Peer Review File

Article information: https://dx.doi.org/10.21037/jhmhp-24-10

Reviewer A Comments

1. Page 3. Line 99-100. "…many value-based care (VBC) programs have been implemented nationwide..." It's correct there are several VBCs. Citing a few examples of its successful implementation would be helpful to readers.

Reply 1: Thank you. We added examples of VBC programs and included one example of Hospital Readmission Reduction Program to show the implement success.

Changes in the text:
"To tie the quality of care delivered to payments for care delivery, multiple value-based care (VBC) programs have been implemented nationwide (2, 3), such as the Hospital Value-Based Purchasing Program, Hospital Readmission Reduction Program, and the Hospital Acquired Conditions Reduction Program Studies have shown that VBC adoption was associated with improved quality of care. For instance, a study examining 3,387 United States (U.S.) hospitals from 2007 to 2015 revealed that the Hospital Readmission Reduction Program decreased readmission rates as a result of program incentives (4). Furthermore, Pandey and colleagues conducted a systematic review and concluded that higher-intensity Value-Based Purchasing Programs were more consistently associated with improvements in desired quality processes, utilization measures, and reductions in spending, compared to lower-intensity programs (5)." (see Page 3, line 61-70).

2. Page 3. Line 105. "…physician-hospital integration…" What is PHI? How about a brief description of the concept for lay readers? Why is it beneficial?

Reply 2: Thank you. We added the definition of "PHI" that included the benefits for implementing PHI. We then reorganized the sentences.

Changes in the text:
"However, efficacious adoption of these strategies requires a mechanism that aligns staff interests and behaviors with the goals of the hospitals. The need to become more efficient has led hospital executives to physician-hospital integration (PHI), which encompasses a closer collaboration between physicians and hospitals through various contractual forms, including employment contracts, joint ventures, and other partnerships (6-8)." (see Page 3, line 73-77).

3. Page 4. Line 142-144. "respectively: Integrated Salary Model (ISM), Independent Practice Association (IPA), and Open & Closed Physician-Hospital Organization (Open-PHO and Closed PHO)…" What do these terms mean? Are they standard terms in the industry?

Reply 3: Thank you. We added a new table (Table 1) to show the American Hospital Association official definitions of each PHI model.

Changes in the text:

In the Table file, we added a new Table 1(see Page 1). In accord with this change, we add the "(the definitions of each PHI model are shown in Table 1)" and add "Insert Table 1 about here" in the manuscript (see Page 6, line 122-124).

4. Page 4. Line 140-141. "measure hospital efficiency at both the hospital level and system level…" Though the authors state this as one of the objectives of the study, there is hardly any discussion on the system and its linkages with X-inefficiency and agency theory, as well as its linkages with financial and clinical performance. In fact, even though the hypotheses mention "system", they are not motivated by system-level characteristics.

Reply 4: Thank you for your insightful comments and for seeking further clarification on our study's objectives.

In our study, guided by data from the AHA survey and illustrative examples, such as from Houston Methodist Hospital, PHI can be implemented either at the individual hospital level or healthcare system level. Our primary focus was to measure the individual hospital efficiency (outcome) difference between individual hospital PHI strategy and healthcare system level PHI strategy.

We recognize your observation regarding the fact that we did not include system-level characteristics. The decision to not delve deeply into system-level characteristics was intentional, aimed at honing our analysis on the nuances of individual hospital efficiency. This approach was predicated on the understanding that while PHI can manifest at both levels, the implications for X-inefficiency and agency theory are directly observable and relevant at the individual hospital level.

In light of your guidance, we rewrote the sentences in order to avoid any confusion.

Changes in the text:
We rewrote the sentence "Our objectives were to 1) measure hospital efficiency of both hospital-level and system-level with three PHI types: financial PHI (represented by Integrated Salary Model (ISM)), clinical PHI (represented by Independent Practice Association (IPA)), and financial and clinical hybrid PHI (represented by Open and Closed Physician-Hospital Organization (Open-PHO and Closed PHO)) (the definitions of each PHI model are shown in Table 1), respectively;" (see Page 5, line 112-117).

