disregarded. Hence, all numerical data presented in Tables 1 and 2 originate from the authors' extraction of the 60 studies included, unless explicitly stated as contextual evidence from prior reviews.

These methodological nuances bear significant ramifications for policy and management. DEA results are increasingly leveraged to guide resource allocation, payment reforms, and performance evaluations. If observed inefficiencies predominantly mirror the discretionary decisions of analysts—such as labor and capital proxies, output definitions, quality omissions, or scale assumptions—rather than authentic performance disparities, benchmarking initiatives may misguide incentives and provoke inappropriate policy actions (Barra et al., 2022; Smith, 1995; Dranove et al., 2003). Additionally, coding practices are pertinent: case-mix systems may facilitate upcoding, which can skew perceived output and efficiency unless robust data validation measures are in place (Silverman & Skinner, 2004; Steinbusch et al., 2007).

This study aims to bridge this gap by synthesizing the existing literature on hospital DEA with a concentrated emphasis on input-output specifications and associated design considerations, including model orientation, returns-to-scale assumptions, case-mix adjustments, quality integration, and advanced DEA variants. Through a systematic review of empirical hospital DEA studies published since 2000, we aspire to: (1) delineate how inputs and outputs are defined, measured, and categorized; (2) summarize the occurrence of critical modeling decisions, encompassing input versus output orientation, CRS versus VRS assumptions, case-mix adjustments, quality indicators, and the employment of SBM or network DEA; (3) evaluate how these specification decisions correlate with reported efficiency estimates and the comparability of findings across studies; and (4) propose a practical interpretative framework that integrates efficiency and quality considerations. By elucidating which methodological choices exert the most substantial influence on efficiency outcomes, this review endeavors to enhance the credibility of hospital DEA research and facilitate safer, more policy-oriented benchmarking.

II. METHODS

https://doi.org/10.69588/mjer.v30i19.5062

We conducted a PRISMA-guided systematic review with a predefined but unregistered protocol covering objectives, eligibility criteria, search strategy, extraction variables, and synthesis methods (Page et al., 2021; Higgins & Green, 2011). The workflow comprised question formulation, database searching, de-duplication, title/abstract screening, full-text eligibility assessment, standardized data extraction, and structured narrative synthesis.

2.1 Data Sources and Search Strategy

We searched PubMed, Web of Science, Scopus, EMBASE, EconLit, and Google Scholar for English-language studies published January 2000-December 2025, and hand-searched reference lists of relevant reviews. Search strings combined hospital and DEA terms (e.g., "hospital*" AND ["efficiency" OR "performance" OR "productivity"] AND ["DEA" OR "Data Envelopment Analysis"]), tailored per database and iteratively refined. We included peer-reviewed journal articles and conference proceedings; we excluded theses, dissertations, and other unpublished literature.

2.2 Inclusion and Exclusion Criteria

Studies were eligible for inclusion if they met all of the following criteria:

1. Setting: Hospitals or hospital subunits assessed within a multi-hospital framework (hospital- or hospital-system level). Primary care–only studies were excluded.
2. Method: Applied DEA to estimate hospital technical efficiency/productivity. Studies using other methods only were excluded; mixed-method studies were included only if DEA results were separable (extracting DEA outputs only).
3. Outcomes: Reported DEA efficiency scores (or sufficient quantitative output to derive them). Purely methodological papers without empirical DEA application were excluded.
4. Timeframe: 2000–2025 inclusive.
5. Language: English.

Additional exclusions were non-hospital providers unless hospital-only results were isolatable; cost/allocative efficiency-focused "cost-DEA" studies; DEA used for non-efficiency purposes; Malmquist-only studies unless static DEA efficiency scores were also reported; and primary-care-only efficiency studies, which were outside the hospital-level focus of this review (Pelone et al., 2015).

2.3 Study Selection and Screening

The search identified 792 records: 776 through database searching and 16 through reference lists and hand-searching. After removing 283 duplicates, 509 records were screened by title and abstract. Of these, 419 were excluded, and 90 full-text articles were assessed for eligibility. Thirty full-text articles were excluded because they were not DEA-specific, did not focus on hospitals, or lacked extractable quantitative efficiency results. Sixty studies were included in

https://doi.org/10.69588/mjer.v30i19.5062

the qualitative synthesis. Two reviewers independently screened titles/abstracts and assessed full texts; disagreements and borderline cases were resolved by team discussion. **Figure 1** presents the corrected PRISMA flow diagram of study selection.

