Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

An Efficient Frontier Approach to Scoring and Ranking Hospital Performance

Daniel Adelman

To cite this article:
Daniel Adelman (2020) An Efficient Frontier Approach to Scoring and Ranking Hospital Performance. Operations Research 68(3):762-792. https://doi.org/10.1287/opre.2019.1972

informs
http://pubsonline.informs.org/journal/opre

OPERATIONS RESEARCH
Vol. 68, No. 3, May–June 2020, pp. 762–792
ISSN 0030-364X (print), ISSN 1526-5463 (online)

Crosscutting Areas

An Efficient Frontier Approach to Scoring and Ranking Hospital Performance

Daniel Adelman[a]

[a] Booth School of Business, The University of Chicago, Chicago, Illinois 60637

Contact: dan.adelman@chicagobooth.edu, https://orcid.org/0000-0001-5534-4511 (DA)

Received: March 5, 2019
Revised: July 18, 2019; November 21, 2019
Accepted: November 24, 2019
Published Online in Articles in Advance: March 17, 2020

Subject Classifications: healthcare: hospitals; economics: input-output analysis; programming: linear, applications
Area of Review: Policy Modeling and Public Sector OR

https://doi.org/10.1287/opre.2019.1972

Abstract. The Centers for Medicare and Medicaid Services (CMS) star rating methodology for publicly evaluating hospitals uses a latent variable model that is based on the presumption of a single, but unobservable, hospital-specific quality factor shared across a group of performance measures. Performance measures are given higher weight if they statistically appear to be more strongly correlated with this hidden factor. We show how this approach, when applied to measures that are weakly or not correlated with each other, can effectively ignore measures and can exhibit "knife-edge" instability, so that even if hospitals improve relative to all other hospitals, they may nonetheless score lower overall because of weight shifting onto different measures than before. In contrast, we provide an approach to scoring and ranking hospitals that, under reasonable conditions, ensures that hospitals that improve relative to all other hospitals obtain higher scores, while also having the capability to autonomously adjust weights as measures are added or subtracted over time. Rather than exploit statistical correlation, we propose a conic optimization framework that offers a new integrated approach in data envelopment analysis for simultaneous efficiency analysis and performance evaluation. We develop theory that explains the behaviour of our approach, including various properties satisfied by hospital scores at optimality. Using data, we apply our approach to score and rank nearly every hospital in the United States and demonstrate the extent to which it agrees or disagrees with the existing approach to the CMS star ratings.

Funding: The author is grateful for financial support from the University of Chicago Booth School of Business.

Keywords: hospitals • healthcare • input-output analysis • data envelopment analysis • conic programming

1. Introduction

Since July 2016, the Centers for Medicaid and Medicare Services (CMS) has publicly released on a biannual basis its hospital star ratings, which assign one to five stars to most (3,000 to 4,000) hospitals in the United States. They were developed to give consumers a simplified way of assessing hospital quality. The underlying methodology was developed by the Yale Center for Outcomes Research and Evaluation in collaboration with other contractors. The challenge was to combine potentially hundreds of measures of hospital quality, as reported on the Hospital Compare website (https://www.medicare.gov/hospitalcompare), into a single five-star rating for each hospital. Ultimately, it was decided to categorize the measures into seven groups: mortality, safety of care, readmission, patient experience, effectiveness of care, timeliness of care, and efficient use of medical imaging. Although stakeholders could accept the weights chosen for these various categories in constructing a composite score, it was unclear how to rationally weight the individual measures within a given group (as well as which measures should be included in the first place). The latent variable model (LVM) was developed to provide an autonomous computational approach[1] to effectively weight the measures within a given group, so that measures could be easily added and subtracted over time without having to arbitrarily assign explicit a priori weights to them. Hospitals' composite scores could then be partitioned to obtain star ratings, but could also be ordered to obtain rankings.

The LVM posits that quality in a given measure group is driven by only a single latent variable or factor, which is not observable but is estimated to construct a hospital's group score. If this were true, then most of the variance in the (observable) measures of the group would be explained by this single latent variable. However, the scree plots in appendix E of the methodology report Venkatesh et al. (2017b) indicate that the principal component of the various groups typically explains only 20%–40% of the variance. Analysts generally look for at least 70%–90% (Jolliffe 2002, Everitt and Hothorn 2011), and for this reason, as well as others, experts have criticized the approach as not being appropriate (Centers for Medicaid and Medicare Services 2017, Castelluci 2018a, Castelluci 2019, Nerenz et al. 2019).

As we will demonstrate, assuming a model that uses only one latent variable, when more than one is needed, results in undesirable behaviour. The implied measure weights either tend to concentrate largely on one measure or can exhibit "knife-edge" instability, in which an infinitesimal change in correlation can result in a dramatic shift in weight from one set of measures to another. This behaviour occurred in the CMS's hospital star ratings that were to be released in July 2018. Whereas in previous releases nearly all weight in the Safety of Care group was placed on the PSI-90 composite score, in the July 2018 release, the weight shifted to Complications from Hip and Knee Surgeries, which impacts many fewer patients. (See Appendix B for detailed descriptions of the measures.) As a result, many hospitals unexpectedly changed in their star ratings, including Rush University Medical Center, which dropped from five stars down to three, even though they had improved along many measures (Castelluci 2018b). Explaining that there had been changes in how the PSI-90 score was calculated, a CMS official stated, "It's not surprising that when you update a measure that hospital performance is going to change, and therefore the star rating is going to change."[2] After stakeholders raised concerns, the CMS decided not to release to the public the July 2018 ratings, but ultimately did not make any significant changes to the methodology for the February 2019 release that would prevent this behaviour moving forward (Castelluci 2018a).