5. Page 4. Line 151-153. "X-inefficiency is concerned with situations when an organization fails to fully utilize its resources to achieve the maximum possible output level, that is, the efficiency frontier (26)." This statement is partially correct. X-inefficiency could be for many other reasons, including employee motivation, competition, supply sources, lack of management control, regulatory environment, etc.

Reply 5: Thank you for the guidance. We rewrote the sentence.

Changes in the text:

We rewrote the sentence “These inefficiencies can arise from numerous sources, such as outdated technology, inefficient production processes, suboptimal management practices, and insufficient competitive pressures, among others. They also manifest when employee behavior deviates from optimal performance, driven by a diminished motivation to pursue efficiency.” (see Page 6, line 130-133).

6. Page 5. Line 213-214. “….1,182 government-owned hospitals, 83 children’s hospitals, and 814 critical access hospitals were excluded from this study.” Why were these hospitals excluded from the study, especially the critical access hospitals (CAH)? CAHs are acute care hospitals, and most are not government owned.

Reply 6: Thank you for requesting clarification. Our initial intention was indeed to include Critical Access Hospitals (CAHs) in our analysis. However, we encountered a significant limitation: the CMS dataset does not provide data on the Central Line-Associated Bloodstream Infection (CLABSI) scores for CAHs. Additionally, CAHs are subject to specific bed number restrictions—according to policy, they typically operate with 15 or fewer beds for acute care inpatient services at any given time, with the allowance for up to 10 additional swing beds. This results in a maximum capacity of 25 inpatient beds, in stark contrast to the larger bed sizes found in acute care hospitals. This discrepancy in bed capacity could potentially skew comparisons between the output and input measures of these two hospital types, making any direct comparison challenging and potentially misleading.

Given these considerations and to ensure the integrity and comparability of our analysis, we made the decision to exclude CAHs from our study.

7. Page 6. Line 224. “…CLBASI…” What is this term? Why not spell it out the first time the term is used in the manuscript. Same with ICU and AMI and several others.

Reply 7: Thank you. We went back to spelling out all the acronyms when mentioned them at the first time.

Changes in the text:
We added “data envelopment analysis (DEA)” (see Page 5, line 109); added Acute Myocardial Infarction (AMI) (see Page 9, line 202); added “central line-associated bloodstream infection (CLBASI) (see Page 9, line 202-203); Intensive care units (ICUs) (see Page 10, line 211-212); Inpatient Prospective Payment Systems (IPPS) (see Page 11, line 247);

8. Data and results sections. The data for the study pertains to 2020, the year of COVID pandemic. This was an unprecedented year of turmoil in the healthcare sector, especially hospitals, which may have affected the data collection and quality. For instance, many hospitals postponed elective procedures for months and were treating COVID-affected patients and patients with COVID symptoms. These may also have impacted results. The authors have not discussed this aspect of the study anywhere in the manuscript.

Reply 8: Thank you for the insightful suggestion. We added the data year limitation to our limitation section.

Changes in the text:
We added “Third, our dataset corresponds to 2020, a year marked by the COVID-19 pandemic, leading to significant disruptions in the healthcare industry, particularly within hospitals. This disruption could have influenced both the gathering and the integrity of the data that might have impacted the generalizability of our findings.” (see Page 20, line 447-450).

Reviewer B Comments
1. The authors should know they cannot use efficiency, effectiveness, and efficacy synonymously (Introduction, lines 94-106) as they have different meanings. I suppose the abbreviation RMU’s for Decision Making Unit is a typo.