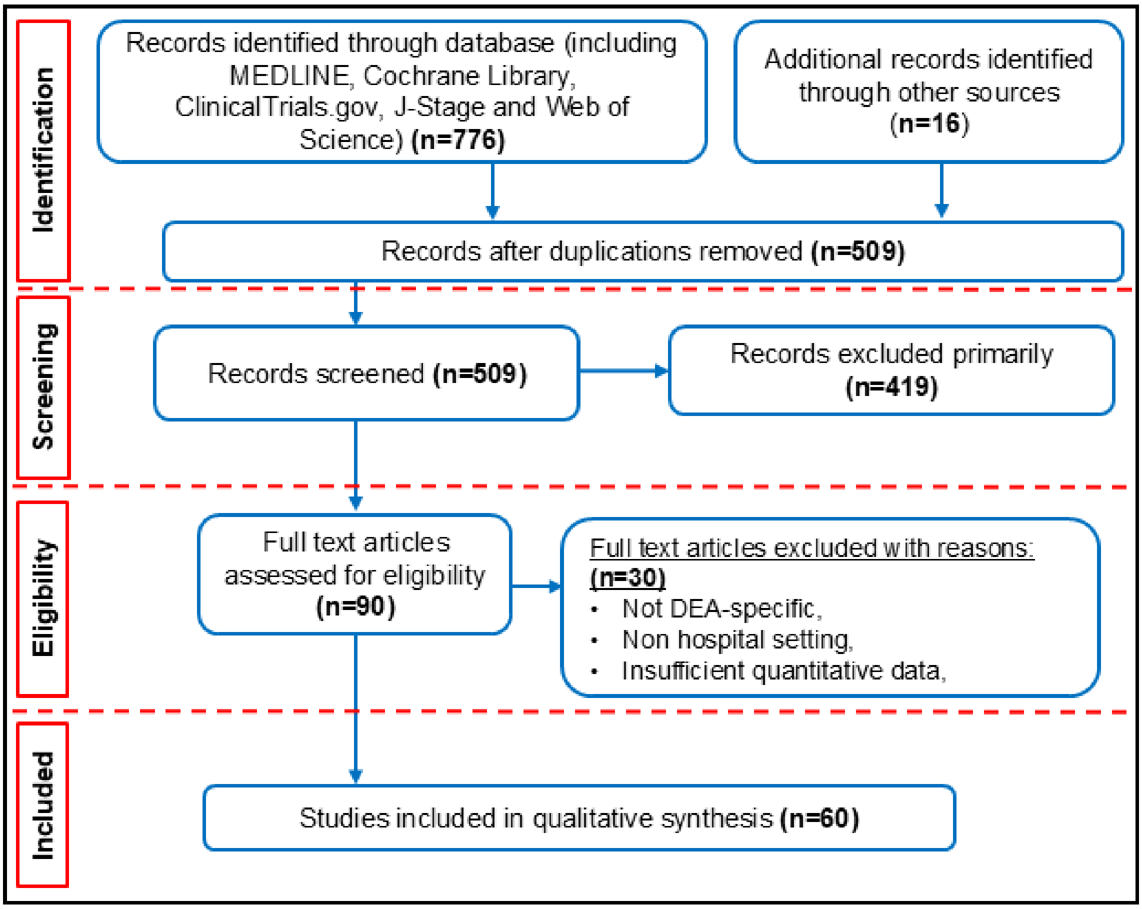

Figure 1. Corrected PRISMA flow diagram of study selection.

2.4 Data Extraction and Synthesis Methods

We developed a standardized extraction form in which each row represented one included study and each column represented a prespecified methodological item: authors, year, country/setting, data period, number and type of hospitals, DEA inputs, DEA outputs, variable definitions and groupings, model orientation (stated or inferred), returns-to-scale assumption (CRS, VRS, or both), DEA variant (CCR/BCC, additive DEA, slack-based measure [SBM], super-efficiency, network DEA, Malmquist, or two-stage designs), case-mix adjustment, quality incorporation, sensitivity analyses, and expert or stakeholder involvement in variable selection. Two reviewers extracted data independently and resolved disagreements by consensus. Counts and percentages in Tables 1 and 2 were calculated from this extraction and are not copied from prior reviews. Because the review was descriptive and methodological, no meta-analysis was attempted. Instead, we conducted a structured narrative and quantitative synthesis, tabulating the frequency of inputs/outputs and key modeling choices and

https://doi.org/10.69588/mjer.v30i19.5062

summarizing how alternative specifications changed efficiency interpretation. In response to the recurrent quality-related limitation, we synthesized a practical efficiency-quality interpretation matrix to guide future DEA reporting.

III. RESULTS

3.1 Study Characteristics

We included 60 studies spanning more than 20 countries across all inhabited continents. Publications increased over time, from relatively few studies in 2000-2005 to substantially more studies in the 2010s and early 2020s. Hospital samples ranged from small regional studies of approximately 10-15 hospitals to national datasets of several hundred hospitals. Most analyses were cross-sectional or repeated annual DEA analyses, with fewer longitudinal productivity studies. Europe and Asia were most represented, including studies from Austria, Ireland, Germany, Italy, Portugal, Turkey, Iran, China, Taiwan, New Zealand, and Türkiye, alongside studies from Africa, the Middle East, and the Americas (Hofmarcher et al., 2002; Gannon, 2004; Herr, 2007; Daidone & D'Amico, 2009; Chowdhury et al., 2014; Fixler et al., 2014; Almiman, 2018; Jahantigh & Ostovare, 2020; Jiang, 2020; Orsini et al., 2021; Sülkü et al., 2023). Prior cross-national taxonomies and reviews were used for contextual comparison only, not as the source of the present extraction counts (O'Neill et al., 2008; Coelli, 2009; Kiadaliri et al., 2013; Hafidz et al., 2018; Mortimer et al., 2017; Pai et al., 2024). Despite wide geographic coverage, input-output choices were broadly similar, although data constraints in some settings led to greater reliance on proxies. Tables 1 and 2 provide the corrected summaries derived from the 60-study extraction.

Table 1: Common input and output variables used in hospital DEA studies (2000–2025).

Variable domain	Specific measures coded in extraction	Studies, n (%)	Interpretation
Inputs: Capacity (beds)	Total hospital beds; staffed beds; ICU or specialized beds when reported	50/60 (83.3%)	Beds remained the dominant capital/capacity proxy, but interpretation varied by whether beds were total, staffed, or specialized.
Inputs: Staff	Physicians; nurses; other clinical staff; administrative/support staff; total full-time equivalent staff	56/60 (93.3%)	Human resources were the most common input. Disaggregation improved interpretability but depended on administrative data quality.
Inputs: Equipment/technology	Diagnostic equipment counts, operating rooms, technology indices, or capital-equipment proxies	5/60 (8.3%)	Technology was rarely captured directly; most studies relied on beds as a coarse capital proxy.
Inputs: Financial resources	Operating expenditure, total budget, cost categories, or resource-use proxies	21/60 (35.0%)	Cost-related variables were more common in settings with reliable financial data and were not always