As a consequence, hospital administrators cannot reasonably anticipate how the weights will change over time, so as to decide on where to focus improvement efforts. The fundamental problem is this: even if a hospital improves along every measure relative to all other hospitals, the LVM offers no guarantee that a hospital's score will not decrease. The purpose of this paper is to propose a new methodology that meets the CMS's requirement to autonomously weight measures as they are added and subtracted over time, while ensuring, under reasonable conditions, that if a hospital improves along every measure relative to all other hospitals, its score improves. A hospital's star rating may still go down under our conditions, but this could only be due to a subsequent clustering step employed by the CMS that takes scores as input and groups hospitals into stars. Our approach thus gives hospitals appropriate incentives to improve over time, and it satisfies a basic property that hospitals would view as being fair.

The current LVM approach depends critically on statistical correlations between measures. In contrast, our approach is based on the following two foundational ideas:

- Hospitals should be evaluated against best performers among all measures, regardless of the statistical correlation between them.
- Measures having the same clinical relevance to a patient but impact more people should be weighted more heavily.

We implement the first by benchmarking a hospital's performance against an efficient frontier of best performance. For the second, we impose constraints on measure weights.

The key to our approach is that every hospital is given its own, unique set of measure weights that score it as close as possible to the best scoring hospitals under the same measure weights, subject to constraints. This contrasts with the CMS's approach, which uses a common set of parameters across all hospitals in estimating the latent variable for a specific hospital. Although giving hospitals their own measure weights may seem to give them unfair advantage, our approach does not choose the weights arbitrarily. Rather, every hospital receives its highest such score possible and is subject to the same collection of constraints that ensure weights obey rational properties. Also, the measure weights for a given hospital flexibly adjust onto the subset of measures it actually reports, addressing the problem of comparing hospitals when measures are missing (e.g., see Chung et al. 2019).

To be specific, our approach calculates a hospital's score by first finding measure weights that score it as close as possible to the highest scores of hospitals on the efficient frontier of best performance, in the direction of ideal performance. It then maximizes the hospital's score among all measure weights that minimize in the first stage. These optimizations are performed subject to problem-specific constraints on the weights, which include the fixed weightings of groups, and the requirement that weights be representative of the number of patients impacted by the measures. These latter constraints are related to the denominator-based weights of Shwartz et al. (2008), but our approach allows slack for optimization over measure weights. The approach is flexible and can accommodate other constraints that policy makers deem relevant.

Our contribution provides a number of benefits relative to the existing LVM:

1. Measure weights obey desirable structural properties under reasonable conditions, including that

scores improve when hospitals improve, where hospital improvement may be either relative to all other hospitals or in absolute terms, depending on how our approach is implemented. We also show that better-performing hospitals score higher. Our approach does not exhibit the unstable knife-edge behaviour discussed above, in which scores and measure weights can change dramatically in response to infinitesimal changes in national data, out of a hospital's control.

2. For many measure groups, the LVM approach concentrates weight on one or a few measures because it relies on correlations between measures, which may be weak or nonexistent. Our approach tends to spread weights across measures even if they are not correlated. Rather than correlation, our approach relies on achieving patient representation through constraints on measure weights.

3. Whereas the LVM approach uses the same loadings in estimating the factor score of a specific hospital when measures are missing, our approach shifts weight to measures that are reported. Thus, for a hospital that does not report a measure that is heavily weighted by the LVM, our approach gives credit or penalizes for measures they do report, whereas the LVM approach can ignore them and effectively reward poor performance or penalize good performance. This has a significant impact on hospital rankings. For example, we find empirically that nearly half of critical access hospitals change the quintile in which they are ranked.

4. We maximize a hospital's score, in the senses described above. This gives hospitals the benefit of the doubt if possible.

5. Our approach does not appear to introduce bias with respect to hospital size, in terms of the relative revising of hospital scores up versus down compared with the CMS approach.

Our approach relies on and substantially extends the range directional model of Portela et al. (2004), which is inspired by the directional distance model of Chambers (1996). Relative to this literature in data envelopment analysis (DEA), we make the following contributions:

1. Although there is much discussion in the literature about evaluating efficiency versus performance, to date there does not appear to be an integrated theory for doing so. We provide a novel two-stage approach using the directional distance model, which first solves for efficiency, and then maximizes the score to evaluate performance subject to first-stage efficiency. We thus contribute to the DEA literature on ranking methods (Adler et al. 2002).

2. We develop new theory around the resulting scores, in particular, perturbation analysis, which does not exist in the DEA literature.

3. Building on Portela et al. (2004) and Chambers (1996), we develop new theoretical properties and interpretations of the directional distance dual formulation. For example, we provide a complete picture of how primal–dual solutions change under affine data transformations.