Reply 1: Thank you for your valuable guidance. Upon a thorough review of the entire manuscript, we verified that the term “effectiveness” was not employed within our document (except references title). However, we did reference phrases like “effective strategy,” which cannot be directly substituted with the concept of efficiency. In addition, the word “efficacious” appears on page 3, line 73, denoting the strategy's efficacy, which similarly, cannot be interchangeably used with efficiency.

Thank you for pointing out the typo. We have made the necessary correction to the acronym for the “Decision Making Unit.” Your attention to detail is greatly appreciated.

Changes in the text:
We corrected two typos for “Decision-Making Unit’s (DMU’s)” (see Page 13, line 288-289).

2. The authors lack providing sound justification for the DEA model:
a. Have the authors checked the effects of the way they transform the bad outputs? This would be interesting to know because the type of transformation can impact the results.
b. Why did the authors not use a sophisticated approach regarding the relationship between efficiency and quality (Daraio C, Simar L (2007) Conditional nonparametric frontier models for convex and nonconvex technologies: a unifying approach, J Prod Anal 28, 13–32)?
c. Why did the authors not use bootstrapping (to provide feasible means for inference in the second stage, in line with Simar L, Wilson P (2007) Estimation and inference in two-stage, semi-parametric models of production processes, J Econom 136, 31–64)?
Overall, using a cross-sectional data set (without bootstrapping), I think the authors run the risk of presenting snapshots rather than reliable associations.

Reply 2: We are genuinely grateful for your insightful feedback and the opportunity to clarify our methodological choices regarding the DEA model. Your questions allow us to better articulate the rationale behind our approach and address the specific concerns raised.

In response to your broader inquiry, the DEA model, despite its classical stature, remains a robust tool for efficiency analysis, as evidenced by its continued application in contemporary research, including many studies published in 2024. Our choice to use an output-oriented DEA approach was driven by its relevance to our research goals, emphasizing the maximization of outputs with fixed inputs, which aligns with the objectives of many sectors we aim to impact.

a. Regarding the transformation of bad outputs, we acknowledge the importance of this process and its potential impact on the results. While our manuscript may not have explicitly detailed our approach to handling bad outputs, we carefully considered various transformation methods. We aimed to select a method that would most accurately reflect the nature of the data and the specific context of our analysis. We understand the significance of this decision and are open to further examining our approach to ensure the robustness and integrity of our results.

b. On the choice of not employing the sophisticated approach to efficiency and quality as suggested by Daraio and Simar (2007), our decision was guided by the specific objectives of our study and the nature of our data. While we recognize the value of conditional nonparametric frontier models, our preliminary assessment suggested that the traditional DEA model suited our immediate analytical needs. However, we appreciate the recommendation and acknowledge that exploring such advanced methodologies could enrich our analysis. We are considering this for future research to deepen our understanding of the efficiency-quality relationship in our domain.

c. We have implemented the Simar and Wilson two-stage DEA analysis method, incorporating bootstrapping techniques to ensure more reliable inference. This methodological refinement, directly responding to your feedback, allows us to address the potential limitations associated with using a cross-sectional data set without bootstrapping. By doing so, we aim to mitigate the risk of presenting mere snapshots of data and instead provide more substantiated and reliable associations. We are grateful for the opportunity to enhance our analysis based on your recommendations and believe that these methodological adjustments will contribute to a more robust and reliable interpretation of our results.

Changes in the text:
We updated the Appendix (P1-P7) with the two stage semi-parametric modeling analysis description and results.

3. The discussion of the results is bumpy and should be revised (cf. '... for-profit hospitals had a 3.3% higher DEA efficiency score (high efficiency score indicates inefficiency)...' lines 374-377); possibly by transforming the output-oriented efficiency so that higher values indicate higher efficiency.

Reply 3: Thanks for your valuable feedback regarding the clarity of our discussion of the results, particularly the interpretation of the DEA efficiency scores. We understand the concern that the traditional output-oriented DEA model's efficiency scores can be counterintuitive, as higher scores, in fact, denote a greater distance from optimal efficiency.