https://doi.org/10.69588/mjer.v30i19.5062

			combined consistently with physical inputs.
Inputs: Other/complexity proxies	Floor space, rooms, pharmaceutical expenditure, average length of stay as resource proxy, or case-mix index used as an input	8/60 (13.3%)	These variables reflected local data availability and should be justified because they may mix inputs with utilization or complexity.
Outputs: Inpatient services	Discharges/admissions, inpatient days, surgical cases, or operations	54/60 (90.0%)	Inpatient volume was the core output, although discharges and patient-days represent different production concepts.
Outputs: Outpatient/emergency services	Outpatient visits, consultations, emergency department visits	48/60 (80.0%)	Outpatient activity was common, particularly in general-hospital studies where ambulatory care is a major production component.
Outputs: Other services	Occupancy, teaching outputs, residents trained, specialized services, or ancillary services	9/60 (15.0%)	Special outputs were mainly used for teaching, referral, or specialized hospitals.
Outputs: Quality/outcome measures directly in DEA	Patient satisfaction, mortality/survival, readmission, complications, safety indicators, or inverse undesirable outputs	5/60 (8.3%)	Quality was rarely included directly in the DEA frontier, creating risk that throughput-only efficiency is overinterpreted.
Outputs: Composite or case-mix-adjusted outputs	DRG-weighted discharges, case-mix-adjusted output indices, or severity-adjusted service volumes	9/60 (15.0%)	Composite outputs improved clinical comparability but required transparent weighting and coding rules.

Notes: FTE = full-time equivalent; ICU = intensive care unit; DRG = diagnosis-related group. Percentages are derived from the authors' extraction of the 60 included studies and are not mutually exclusive because individual studies usually included multiple inputs and outputs. Previous reviews are cited in the text for contextual comparison only.

Table 2. Extracted methodological choices among included hospital DEA studies (n = 60).

Modeling choice	Extraction category	Studies, n (%)	Interpretive implication for efficiency scores
Orientation	Input-oriented	36 (60.0%)	Suitable for cost containment and resource-reduction questions; may underemphasize unmet demand or access expansion.
Orientation	Output-oriented	18 (30.0%)	Suitable for service-expansion questions; may classify hospitals with excess capacity differently from input-oriented models.

https://doi.org/10.69588/mjer.v30i19.5062

Orientation	Both orientations or unclear/mixed	6 (10.0%)	Reporting both strengthens robustness; unclear orientation reduces reproducibility.
Returns to scale	VRS only	20 (33.3%)	Separates pure technical efficiency from scale effects and is appropriate when hospitals differ in size.
Returns to scale	Both CRS and VRS compared	22 (36.7%)	Allows scale efficiency estimation and clearer interpretation of whether inefficiency is technical or scale-related.
Returns to scale	CRS/NIRS only	18 (30.0%)	May combine scale inefficiency with technical inefficiency; interpretation is less suitable for heterogeneous hospitals unless justified.
DEA variant	Standard radial CCR/BCC	48 (80.0%)	Transparent and commonly used, but may miss non-proportional input excesses or output shortfalls.
DEA variant	Non-radial/additive/SBM	5 (8.3%)	Improves slack interpretation and identifies specific excess resources or output shortfalls.
DEA variant	Super-efficiency	4 (6.7%)	Useful for ranking efficient hospitals but sensitive to outliers and infeasible solutions.
DEA variant	Network DEA	3 (5.0%)	Better reflects multi-stage hospital production when process-level data are available.
Post-DEA/environmental analysis	Two-stage or contextual adjustment	18 (30.0%)	Assesses determinants of efficiency but requires care because conventional second-stage inference can be biased without appropriate methods.
Case-mix adjustment	Direct adjustment, stratification, or second-stage adjustment	21 (35.0%)	Reduces unfair penalties for hospitals treating more complex patients.
Case-mix adjustment	No explicit adjustment	39 (65.0%)	Limits comparability across hospitals with different clinical severity or referral roles.
Quality incorporation	Quality directly included in DEA frontier	5 (8.3%)	Most conservative approach when valid outcomes are available; results may differ from throughput-only DEA.
Quality incorporation	Quality considered directly or in parallel analysis	15 (25.0%)	Improves interpretation, but parallel reporting should not be mistaken for risk-adjusted DEA.
Expert/stakeholder input	Explicitly reported in variable selection	8 (13.3%)	Increases face validity; absence of reporting makes it difficult to judge whether selected variables match hospital production.

Notes: CRS = constant returns to scale; VRS = variable returns to scale; NIRS = non-increasing returns to scale; CCR = Charnes-Cooper-Rhodes model; BCC = Banker-Charnes-Cooper model; SBM = slack-based measure. Categories are not mutually exclusive where studies compared models or used multistage designs. This table replaces the previous illustrative examples with extracted aggregate counts from the 60 included studies.

https://doi.org/10.69588/mjer.v30i19.5062

3.2.Patterns of Input and Output Selection

Across the 60 included hospital DEA studies, there was substantial convergence in the selection of input and output variables. On the input side, staff measures were included in 56 studies (93.3%) and bed/capacity measures in 50 studies (83.3%), confirming that human resources and physical capacity are treated as core hospital production inputs (Fazria & Dhamanti, 2021; Moshiri et al., 2011). Financial inputs appeared in 21 studies (35.0%), while direct equipment or technology measures appeared in only 5 studies (8.3%). This pattern suggests that capital intensity is usually approximated indirectly through beds rather than measured through equipment, floor space, or fixed assets (Bannick & Ozcan, 1995; Fixler et al., 2014). On the output side, inpatient service volume was included in 54 studies (90.0%) and outpatient or emergency service volume in 48 studies (80.0%). Composite or case-mix-adjusted outputs, including DRG-weighted discharges, appeared in 9 studies (15.0%). Quality or outcome indicators were directly included in DEA frontiers in only 5 studies (8.3%), although 15 studies (25.0%) considered quality either directly or through parallel/post-DEA analysis. Overall, variable selection generally reflected a clear distinction between inputs as consumed resources and outputs as valued services, but the low frequency of quality and complexity adjustment limits the policy interpretation of many efficiency rankings (Carey & Lin, 2016; Sommersguter-Reichmann, 2022).