4. We integrate conic constraints directly into the formulation from the start, and nearly all results and proofs are tailored to incorporate them. This paper appears to be one of the first to apply conic duality (Ben-Tal and Nemirovski 2001) to a DEA model. In the literature, theory for a DEA model is developed first, and when cone constraints are used, they are added afterward for the application at hand (Cooper et al. 2007). Allen et al. (1997) discusses how constraining the multipliers is geometrically equivalent to tightening the cone of normals for the supporting hyperplanes that define the piecewise linear frontier.

5. We add a compelling, significant new application of efficient frontier methods in healthcare (Hollingsworth 2008, Ozcan 2008).

Although we propose an alternative methodology to replace the LVM, we do not offer a comprehensive solution that addresses every issue that has been raised with the current CMS hospital star rating system. Many concerns are merely out of scope for this work, such as the validity of and improvements to underlying measures, the role of patient mix, alternatives to clustering hospital overall scores to obtain ratings, selection of which measures to include, ways of peer grouping hospitals when computing scores and ratings, incentivizing hospitals to report more measures, etc. However, we do address a foundational problem that sits at the core of the current approach. We offer a more "explicit" approach to computing hospital overall scores, which incorporates volume-based weighting with comparisons against an efficient frontier of hospital performance. Because it is based on an optimization framework, it offers more flexibility, stability, and interpretability than the current approach.

The rest of this paper is organized as follows. In Section 2, we describe the latent variable model and give examples of scoring behaviours, including the knife-edge effect mentioned above. In Section 3, we develop our proposed efficient frontier model in a general context, where weights are required to reside in a cone. This model is not specialized to our hospital context until Section 4, where we specify the details of the cone we employ and various properties that are satisfied by optimal hospital scores. In Section 5, we present results from running our model to score; and rank every hospital in the United States that is ranked in the CMS hospital star ratings. Last, in Section 6, we conclude. To make our key insights easier for readers to identify and understand, throughout this paper we highlight properties as they are discussed.

2. Latent Variable Model

In this section, we describe the LVM used by the CMS in their hospital star ratings and then illustrate a variety of scoring behaviours it exhibits. Our discussion is more technical than described in the CMS methodology report (Venkatesh et al. 2017b). Much of the material in Section 2.2 is either in this report or can be found in the literature on factor analysis (Mardia et al. 1979, Everitt 1984, Everitt and Hothorn 2011). We present it concisely here to make this paper self-contained and for reference later.

2.1. Data

We are given the following data on hospital performance. Suppose a set $\mathcal{I}$ of I measures are gathered on a collection $\mathcal{J}$ of J hospitals. For each measure i and hospital j, denote its value by y_i^j, not restricted in sign, and let N_i^j denote the total number of patients included in the measure's denominator, that is, impacted, for this hospital. Thus, $N_i^j = 0$ if measure i is not collected or available for hospital j, and we can denote by $\mathcal{I}^j = \{i \in \mathcal{I} : N_i^j > 0\}$ the set of available measures for hospital j, so that $\mathcal{I} = \bigcup_{j \in \mathcal{J}} \mathcal{I}^j$. Without loss of generality, we assume that all measures are oriented so that larger values represent better performance. Thus, for measure $i \in \mathcal{I}$, let

$$\bar{y}_i = \max\left\{y_i^j : j \in \mathcal{J}, i \in \mathcal{I}^j\right\}$$

denote the maximum performance in measure i among all hospitals that collect it.

Measures $\mathcal{I}$ are also partitioned into different groups, to be discussed later, which play an important role in the CMS methodology (Venkatesh et al. 2017b). Each group has a separate latent variable model, but for simplicity in the following discussion one could imagine $\mathcal{I}$ as containing measures from only one group.

2.2. Brief Description

The methodology assumes that each hospital's reported data represent a single sample from an underlying LVM. The model posits that each hospital possesses a single and unique hidden factor, or latent variable, that cannot be directly observed but drives performance on all measures we can observe. Ultimately, the approach assigns a hospital's score to be the estimate of its latent variable.

Specifically, let $\boldsymbol{y} = \langle y_i \rangle_{i \in \mathcal{I}}$ be a vector that denotesthe random outcome for a hospital across the performance measures. Thus, the $y_i^j \; \forall j \in \mathcal{J}, i \in \mathcal{I}^j$ are samples; that is, whenever the index j is removed, the notation represents a random variable. As discussed above, not all hospitals report all measures. And so, in reality, there are missing data that are merely ignored; that is, there is no model employed to capture the probability a hospital reports a given measure. For simplicity, assume below that all measures $\mathcal{I}$ are reported.

Let f denote the latent variable for a random hospital. The LVM is

$$y_i = \nu_i + \delta_i f + \epsilon_i \quad \forall i \in \mathcal{I}, \tag{1}$$

where the errors ϵ_i are assumed to be normally distributed with mean 0 and variance σ_i^2, and independent across measures i. Hence, performance on measure i is an affine function of the latent variable f, where the parameters ν_i and δ_i represent the intercept and slope, respectively. The parameters ν_i, δ_i, and $\sigma_i^2 \; \forall i \in \mathcal{I}$ are in common across all hospitals, and instead what changes by hospital are the realizations of the latent variable f. This factor f is assumed to be normally distributed with mean 0 and variance 1. Furthermore, it is assumed that f is independent of the random errors $\epsilon_i \; \forall i \in \mathcal{I}$. Hence, when the input data are standardized, the parameter δ_i represents the correlation that performance on measure i has with the latent variable f, and it is called the loading of measure i. Conditional on f, performance is independent across measures.