To address this, we have revised our discussion to more clearly articulate that within the context of DEA output-oriented models, a score of 1 represents optimal efficiency. Scores above 1, therefore, indicate a potential for improvement or a deviation from this optimal efficiency. Specifically, the observed 3.3% higher DEA efficiency score for for-profit hospitals, compared to non-profit hospitals, signals those for-profit hospitals, on average, exhibit a greater margin for enhancing their operational efficiency.

We believe this clarification will make the interpretation of the efficiency scores more intuitive for readers and align with the expectation that higher values typically signify better performance. By explicitly stating that higher scores indicate a greater potential for improvement, we aim to smooth the discussion of the results and enhance the manuscript's readability.

Reviewer C Comments

1. Was a time lag taken into consideration as to start date of the PHI and measured outcomes?

Reply 1: Thank you for highlighting the importance of considering time lags between the start date of PHI and the measured outcomes in our analysis. In our study, we focused on a cross-sectional analysis due to the constraints posed by the availability of longitudinal data. This limitation indeed restricted our ability to directly assess the time lag effects on the outcomes associated with PHI. Recognizing the significance of this aspect, we will consider exploring methodologies that could potentially accommodate or adjust for such effects within the scope of cross-sectional data in future research. We value your feedback as it provides a crucial perspective for enhancing the rigor and depth of our study.

2. Why was system-level considered when PHI contracting and administration is done at the local hospital level?

Reply 2: Thank you for seeking further clarification on our decision to conduct our analysis at the system level. Our rationale for this approach is grounded in insights from the AHA Survey, which indicates that PHI can be strategized and implemented at either the individual hospital or healthcare system level. Recognizing the limited research on the comparative outcomes of PHI across these different levels, our study aims to address this gap.

Our other ongoing research on PHI has revealed distinct patterns in outcomes between hospital-level and system-level PHI. This observation led us to treat PHI not merely as an operational tactic but as a strategic initiative, with the anticipation that its implementation in various contexts yields divergent outcomes. In order to better justify the rationality, we added justification.

Changes in the text:
We added “Additionally, Li (29) found hospital-level PHI and system-level PHI demonstrated distinctly different hospital performance outcomes. These differences underscore the importance of examining the relationships between PHI and hospital efficiency in different organizational contexts. We thus examine hospital efficiency from both hospital level PHI and system level PHI” (see Page 5, line 102-106).

3. Agency theory – aren’t physician and hospital administrators goals the same – quality care? As such, not convinced this theory applies.

Reply 3: Thank you for requesting clarification on our choice of the conceptual framework.
Agency theory, which explores the alignment of interests between physicians and hospitals, has been widely applied (probably the most important one) in research on PHI. Studies shows the goals between physician and hospitals are not always aligned. This theory provides a foundational basis for understanding how strategic alignments can be achieved between these

two entities. PHI is often strategized as a means to synchronize the objectives of physicians and hospitals, aiming to enhance collaboration and improve overall healthcare delivery. Given its extensive use and relevance in the domain of PHI, we consider agency theory as an appropriate conceptual framework for our study.

4. Since the category financial PHI is used, why wasn't operating or total margin used?

Reply 4: We appreciate the opportunity to clarify our approach and the economic principles that guide our analysis.

In our examination of "efficiency," we adopt a perspective rooted in economic science, which posits that financial performance is not a measure of efficiency. One can be doing very well financially without being efficient from either an allocative efficiency perspective or technical/production efficiency perspective. Both are required for Pareto Optimality (an allocation of resources in which it is impossible to improve the level of welfare of one person without hurting the welfare of another person).

5. I do not understand the relevance of the narratives in lines 177-182 and lines 183-186 – both appear to be out of place.

Reply 5: Thank you for highlighting this concern regarding the narratives.