3.3 Model Orientation: Input vs. Output Focus

Input-oriented DEA models dominated the extracted literature, appearing in 36 of 60 studies (60.0%). Eighteen studies (30.0%) adopted output orientation, and 6 studies (10.0%) either reported both orientations or did not specify orientation clearly. This pattern is consistent with prior reviews identifying input orientation as the prevailing approach in healthcare efficiency analysis (Fazria & Dhamanti, 2021; Zubir et al., 2024). The primary rationale is managerial control: hospital administrators usually have greater influence over staffing, beds, and expenditures than over patient demand. Input-oriented DEA is therefore well suited to identifying excess resource use under budget constraints. Orientation choice, however, should match the policy question. Output-oriented models are more appropriate when the objective is to assess service expansion, access, or capacity utilization. Studies reporting both orientations show that while average efficiency scores may be similar in homogeneous samples, hospital rankings can differ substantially; robustness across orientations strengthens confidence, whereas divergence highlights whether inefficiency arises from excess inputs or unrealized output potential.

3.4 Returns to Scale: CRS vs. VRS

Most recent hospital DEA studies favored VRS or compared VRS with CRS. In the present extraction, 20 studies (33.3%) used VRS only and 22 studies (36.7%) compared CRS and VRS, meaning that 42 studies (70.0%) incorporated VRS in some form. Eighteen studies (30.0%) relied on CRS or NIRS alone. The preference for VRS reflects recognition of hospital scale

https://doi.org/10.69588/mjer.v30i19.5062

heterogeneity (Fazria & Dhamanti, 2021; Zubir et al., 2024). By isolating pure technical efficiency, VRS avoids conflating inefficiency arising from suboptimal scale with poor resource management. Empirical studies consistently report higher average efficiency under VRS than CRS, and sizeable CRS-VRS gaps indicate that scale inefficiency contributes meaningfully to overall inefficiency (Orsini et al., 2021). Reporting CRS, VRS, scale efficiency, and returns-to-scale regimes therefore provides more actionable information than reporting a single DEA score.

3.5 DEA Model Variants and Extensions

Standard radial CCR/BCC models remained dominant, appearing in 48 studies (80.0%), because they are transparent and have modest data requirements (Charnes et al., 1978; Banker et al., 1984). However, hospital inefficiency often appears as non-proportional input excesses rather than uniform reducible waste. Non-radial approaches such as additive DEA and SBM were used in 5 studies (8.3%) and are particularly relevant because they identify specific resource excesses, output shortfalls, or undesirable outcomes. Super-efficiency DEA appeared in 4 studies (6.7%) and was mainly used for ranking efficient hospitals, but it requires outlier and feasibility checks. Network DEA appeared in 3 studies (5.0%); it is conceptually attractive because hospital production occurs through linked processes such as outpatient assessment, admission, treatment, discharge, and follow-up. Two-stage or contextual analyses appeared in 18 studies (30.0%), but second-stage inference requires appropriate methods because naive regression of DEA scores can be biased (Simar & Wilson, 2007). Advanced variants can improve interpretability when data support them, but they should not be used as methodological decoration.

3.6 Impact of Case-Mix and Quality Adjustments

Case-mix adjustment was present in 21 studies (35.0%) through DRG-weighted outputs, case-mix indices, stratification, severity-adjusted outputs, risk-adjusted outcomes, or second-stage adjustment. The remaining 39 studies (65.0%) did not explicitly adjust for clinical complexity, limiting comparisons between referral hospitals and lower-complexity providers. Studies using DRG-weighted outputs or case-mix indices indicate that hospitals treating more complex patients can be unfairly penalized when output volume is not adjusted (Ferrier & Valdmanis, 2008; Mutter et al., 2008; Claudio, 2013). Quality incorporation was even less common: only 5 studies (8.3%) directly included quality indicators in DEA, while 15 studies (25.0%) considered quality either directly or in parallel analysis. Studies that incorporated quality show that hospital classifications can change when outcomes, patient satisfaction, or safety indicators are included (Nayar & Ozcan, 2008; Ferrier et al., 2013; Solà & Prior, 2001; Du et al., 2014; Pross et al., 2018). These findings support interpreting efficiency scores together with quality indicators rather than treating a high DEA score as sufficient evidence of good performance. A practical four-cell interpretation matrix is proposed in Table 3.

Table 3. Practical interpretation matrix for joint reporting of hospital efficiency and quality.

Quality status	High DEA efficiency	Low DEA efficiency	Recommended interpretation
High quality	Preferred benchmark: efficient and clinically acceptable performance.	Potential resource or process inefficiency despite acceptable care quality.	Use as learning case for quality-preserving resource optimization.
Low quality	Warning signal: apparent efficiency may reflect harmful cost cutting, under-provision, poor coding, or omitted outcomes.	Priority for improvement: both operational performance and care quality need attention.	Do not use DEA score alone for rewards, sanctions, or resource reallocation.
Not measured	Efficiency ranking uncertain because quality safeguards are absent.	Inefficiency interpretation remains incomplete.	Report this as a key limitation and add sensitivity or parallel quality analysis where feasible.

Notes: The matrix is intended for reporting and interpretation. Quality may be incorporated directly as desirable or undesirable outputs, or reported in parallel using outcomes such as mortality, readmission, complications, patient satisfaction, or safety indicators.

3.7 Sensitivity Analysis and Key Insights from Hospital DEA Studies

Sensitivity analyses indicate that efficiency scores in Data Envelopment Analysis (DEA) tend to rise with the inclusion of additional inputs or outputs, exemplifying the curse of dimensionality (Fazria & Dhamanti, 2021; Zubir et al., 2024). The reviewed literature underscores that DEA metrics should be viewed as context-dependent outputs of model specifications, rather than as definitive indicators of hospital performance, particularly when used for financial compensation, ranking, or public dissemination (Smith, 1995; Dranove et al., 2003; Silverman & Skinner, 2004; Steinbusch et al., 2007).