Although the model has three sets of underlying parameters, that is, δ_i, ν_i, and σ_i^2, the CMS only reports estimates for the loadings in Venkatesh et al. (2017a). Estimates for the others must be obtained by running the model on data using SAS code they make publicly available (discussed in Appendix C).

Without the i subscript, let $\boldsymbol{\delta}$, $\boldsymbol{\nu}$, and $\boldsymbol{\sigma}^2$ denote parameter (column) vectors. Let $\boldsymbol{\Psi}$ denote the $|\mathcal{I}| \times |\mathcal{I}|$ diagonal matrix of variances $\boldsymbol{\sigma}^2$; that is, for each $i \in \mathcal{I}$, the ith diagonal element equals σ_i^2. Thus, we can write (1) in vector form as

$$\boldsymbol{y} = \boldsymbol{\nu} + \boldsymbol{\delta} f + \boldsymbol{\epsilon}. \tag{2}$$

The assumptions discussed above imply that the distribution of $\boldsymbol{y}$ is multivariate normal, that is,

$$\boldsymbol{y} \sim \text{Normal}(\boldsymbol{\nu}, \boldsymbol{\delta}\boldsymbol{\delta}^T + \boldsymbol{\Psi}),$$

where T denotes the transpose operator. Because in this model there is only one factor, there is a unique $\boldsymbol{\delta}$ (ignoring sign changes) and $\boldsymbol{\Psi}$ for every covariance matrix $\boldsymbol{\Sigma}$, that is, such that

$$\boldsymbol{\Sigma} = \boldsymbol{\delta}\boldsymbol{\delta}^T + \boldsymbol{\Psi}.$$

Therefore, the model is identified.

After taking the estimate of $\boldsymbol{\nu}$ to be the sample average $\hat{\nu}_i = (1/|\mathcal{J}|) \sum_{j \in \mathcal{J}} y_i^j \; \forall i \in \mathcal{I}$, model estimation amounts to finding parameter estimates $\hat{\boldsymbol{\delta}}$ and $\hat{\boldsymbol{\Psi}}$ to fit the sample covariance $\hat{S}$, that is,

$$\hat{S} \approx \hat{\boldsymbol{\delta}}\hat{\boldsymbol{\delta}}^T + \hat{\boldsymbol{\Psi}}. \tag{3}$$

This is often done using maximum likelihood estimation, subject to the constraint that $\hat{\boldsymbol{\Psi}}$ be nonnegative. The CMS methodology employs a weighted likelihood function, accounting for missing data; that is,

$$\mathcal{L}\left(\left\{y_i^j\right\}_{j\in\mathcal{J},i\in\mathcal{I}^j}\right) \equiv \prod_{j\in\mathcal{J}}\prod_{i\in\mathcal{I}^j} L\left(y_i^j\right)^{w_i^j},$$

where the likelihood weights satisfy

$$w_i^j = \left(\frac{N_i^j}{\sum_{\{j\in\mathcal{J}:i\in\mathcal{I}^j\}} N_i^j}\right)\left|\left\{j\in\mathcal{J} : i\in\mathcal{I}^j\right\}\right|,$$

and $L(\cdot)$ is the usual normal likelihood function integrated over f under its Bayesian prior. These likelihood weights represent the number of "average" hospitals corresponding to the volume of patients impacted by measure i at hospital j.

Once the parameters $(\hat{\boldsymbol{\nu}}, \hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\Psi}})$ are estimated, the factor score $\hat{f}^j$ for each hospital j is computed using empirical Bayes estimation as follows. Let $p(\cdot)$ be the density function for a standard normal distribution. Conditional on the latent variable f in (1) equaling f^j, y_i is assumed to be normally distributed, with estimated mean $\hat{\nu}_i + \hat{\delta}_i f^j$ and estimated variance $\hat{\sigma}_i{}^2$. Thus, given hospital j's reported measures $y_i^j\ \forall i\in\mathcal{I}^j$, empirical Bayes estimation solves

$$\max_{f^j} \prod_{i\in\mathcal{I}^j}\left(\mathsf{Prob}\left\{y_i = y_i^j \,\middle|\, f^j\right\}\cdot p\left(f^j\right)\right)^{w_i^j}$$

to obtain an estimate $\hat{f}^j$, where $\mathsf{Prob}\{\cdot\,|\,\cdot\}$ denotes the conditional probability operator. Consider the "balanced" case in which the likelihood weights w_i^j all equal 1, that is, representing an average hospital. This estimation can be solved in closed form. When the hospital reports all measures $\mathcal{I}$, written as a (column) vector $\boldsymbol{y}^j$, we can derive

$$\hat{f}^j = \hat{\boldsymbol{\delta}}^T\left[\frac{\hat{\boldsymbol{\Psi}}^{-1}}{1+\hat{\boldsymbol{\delta}}^T\hat{\boldsymbol{\Psi}}^{-1}\hat{\boldsymbol{\delta}}}\right]\left(\boldsymbol{y}^j - \hat{\boldsymbol{\nu}}\right). \tag{4}$$