In response to your feedback, we have taken steps to clarify their relevance and connection to our study's overarching themes. Specifically, we introduced subheadings within the section to delineate more clearly between the conceptual framework, focusing on agency and X-inefficiency, and the development of our hypotheses.

Changes in the text:
We revised "Among the three PHI types (the specific PHI models were defined in Table 1)..." (see Page 7, line 155).

6. I'm not convinced that an Open PHO can be compared to a Closed PHO regarding efficiency – each operate differently regarding management, control, and oversight.

Reply 6: Thank you for seeking further clarification on this matter.

We recognize the distinctions between Open PHOs and Closed PHOs, particularly in aspects of management and control. In our study, we have categorized both models under the umbrella of "financial and clinical hybrid PHI" due to their shared characteristics, such as the formation of joint ventures, among others. This classification aligns with the viewpoint of Alexander et al. (1996), who posited that the differences between Open-PHOs and Closed-PHOs may be more semantic than substantive.

Methodologically, while we grouped these models together for theoretical discussion, we have ensured that both are represented in our data analysis. This approach allows us to capture any

nuanced differences in outcomes that may arise from the operational variances between Open and Closed PHOs.

To clarify this point, we added sentences for explaining our rationality.

Changes in the text:
We added “It’s important to acknowledge the distinctions between Open PHOs and Closed PHOs, especially in terms of management and control. However, as noted by Alexander et al., the differences between Open-PHOs and Closed-PHOs might predominantly be semantic rather than substantive (43). We thus categorize both Open-PHOs and Closed-PHOs into financial and clinical hybrid PHI.” (see Page 8, line 171-175).

7. Were specialty hospitals excluded from your study sample? It is unclear how the study sample went from 6,156 to 434. Please advise.

Reply 7: Thank you for requesting clarification. We added detailed description.

Changes in the text:
We rewrote “This study focused on short-term general acute care hospitals in the U.S. There were 6,156 hospitals that participated in the 2020 AHA Survey. From this initial group, 257 hospitals were removed due to missing Medicare provider numbers. Additionally, 1,887 hospitals were excluded for lacking data on either IPA, ISM, Open-PHO, or Closed-PHO at the hospital or system level. Furthermore, 1,998 hospitals were omitted because they had not implemented any of the aforementioned models at either level. After merging data from the AHA Survey with the 2020 CMS Hospital Compare dataset (specifically, the Hospital General Information dataset), several other exclusions were made: 455 government-owned hospitals, 74 hospitals not included in the merged dataset, and 389 non-acute care hospitals (comprising 31 children’s hospitals, 298 critical access hospitals, and 60 psychiatric hospitals). Additionally, 422 hospitals were excluded for not being classified as “short-term hospitals” in the Hospital Cost Report of CMS Hospital Compare dataset. Finally, exclusions were made for hospitals missing specific outcome data: 165 for Acute Myocardial Infarction (AMI) Mortality, 38 for AMI readmission, and 37 for central line-associated bloodstream infection (CLABSI) data. Consequently, the final sample comprised 434 non-government-owned short-term general acute care hospitals.” (see Page 9, line 190-204).

8. I’m not convinced that facility admissions and surgical operations are “good” output indicators if the hospital is a safety-net or if high percent of revenues are paid under capitation arrangements or a value-based plans.

Reply 8: Thank you for this insightful perspective regarding our selection of output indicators.

We understand your concerns about the appropriateness of these metrics given the unique operational and financial frameworks within which these hospitals operate. Our rationale for selecting these indicators is based on their broad applicability and established precedence in evaluating hospital performance across a diverse range of healthcare settings. These measures provide a generalized view of hospital activity and are commonly used benchmarks in healthcare research, offering a valuable perspective on the volume of care provided.