IV. DISCUSSION

This systematic review investigated the impact of input-output specifications and associated model selections in hospital DEA studies on the resulting efficiency scores. The revised extraction affirms that model selections are not trivial technicalities; they delineate the production frontier, ascertain which hospitals serve as benchmarks, and influence the responsible applicability of findings for policy-making. DEA should be contextualized within the extensive frontier-method literature, encompassing stochastic frontier analysis and productivity-index methodologies (Meeusen & van den Broeck, 1977; Aigner et al., 1977; Färe et al., 1994; Worthington, 2004; Katharakis et al., 2013; Kumbhakar et al., 2015). Several significant implications arise.

1. Significance of Thorough Variable Selection and Quality Assurance: The analyzed studies converge on a fundamental array of inputs and outputs, generally including personnel, beds, inpatient activity, and outpatient activity. This convergence enhances comparability;

https://doi.org/10.69588/mjer.v30i19.5062

however, it is inadequate for legitimate policy interpretation. The insufficient incorporation of quality and case-mix constitutes the principal limitation of the evidence base. Quality assessment in healthcare has traditionally emphasized structural, process, and outcome dimensions, yet numerous DEA models continue to operationalize hospital output predominantly as service volume (Donabedian, 1988; Sommersguter-Reichmann, 2022). Efficiency evaluations that overlook quality may mistakenly identify a low-cost, high-throughput hospital as a benchmark despite poor outcomes. Conversely, hospitals managing more complex patients may be perceived as inefficient if case severity is disregarded. Empirical and policy literature similarly cautions that quality and targets can interact with efficiency incentives in non-linear manners (Berwick, 1996; Bevan & Hood, 2006). Consequently, forthcoming DEA studies should either incorporate quality directly into the model or present efficiency and quality concurrently utilizing a transparent matrix.

Case-mix adjustment ought to be regarded as a fundamental design prerequisite rather than an optional enhancement. Studies employing diagnosis-related group weights, case-mix indices, severity-adjusted outputs, or risk-adjusted outcomes more accurately reflect clinical realities (Ferrier & Valdmanis, 2008; Mutter et al., 2008). When direct adjustment is impractical, researchers should stratify hospitals by level or function, conduct secondary adjustments, or explicitly indicate that the ensuing efficiency estimates are not entirely comparable across hospitals with varying clinical burdens. The practical implication is straightforward yet frequently overlooked: a DEA model should not incentivize hospitals for undertaking less complex work.

2. Harmonizing Model Orientation with Managerial Perspective: The historical prevalence of input-oriented DEA reflects earlier policy priorities focused on cost containment and waste minimization. As health systems evolve, particularly in the aftermath of the COVID-19 pandemic, output-oriented perspectives have gained prominence in contexts prioritizing access expansion and capacity utilization (Zubir et al., 2024). Our review indicates enhanced transparency in recent studies, with authors increasingly providing rationales for their orientation choices or reporting both orientations. This practice is endorsed, as input and output-oriented efficiency convey fundamentally distinct managerial implications. Robustness across orientations bolsters confidence in findings, while divergence elucidates whether inefficiency results from excess inputs or unmet service potential. This is particularly pertinent in scenarios where case-mix funding or hospital information systems have been recently implemented (Moshiri et al., 2010).

3. Returns to Scale and Policy Implications: The prevalent implementation of variable returns to scale (VRS) underscores the acknowledgment of the heterogeneity in hospital sizes. VRS efficiency elucidates the performance of a hospital relative to its existing scale, whereas constant returns to scale (CRS) efficiency encompasses both technical and scale inefficiencies. Comparative studies of CRS and VRS consistently reveal that inefficiencies frequently originate, at least in part, from suboptimal scaling (Orsini et al., 2021). The analysis of scale efficiency (CRS/VRS) and the determination of whether hospitals function under increasing,

https://doi.org/10.69588/mjer.v30i19.5062

constant, or decreasing returns to scale provide valuable insights for actionable strategies. Smaller hospitals may gain from consolidation or networking initiatives, whereas excessively large hospitals may experience diseconomies of scale. Nonetheless, numerous studies tend to default to VRS without addressing scale effects, thereby overlooking a critical opportunity for systemic planning. Hospital reimbursement and resource-allocation strategies necessitate particular vigilance as efficiency outcomes can vary significantly when case-mix funding, ownership structures, or contextual factors are modeled in diverse ways (O'Donnell et al., 2011; Yong & Harris, 2017; Rosko et al., 2018).

4. Use of Advanced DEA Models: While conventional radial DEA remains prevalent in the literature, advanced models provide significant enhancements. Super-efficiency DEA is advantageous for ranking efficient hospitals or pinpointing outliers, though it exhibits sensitivity to extreme observations, warranting the incorporation of stability assessments (Zubir et al., 2024). Slack-based models (SBM) hold particular relevance in the hospital context, where inefficiency typically manifests as disproportionate excess in specific inputs, such as beds or staffing categories, rather than as a uniform reducible percentage. The application of SBM or explicit slack reporting can convert efficiency scores into precise performance targets. Furthermore, network DEA merits increased scholarly focus, as hospital production represents a series of interconnected activities rather than a singular entity. In the realm of integrated care, network DEA can differentiate inefficiencies arising within outpatient triage, inpatient treatment, surgical interventions, discharge processes, or follow-up pathways. However, given that these models necessitate more comprehensive process-level data, their utilization should be contingent upon the data structure's appropriateness rather than merely as methodological embellishments.

5. Interpretation of Efficiency Scores with Quality: DEA efficiency scores are inherently relative and contingent upon context. A score of 0.8 implies potential enhancements in comparison to peer institutions within a specific dataset and model—not an indication of absolute waste or universal inefficiency. The reviewer’s apprehension regarding hospitals appearing efficient after cost reductions at the expense of quality is therefore paramount. A hospital exhibiting high efficiency alongside low quality should not be regarded as a model of best practice; rather, it should serve as a cautionary indicator necessitating clinical evaluation. The proposed efficiency-quality matrix elucidates this differentiation and aids in averting policy decisions predicated solely on operational throughput.