Hence, the relationship between loadings and the weights on measures is rather complex, though the score $\hat{f}^j$ is an affine function of performance $\boldsymbol{y}^j$. When the likelihood weights are not constrained to equal 1, $\hat{\boldsymbol{\Psi}}^{-1}$ in (4) is replaced with $(\boldsymbol{w}^j)^T\hat{\boldsymbol{\Psi}}^{-1}$, where $\boldsymbol{w}^j$ is the (column) vector of likelihood weights for hospital j given above. Furthermore, when the fit (3) is exact, it can be shown that this expression (4) reduces to

$$\hat{f}^j = \hat{\boldsymbol{\delta}}^T\hat{\mathbf{S}}^{-1}\left(\boldsymbol{y}^j - \hat{\boldsymbol{\nu}}\right).$$

For each measure group, the CMS applies the likelihood-weighted version of (4) only on the subset of measures that a given hospital j reports within that group. It uses the same parameter estimates $(\hat{\boldsymbol{\nu}}, \hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\Psi}})$ across all hospitals, but the likelihood weights as given above are hospital specific. This estimate of the hospital's latent variable becomes an estimate of the hospital's score on that group. After repeating these calculations for each group separately, the hospital's overall score equals a weighted average of these group scores.

2.3. Examples of Scoring Behaviours

Using (4), we can compute how the weights on measures change as a function of the correlation structure. In particular, the weight of a measure i equals the constant given in (4) associated with y_i^j. We present two simple examples to demonstrate some of the scoring behaviours exhibited by the CMS approach. Because the latent variable model is scale invariant (Mardia et al. 1979), it suffices to consider the correlation matrix.

In Figure 1, we show an example with four measures whose correlation structure partitions into two disjoint sets. For symmetry, we suppose the correlation of A and B equals $0.5-\epsilon$, whereas the correlation of C and D equals $0.5+\epsilon$. So, when $\epsilon = 0$, there is complete symmetry. When $\epsilon > 0$, the correlation of C and D dominates that of A and B, and vice versa when $\epsilon < 0$.

As shown in Figure 2, there is a knife edge at $\epsilon = 0$, in which the model either puts all weight on A and B, and none on C and D, or vice versa. This occurs because there is only one factor in the model, but two are needed to capture both correlations. The model picks the component having the greatest correlation in the maximum likelihood estimation of parameters. This effect occurs regardless of how small ϵ may be. A slight perturbation can cause the weights to shift dramatically.

To see the impact of this knife-edge effect on hospital scores over time, suppose $\epsilon < 0$ and a hospital performs better on A and B than C and D. The hospital scores well because only its strongest measures are

Figure 1. Correlation Structure for the Example with Four Measures

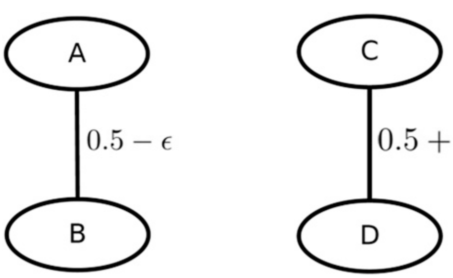


Figure 2. Measure Weights for the Example with Four Measures, as a Function of the Correlation Offset

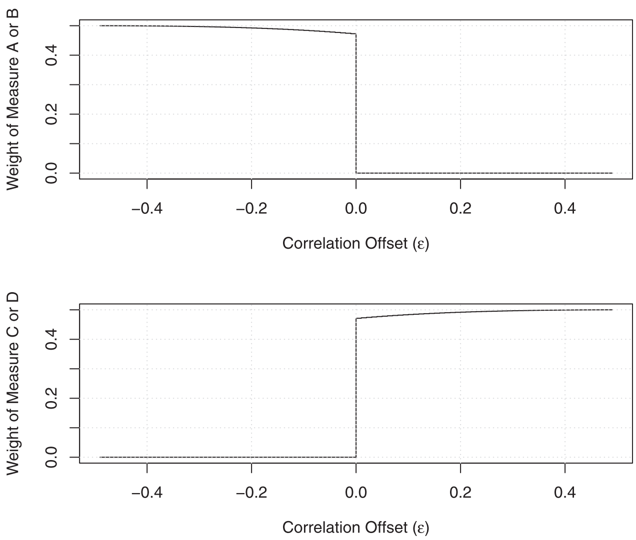
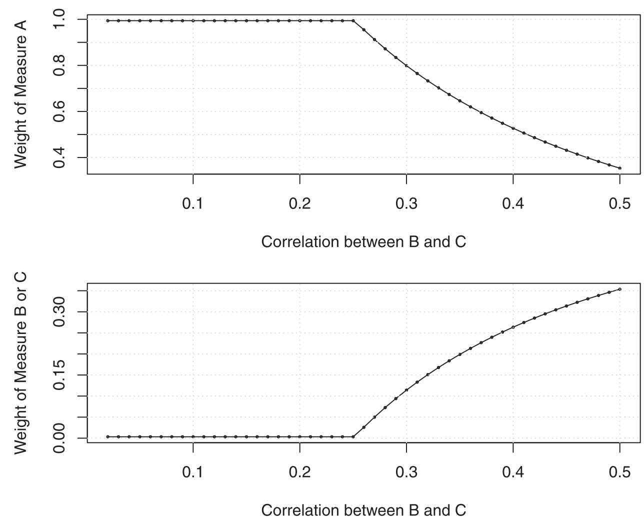