Changes in the text:
We added "..,aligning with the previous PHI efficiency research (18)." " (see Page 9, line 209-210); We added "Furthermore, consistent with prior studies, we selected facility admissions, along with inpatient and outpatient surgical procedures, as the desired outputs for our analysis. However, it is important to note a potential limitation: in hospitals that serve as safety nets or where a significant portion of revenue comes from capitation arrangements or value-based plans, these outputs might not be as indicative of desired outcomes." (see Page 20-21, line 453-457).

9. Suggest using "undesirable" output indicators versus "bad"

Reply 9: Thank you for the suggestion. We changed the term of "good" to "desirable" and "bad" to "undesirable."

Changes in the text:
We changed "good" to "desirable" (see Page 1, line 13); changed "bad" to "undesirable" (see Page 1, line 13); changed "good" to "desirable" (see Page 1, line 13); changed "good" to "desirable" (see Page 4, line 81); changed "bad" to "undesirable" (see Page 9, line 200); changed "good" to "desirable" (see Page 9, line 201); changed "bad" to "undesirable." (see Page 9, line 204); changed "bad" to "undesirable" (see Page 13, line 281); changed "good" to "desirable" (see Page 13, line 282); changed "bad" to "undesirable" (see Page 14, line 307); changed "good" to "desirable" (see Page 14, line 307); changed "bad" to "undesirable" (see Page 14, line 308); and changed "good" to "desirable" (see Page 14, line 311).

10. Why were CLBASI and AMI used versus other quality indicators?

Reply 10: Thank you for asking for clarification. We added the justification for using CLBASI and AMI as care quality indicators.

Changes in the text:
We added "The selection of CLABSI score, along with AMI mortality and readmission rates, as indicators of care quality was made to maintain consistency with prior research on PHI (11, 20, 46, 47). This alignment ensures that our study builds on the established body of knowledge and facilitates comparability with existing findings in the field." (see Page 10, line 226-229).

11. Should case mix index be considered a control variable versus an input measure?

Reply 11: We included case mix index as an input measure for DEA to align with the previous research.

Changes in the text:
We added "To keep consistency with prior studies on PHI and its impact on efficiency, we have incorporated the case mix index into the input measures (18).." (see Page 11, line 242-244).
12. Number of beds – does this mean total or staffed?

Reply 12: Thank you for asking for clarification. We added the details of the number of beds.

Changes in the text:
We added “Number of beds refers to the facility beds set up and staffed at the end of the reporting period (2020)...” (see Page 11, line 248-249).

13. Market characteristics – why not controlling for location differences such as population demographics, wealth of community, education levels (all influence quality and financial outcomes

Reply 13: Thank you for the thoughtful feedback regarding the inclusion of market characteristics such as population demographics, community wealth, and education levels in our analysis. We understand the significance of these factors in influencing quality and financial outcomes within healthcare settings.

Our decision not to incorporate these specific market characteristics was guided by a focused research objective and the scope defined for this particular study. This approach was chosen to isolate the effects of PHI from broader external factors, allowing us to attribute observed outcomes more precisely to PHI practices.

14. I do not understand the relevance of the attached Appendix.

Reply 14: Thank you for your feedback regarding the relevance of the attached Appendix. We initially included an examination of telehealth adoption within our sensitivity analysis, employing both DEA and multiple linear regression models. This focus was driven by the significant increase in telehealth usage during the COVID-19 pandemic, underscoring its potential as a pivotal component of the healthcare system in 2020, the period our study examines. However, upon reflecting on your comments and conducting a thorough comparison between our main results and the sensitivity analysis, we acknowledged the necessity to refine our approach. Consequently, we have opted to omit these analyses from our study. This decision was made to ensure the coherence and focus of our research findings. We appreciate your insights, as they have been instrumental in guiding our revisions for greater clarity and relevance.
d15. Various typos throughout paper – needs proofreading (example HASF was used when referring to the Health Service Areas Files) Thank you. We fixed the typos.

Changes in the text:
We fixed the typo of “HSAF” (see Page 5, line 111); fixed the typo of “HAS” (see Page13, line 278; line 280; line 282).