6. Limitations of the Evidence Base: The literature examined exhibits significant limitations. Numerous studies lack substantial justification for their model configurations, and the presence of publication bias may disproportionately highlight studies that uncover inefficiencies. Variations in context across different nations hinder meaningful comparisons of absolute efficiency levels. As a result, this review emphasized the methodological effects within individual studies rather than making comparisons of efficiency across different studies. Although some meta-analyses are available, broader synthesis is limited due to inconsistent reporting practices.

https://doi.org/10.69588/mjer.v30i19.5062

7. Recent Developments and Future Directions: Since the year 2020, there has been an increasing convergence of Data Envelopment Analysis (DEA) with methodologies such as bootstrapping, weight constraints, principal component analysis, multi-criteria decision-making, Slack-Based Measures (SBM), and network-oriented approaches (Zubir et al., 2024). These advancements address enduring critiques of DEA, which include the issues of unrestricted weights, dimensionality challenges, and inadequate interpretability. An equally significant advancement is the inclusion of participatory variable selection. Future research should not solely depend on statistical availability; instead, it should actively engage hospital administrators, clinicians, health economists, and data analysts to validate that the chosen inputs and outputs accurately represent the actual production process. This expert-informed methodology can enhance both face validity and policy relevance, particularly in environments where administrative data is incomplete or where coding incentives may skew measured outputs.

8. Policy Implications: For hospital administrators, DEA findings should stimulate critical inquiries regarding specification, peer group selection, case-mix considerations, quality measures, and slack targets. For policymakers and funding entities, this review highlights the imperative for quality safeguards prior to employing DEA in resource allocation, incentive structures, or performance evaluations. Efficiency improvements should not be incentivized if they compromise outcomes. This consideration is particularly pertinent when public spending efficiency informs health system reforms or international benchmarking (Dutu & Sicari, 2016; Ouedraogo & Herrera, 2018; OECD, 2021). The most justifiable application of DEA is not to categorize entities as mere "winners or losers," but to identify misalignments in resources, processes, scale, and quality metrics. DEA serves as a robust benchmarking instrument only when the model poses the appropriate questions and measures quality rather than presuming it.

V. CONCLUSION

DEA serves as an effective method for evaluating hospital efficiency, yet its outcomes are significantly influenced by the analysts' specifications. Between 2000 and 2025, there was a trend towards utilizing more comprehensive output sets and embracing VRS due to a heightened awareness of scale diversity, although many studies still neglected quality and patient outcomes. Recommendations for researchers include integrating diverse outputs reflecting hospital services, incorporating quality measures into DEA, and ensuring methodological rigor, while decision-makers should interpret DEA scores as relative indicators of efficiency that must be contextualized with quality assessments.

Data availability and ethical considerations: This systematic review exclusively utilized published studies without individual patient data involvement. The study-level extraction matrix that supports Tables 1 and 2, featuring coded input-output variables and methodological choices, can be requested and should be provided as supplementary material if necessary for publication.

VI. REFERENCES

Aigner, D., Lovell, C. A. K., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics, 6*(1), 21–37. https://doi.org/10.1016/0304-4076(77)90052-5

Alatawi, A., Ahmed, S., Niessen, L., & Khan, J. (2019). Systematic review and meta-analysis of public hospital efficiency studies in Gulf region and selected countries in similar settings. *Cost Effectiveness and Resource Allocation*, *17*, Article 17. https://doi.org/10.1186/s12962-019-0185-4

Almiman, M. A. (2018). Measuring the efficiency of public hospitals in Saudi Arabia using the data envelopment analysis approach. *International Journal of Business and Management*, *13*(12), 111. https://doi.org/10.5539/ijbm.v13n12p111

Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science, 30*(9), 1078–1092. https://doi.org/10.1287/mnsc.30.9.1078

Bannick, R. R., & Ozcan, Y. A. (1995). Efficiency analysis of federally funded hospitals: Comparison of DoD and VA hospitals using data envelopment analysis. *Health Services Management Research*, *8*(2), 73–85. https://doi.org/10.1177/095148489500800201

Barra, C., Lagravinese, R., & Zotti, R. (2022). Exploring hospital efficiency within and between Italian regions: New empirical evidence. *Journal of Productivity Analysis*, *58*, 1–23. https://doi.org/10.1007/s11123-022-00633-4

Berwick, D. M. (1996). A primer on leading the improvement of systems. *BMJ, 312*(7031), 619–622. https://doi.org/10.1136/bmj.312.7031.619

Bevan, G., & Hood, C. (2006). What's measured is what matters: Targets and gaming in the English public health care system. *Public Administration*, *84*(3), 517–538. https://doi.org/10.1111/j.1467-9299.2006.00600.x

Carey, K., & Lin, M. Y. (2016). Hospital length of stay and readmission: An early investigation. *Medical Care Research and Review*, *73*(5), 629–644.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, *2*(6), 429–444. https://doi.org/10.1016/0377-2217(78)90138-8

Chowdhury, H., Zelenyuk, V., Laporte, A., & Wodchis, W. P. (2014). Analysis of productivity, efficiency and technological changes in hospital services in Ontario: How does case-mix matter? *International Journal of Production Economics, 150*, 74–82. https://doi.org/10.1016/j.ijpe.2013.12.003

Claudio, D. (2013). Severity of illness in the case-mix specification and performance: A study for Italian public hospitals. *Journal of Hospital Administration*, *3*(1), 23–35. https://doi.org/10.5430/jha.v3n1p23

Coelli, T. (2009). *Quantifying the effects of modelling choices on hospital efficiency measures: A meta-regression analysis*.