Figure 4. Measure Weights for the Example with Three Measures, as a Function of the Correlation Between B and C

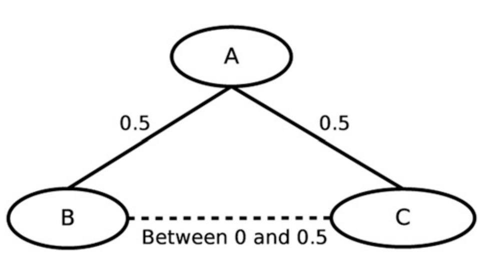


counted, that is, measures C and D are ignored. Over the course of time, suppose the hospital improves along all measures, but its improved performances on C and D are still below its previous performances on A and B. Suppose that in the next round of ratings, $\epsilon > 0$ for reasons outside their control, and possibly even because of spurious floating point calculations. Even though the hospital has improved along every dimension, its score will have decreased, because now measures A and B are ignored, and only its (improved) performances on C and D matter.

In this example, the measures partition cleanly into two distinct groups, but similar undesirable behaviour can occur even when measures are all correlated with one another. In the next example, shown in Figure 3, there are measures labeled A, B, and C. Whereas A and B have correlation 0.5, as do A and C, we vary the correlation between B and C from 0 to 0.5. In Figure 4, we plot how the weights on measures A, B, and C vary as a result. Because of symmetry, the weights on measures B and C equal one another.

When the correlation between B and C is between 0 and 0.25, only measure A is given any weight, that is, measures B and C receive no weight. From 0.25 to 0.5, the weights on measures B and C increase until all three weights equal one another at 0.354, when the correlations all equal 0.5. This effect occurs because if B and C were correlated only with the same latent factor as A, they would need to be correlated with each other at least at the level of 0.25, given their correlations of 0.5 with A. Hence, they matter less in the scoring system to the point of being completely ignored even when they are still significantly correlated at 0.25. This concentration of weight on measure A may not be desirable behaviour if it is believed that B and C are meaningful measures that impact many patients, regardless of their correlations.

Figure 3. Correlation Structure for the Example with Three Measures



In both of the examples, weaker or nonexistent correlations indicate the need for more than one latent factor. By contrast, our proposed approach does not rely on correlations.

3. Efficient Frontier Model

Our approach consists of two steps. Fixing a hospital under consideration, our first model, presented in Section 3.1, calculates an overall score that is as close as possible to that of the best scoring hospitals, in particular a virtual hospital, which are located on the efficient frontier in the direction of ideal performance. Although this initial score can be unique, and hence maximal, there are cases when it is not. Therefore, as a second stage, presented in Section 3.2, we deduce a model that computes a maximal such score, taking as input the distance to the efficient frontier calculated from the first stage.

Our first model constrains measure weights to reside in a convex cone, which, for the moment, we do not specify, but which is intended to impose "agreed-upon" constraints on measure weights. (We map this cone into a bounded convex set for the second-stage problem.) This cone will capture context-specific constraints on the measure weights, for example, constraints requiring that CMS's weightings of the measure groups are satisfied. We postpone its definition for several reasons. First, our approach and underlying theory do not rely on any particular specification for this convex cone, whose details may be controversial but not essential to the approach's foundation. Second, separating out the cone's specification from the model provides a parsimonious description of the approach, making it more understandable and streamlining its theoretical development. Third, this expository approach will hopefully make it easier for others to transfer our underlying approach to other nonhospital application settings.

3.1. Stage 1: Minimizing the Distance to the Efficient Frontier

Our goal is to calculate a score S^j for each hospital $j \in \mathcal{J}$, which can be used to rank hospitals. Let $o \in \mathcal{J}$ denote a hospital for which a score will be calculated. Given a *nonnegative* convex cone C^o, which may depend on the hospital o being assessed, our first math program minimizes the distance to the efficient frontier:

(Stage 1)

$$(\mathbf{D}^o) \quad \Delta^o = \min_{\theta \in \mathbb{R}, \mu \in \mathbb{R}^{|\mathcal{I}^o|}} \theta - \sum_{i \in \mathcal{I}^o} y_i^o \mu_i \tag{5a}$$

$$\text{s.t.} \quad \sum_{i \in \mathcal{I}^o \cap \mathcal{I}^j} y_i^j \mu_i \le \theta \quad \forall j \in \mathcal{J}, \tag{5b}$$

$$\sum_{i \in \mathcal{I}^o} (\bar{y}_i - y_i^o)\mu_i \ge 1, \tag{5c}$$

$$\mu \in C^o. \tag{5d}$$

When the hospital $o \in \mathcal{J}$ being evaluated is clear from context, we will denote an optimal solution by (θ^*, μ^*), and otherwise by (θ^o, μ^o). Letting $\Gamma^o \equiv \sum_{i \in \mathcal{I}^o} \mu_i^o$, we take hospital o's first-stage score to be

$$S^o = \sum_{i \in \mathcal{I}^o} y_i^o \cdot (\mu_i^o / \Gamma^o), \tag{6}$$

which is a convex combination of measures. When the scaling factor Γ^o is not present, we refer to the score as unscaled and write in script font $\mathcal{S}^o$.

To interpret this optimization, we formulate its dual. Define the dual cone of C^o, for each $o \in \mathcal{J}$, by

$$C_\star^o = \left\{ s \in \mathbb{R}^{|\mathcal{I}^o|} : s^T \mu \ge 0 \quad \forall \mu \in C^o \right\},$$

where $s^T \mu$ is the scalar product of s and μ. Then, we can write the dual program as

$$(\mathbf{P}^o) \quad \max_{\beta \in \mathbb{R}, \lambda \in \mathbb{R}^{|\mathcal{J}|}, s \in \mathbb{R}^{|\mathcal{I}^o|}} \beta \tag{7a}$$

$$\text{s.t.} \quad y_i^o + \beta \cdot (\bar{y}_i - y_i^o) + s_i = \sum_{\{j \in \mathcal{J} : i \in \mathcal{I}^j\}} y_i^j \lambda_j \quad \forall i \in \mathcal{I}^o, \tag{7b}$$

$$\sum_{j \in \mathcal{J}} \lambda_j = 1, \tag{7c}$$

$$\beta, \lambda \ge \mathbf{0}, \tag{7d}$$

$$s \in C_\star^o. \tag{7e}$$

Again, when the hospital under assessment $o \in \mathcal{J}$ is clear from context, we will express an optimal solution as $(\beta^*, \lambda^*, s^*)$, but otherwise write $(\beta^o, \lambda^o, s^o)$. When $C^o = \mathbb{R}_+^{|\mathcal{I}^o|}$, then $C_\star^o = \mathbb{R}_+^{|\mathcal{I}^o|}$. In this case, $(\mathbf{P}^o)$ is the same as the range directional model given in Portela et al. (2004), because with $s_i \ge 0 \ \forall i \in \mathcal{I}^o$, the constraints (7b) become

$$y_i^o + \beta \cdot (\bar{y}_i - y_i^o) \le \sum_{\{j \in \mathcal{J} : i \in \mathcal{I}^j\}} y_i^j \lambda_j \quad \forall i \in \mathcal{I}^o. \tag{8}$$

More generally, when $C^o \subseteq \mathbb{R}_+^{|\mathcal{I}^o|}$, we have $C_\star^o \supseteq \mathbb{R}_+^{|\mathcal{I}^o|}$, and so the above inequality need not hold depending on the solution.

When the cone C^o is polyhedral, the above two programs are linear programs, and we can appeal to the usual optimality conditions. In general, C^o is not required to be polyhedral, and may contain, for example, second-order cone constraints. In order for strong duality to hold, meaning that both primal and dual programs have optimal solutions with no duality gap, it suffices that the cone C^o be closed, pointed, and have a nonempty interior. See Ben-Tal and Nemirovski (2001) for more details. We will revisit this issue later when we specify the cone C_o that we use, but for now we assume that C^o is defined such that strong duality holds.

Feasibility in $(\mathbf{D}^o)$ and $(\mathbf{P}^o)$ is easy to achieve. To avoid degenerate cases, we make the following assumption.

Assumption 1 (No Perfect Hospital)**.** *For every hospital $j \in \mathcal{J}$, there exists $i \in \mathcal{I}^j$ such that $y_i^j < \bar{y}_i$.*

Lemma 1. *Under Assumption 1, if C^o contains at least one strictly positive element, then there exist feasible solutions to $(\mathbf{D}^o)$ and $(\mathbf{P}^o)$.*

Proof. Let $\mu' \in C^o$ be strictly positive. If (5c) is violated, because C^o is a nonnegative cone, and because Assumption 1 ensures the left-hand side is positive, we can multiply μ' by any positive constant until the inequality is satisfied. Therefore, assume μ' satisfies (5c).

Now set $\theta' = \max_{j\in\mathcal{J}} \sum_{i\in\mathcal{I}^o\cap\mathcal{I}^j} y_i^j \mu_i'$. This solution $(\theta', \boldsymbol{\mu}')$ is a feasible solution to $(\mathbf{D}^o)$ by construction. To obtain a feasible solution to $(\mathbf{P}^o)$, let $\lambda_j' = 0$ for all $j \in \mathcal{J} \setminus \{o\}$ and $\lambda_o' = 1$. With $\beta' = 0$ and $s_i' = 0\ \forall i \in \mathcal{I}^o$, so that $\mathbf{s}' \in C_\star^o$, we obtain feasibility. □