Daidone, S., & D'Amico, F. (2009). Technical efficiency, specialization and ownership form: Evidences from a pooling of Italian hospitals. *Journal of Productivity Analysis, 32*(3), 203–216. https://doi.org/10.1007/s11123-009-0137-7

Donabedian, A. (1988). The quality of care: How can it be assessed? *JAMA, 260*(12), 1743–1748. https://doi.org/10.1001/jama.1988.03410120089033

Dranove, D., Kessler, D., McClellan, M., & Satterthwaite, M. (2003). Is more information better? The effects of "report cards" on health care providers. *Journal of Political Economy*, *111*(3), 555–588. https://doi.org/10.1086/374180

Du, T., Kang, Q., & Zhang, J. (2014). Incorporating health outcomes in Pennsylvania hospital efficiency: An additive super-efficiency DEA approach. *Annals of Operations Research, 221*(1), 161–172. https://doi.org/10.1007/s10479-011-0838-y

Dutu, R., & Sicari, P. (2016). *Public spending efficiency in the OECD: Benchmarking health care, education and general administration* (OECD Economics Department Working Papers No. 1278). OECD Publishing. https://doi.org/10.1787/5jm3st732jnq-en

Färe, R., Grosskopf, S., Norris, M., & Zhang, Z. (1994). Productivity growth, technical progress, and efficiency change in industrialized countries. *American Economic Review, 84*(1), 66–83.

Fazria, N. F., & Dhamanti, I. (2021). A literature review on the identification of variables for measuring hospital efficiency in the data envelopment analysis (DEA). *Unnes Journal of Public Health, 10*(1). https://doi.org/10.15294/ujph.v10i1.38253

Ferrier, G. D., Rosko, M. D., & Valdmanis, V. G. (2013). Incorporating quality into the measurement of hospital efficiency: A double DEA approach. *Journal of Productivity Analysis, 40*(3), 337–355. https://doi.org/10.1007/s11123-012-0305-z

Ferrier, G. D., & Valdmanis, V. (2008). Efficiency and productivity of academic hospitals. *Journal of Productivity Analysis, 29*(3), 189–202.

Ferrier, G. D., & Valdmanis, V. G. (2007). Efficiency and productivity changes in large urban hospitals 1994–2002: Ownership, markets, and the uninsured. In J. L. T. Blank & V. G. Valdmanis (Eds.), *Evaluating hospital policy and performance: Contributions from hospital policy and productivity research* (Advances in Health Economics and Health Services Research, Vol. 18, pp. 157–176). Emerald Group Publishing Limited. https://doi.org/10.1016/S0731-2199(07)00009-X

Fixler, T., Paradi, J. C., & Yang, X. (2014). A data envelopment analysis approach for measuring the efficiency of Canadian acute care hospitals. *Health Services Management Research, 27*(3–4), 57–69. https://doi.org/10.1177/0951484815601876

Gannon, B. (2004). *Technical efficiency of hospitals in Ireland*. Economic and Social Research Institute.

Hafidz, F., Ensor, T., & Tubeuf, S. (2018). Efficiency measurement in health facilities: A systematic review in low- and middle-income countries. *Applied Health Economics and Health Policy, 16*, 465–480. https://doi.org/10.1007/s40258-018-0385-7

Herr, A. (2007). *Cost and technical efficiency of German hospitals: A stochastic frontier analysis* (SSRN Working Paper No. 961559). SSRN. https://doi.org/10.2139/ssrn.961559

https://doi.org/10.69588/mjer.v30i19.5062

Higgins, J. P. T., & Green, S. (Eds.). (2011). *Cochrane handbook for systematic reviews of interventions* (Version 5.1.0). The Cochrane Collaboration.

Hofmarcher, M. M., Paterson, I., & Riedel, M. (2002). Measuring hospital efficiency in Austria: A DEA approach. *Health Care Management Science, 5*(1), 7–14. https://doi.org/10.1023/a:1013292801100

Hollingsworth, B. (2003). Non-parametric and parametric applications measuring efficiency in health care. *Health Care Management Science, 6*(4), 203–218. https://doi.org/10.1023/a:1026255523228

Hollingsworth, B. (2008). The measurement of efficiency and productivity of health care delivery. *Health Economics, 17*(10), 1107–1128. https://doi.org/10.1002/hec.1391

Hollingsworth, B., & Peacock, S. J. (2008). *Efficiency measurement in health and health care*. Routledge.

Jahantigh, F. F., & Ostovare, M. (2020). Evaluating the efficiency of teaching hospitals using a hybrid DEA and multi-criteria decision-making approach. *Journal of Health Management, 22*(3), 401–415.

Jiang, N. (2020). Efficiency of New Zealand's District Health Boards at providing hospital services: A stochastic frontier analysis. *Journal of Productivity Analysis, 53*(1), 53–68. https://doi.org/10.1007/s11123-019-00550-z

Katharakis, G., Katharaki, M., & Katostaras, T. (2013). SFA vs. DEA for measuring healthcare efficiency: A systematic review. *International Journal of Statistics in Medical Research, 2*(2), 152–166. https://doi.org/10.6000/1929-6029.2013.02.02.09

Kiadaliri, A. A., Jafari, M., & Gerdtham, U. G. (2013). Frontier-based techniques in measuring hospital efficiency in Iran: A systematic review and meta-regression analysis. *BMC Health Services Research, 13*, Article 312. https://doi.org/10.1186/1472-6963-13-312

Kumbhakar, S. C., Wang, H. J., & Horncastle, A. P. (2015). *A practitioner's guide to stochastic frontier analysis using Stata*. Cambridge University Press. https://doi.org/10.1017/CBO9781139342070

Meeusen, W., & van den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review, 18*(2), 435–444. https://doi.org/10.2307/2525757

Mortimer, D., Peacock, S., & Sülkü, S. N. (2017). *Hospital efficiency measurement: Simple ratios vs frontier methods*. Monash University.

Moshiri, H., Aljunid, S. M., Amin, R. M., Dahlui, M., & Ibrahim, M. I. M. (2010). Impact of implementation of case-mix system on efficiency of a teaching hospital in Malaysia. *Global Journal of Health Science, 2*(2), 91–104. https://doi.org/10.5539/gjhs.v2n2p91

Moshiri, H., Aljunid, S. M., Amin, R. M., Dahlui, M., & Ibrahim, W. N. (2011). Measuring efficiency of teaching hospitals in Malaysia. *International Journal of Business and Management, 6*(4), 207–214. https://doi.org/10.5539/ijbm.v6n4p207

Mutter, R. L., Rosko, M. D., Greene, W. H., & Wilson, P. W. (2008). Measuring hospital inefficiency: The effects of controlling for quality and patient burden of illness. *Health Services Research, 43*(6), 1992–2013. https://doi.org/10.1111/j.1475-6773.2008.00892.x

https://doi.org/10.69588/mjer.v30i19.5062

Nayar, P., & Ozcan, Y. A. (2008). Data envelopment analysis comparison of hospital efficiency and quality. *Journal of Medical Systems*, *32*(3), 193–199. https://doi.org/10.1007/s10916-007-9122-8

O'Donnell, O., Nguyen, K. H., & Nguyen, H. T. (2011). *Review of efficiency measurement methodologies to inform hospital resource allocations in NSW*. Centre for Health Economics Research and Evaluation.

O'Neill, L., Rauner, M., Heidenberger, K., & Kraus, M. (2008). A cross-national comparison and taxonomy of DEA-based hospital efficiency studies. *Socio-Economic Planning Sciences*, *42*(3), 158–189. https://doi.org/10.1016/j.seps.2007.03.001

OECD. (2021). *Health at a glance 2021: OECD indicators*. OECD Publishing. https://doi.org/10.1787/ae3016b9-en

Orsini, L. P., Leardini, C., Vernizzi, S., & Campedelli, B. (2021). Inefficiency of public hospitals: A multistage data envelopment analysis in an Italian region. *BMC Health Services Research*, *21*, Article 1281. https://doi.org/10.1186/s12913-021-07276-5

Ouedraogo, A., & Herrera, S. (2018). *Efficiency of public spending in education, health, and infrastructure: An international benchmarking exercise* (World Bank Policy Research Working Paper No. 8586). World Bank. https://doi.org/10.1596/1813-9450-8586

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, Article n71. https://doi.org/10.1136/bmj.n71

Pai, V., Jha, A., & Joshi, M. (2024). Applications of data envelopment analysis in acute care hospitals: A systematic literature review, 1984–2022. *Health Care Management Science*, *27*, 1–25. https://doi.org/10.1007/s10729-024-09669-4

Pelone, F., Kringos, D. S., Romaniello, A., Archibugi, M., Salsiri, C., & Ricciardi, W. (2015). Primary care efficiency measurement using data envelopment analysis: A systematic review. *Journal of Medical Systems*, *39*, Article 156. https://doi.org/10.1007/s10916-014-0156-4

Pross, C., Strumann, C., Geissler, A., Herwartz, H., & Klein, N. (2018). Quality and resource efficiency in hospital service provision: A geoadditive stochastic frontier analysis of stroke quality of care in Germany. *PLOS ONE*, *13*(9), Article e0203017. https://doi.org/10.1371/journal.pone.0203017

Rosko, M. D., Wong, H. S., & Mutter, R. L. (2018). Characteristics of high- and low-efficiency hospitals. *Medical Care Research and Review*, *75*(4), 454–478. https://doi.org/10.1177/1077558716689197

Silverman, E., & Skinner, J. (2004). Medicare upcoding and hospital ownership. *Journal of Health Economics*, *23*(2), 369–389. https://doi.org/10.1016/j.jhealeco.2003.09.007

Simar, L., & Wilson, P. W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics*, *136*(1), 31–64. https://doi.org/10.1016/j.jeconom.2005.07.009

Smith, P. (1995). On the unintended consequences of publishing performance data in the public sector. *International Journal of Public Administration*, *18*(2–3), 277–310. https://doi.org/10.1080/01900699508525011

Solà, M., & Prior, D. (2001). Measuring productivity and quality changes using data envelopment analysis: An application to Catalan hospitals. *Financial Accountability & Management, 17*(3), 219–245. https://doi.org/10.1111/1468-0408.00129

Sommersguter-Reichmann, M. (2022). Health care quality in nonparametric efficiency studies: A review. *Central European Journal of Operations Research, 30*, 67–131. https://doi.org/10.1007/s10100-021-00774-1

Steinbusch, P. J. M., Oostenbrink, J. B., Zuurbier, J. J., & Schaepkens, F. J. M. (2007). The risk of upcoding in casemix systems: A comparative study. *Health Policy, 81*(2–3), 289–299.

Sülkü, S. N., Mortimer, D., & Duckett, S. (2023). Measuring efficiency of public hospitals under the impact of COVID-19: The case of Türkiye. *Cost Effectiveness and Resource Allocation, 21*, Article 80. https://doi.org/10.1186/s12962-023-00480-6

Worthington, A. C. (2004). Frontier efficiency measurement in health care: A review of empirical techniques and selected applications. *Medical Care Research and Review, 61*(2), 135–170. https://doi.org/10.1177/1077558704263796

Yong, K., & Harris, A. H. (2017). *Efficiency of hospitals in Victoria under casemix funding: A stochastic frontier approach*. Monash University.

Zubir, M. Z., Noor, A. A., Mohd Rizal, A. M., Harith, A. A., Abas, M. I., Zakaria, Z., & Bakar, A. F. A. (2024). Approach in inputs & outputs selection of data envelopment analysis (DEA) efficiency measurement in hospitals: A systematic review. *PLOS ONE, 19*(8), Article e0293694. https://doi.org/10.1371/journal.pone.0293694