Given optimal solutions $(\theta^*, \boldsymbol{\mu}^*)$ and $(\beta^*, \boldsymbol{\lambda}^*, \mathbf{s}^*)$ to $(\mathbf{P}^o)$ and $(\mathbf{D}^o)$, respectively, the complementary slackness conditions (Ben-Tal and Nemirovski 2001) are

$$\left[\theta^* - \sum_{i\in\mathcal{I}^o\cap\mathcal{I}^j} y_i^j \mu_i^*\right] \cdot \lambda_j^* = 0 \quad \forall j \in \mathcal{J}, \tag{9}$$

$$\left[\sum_{i\in\mathcal{I}^o} (\bar{y}_i - y_i^o)\mu_i^* - 1\right] \cdot \beta^* = 0, \quad \text{and} \tag{10}$$

$$\mathbf{s}^{*T}\boldsymbol{\mu}^* = 0. \tag{11}$$

These conditions, together with (strict) primal–dual feasibility, imply no duality gap, that is,

$$\theta^* - \sum_{i\in\mathcal{I}^o} y_i^o \mu_i^* = \beta^*, \tag{12}$$

so that the optimal objective values of $(\mathbf{P}^o)$ and $(\mathbf{D}^o)$ are equal.

When $C^o = C_\star^o = \mathbb{R}_+^{|\mathcal{I}^o|}$, $s_i \ge 0$ and $\mu_i \ge 0$ imply that (11) becomes $s_i^*\mu_i^* = 0\ \forall i \in \mathcal{I}^o$. Thus, in this case, if a measure $i \in \mathcal{I}^o$ is positively weighted, that is, $\mu_i^* > 0$, then $s_i^* = 0$. In the general case, a feasible solution may have either inequality $\le$ or $\ge$ in (8), depending on $C_\star^o$, provided the $\boldsymbol{\mu}^*$-weighted slacks sum to zero. This has implications for interpreting optimal solutions.

Aggregating (9) over $j \in \mathcal{J}$, using (7c), yields

$$\theta^* = \theta^* \sum_{j\in\mathcal{J}} \lambda_j^* = \sum_{j\in\mathcal{J}} \lambda_j^* \sum_{i\in\mathcal{I}^o\cap\mathcal{I}^j} y_i^j \mu_i^*.$$

Reversing the last summations on the right-hand side, we obtain

$$\theta^* = \sum_{i\in\mathcal{I}^o} \mu_i^* \left(\sum_{\{j\in\mathcal{J}: i\in\mathcal{I}^j\}} y_i^j \lambda_j^*\right). \tag{13}$$

Only hospitals $j \in \mathcal{J}$ for which $\lambda_j^* > 0$ enter into this calculation. From (9), for such hospitals, we have

$$\theta^* = \sum_{i\in\mathcal{I}^o\cap\mathcal{I}^j} y_i^j \mu_i^*.$$

However, because θ enters the constraints in $(\mathbf{D}^o)$ only through (5b), and the objective (5a) minimizes, an optimal solution must satisfy

$$\theta^* = \max_{j\in\mathcal{J}} \sum_{i\in\mathcal{I}^o\cap\mathcal{I}^j} y_i^j \mu_i^*.$$

Thus,

$$\arg\max_{j\in\mathcal{J}} \sum_{i\in\mathcal{I}^o\cap\mathcal{I}^j} y_i^j \mu_i^* \supseteq \{j \in \mathcal{J} : \lambda j^* > 0\}, \tag{14}$$

with set equality for any optimal solution satisfying strict complementarity. As a consequence, through (13), we see that θ^* equals the unscaled score of a virtual hospital that is a convex combination of hospitals all having maximum unscaled scores under $\boldsymbol{\mu}^*$.

On the other hand, after substituting s_i^* from (7b), (11) reads

$$\sum_{i\in\mathcal{I}^o} \mu_i^* \cdot \left(y_i^o + \beta^* \cdot (\bar{y}_i - y_i^o)\right) = \sum_{i\in\mathcal{I}^o} \mu_i^* \left(\sum_{\{j\in\mathcal{J}: i\in\mathcal{I}^j\}} y_i^j \lambda_j^*\right), \tag{15}$$

where the right-hand side matches that of (13) and thus equals θ^*. However, when (11) implies $s_i^*\mu_i^* = 0\ \forall i \in \mathcal{I}^o$, we see that $\mu_i^* > 0$ only for measures $i \in \mathcal{I}^o$ such that

$$y_i^o + \beta^* \cdot (\bar{y}_i - y_i^o) = \sum_{\{j\in\mathcal{J}: i\in\mathcal{I}^j\}} y_i^j \lambda_j^* \quad \forall i \in \mathcal{I}^o. \tag{16}$$

Because $\beta^* \ge 0$ in $(\mathbf{D}^o)$, we can interpret β^* as the distance from hospital o to this virtual hospital given by the right-hand side of (16), in the vector direction $\langle \bar{y}_i - y_i^o \rangle_{\forall i\in\mathcal{I}^o}$ of best performance, that is, toward the ideally performing hospital with measures $\bar{y}_i\ \forall \in \mathcal{I}^o$. Thus, by construction, this virtual hospital dominates hospital o along every measure dimension that receives positive weight, and because it is a mixture of real hospitals, as opposed to $\bar{\mathbf{y}}$, it makes for a reasonable benchmark. In the general case of (11), this virtual hospital is offset by s_i^* appearing on the left-hand side of (16), but we retain the weighted version (15) of (16).

To complete the picture, we need to account for the action of inequalities (5c). From (10), when $\beta^* > 0$, we must have equality in (5c). However, we next show that we can obtain equality in (5c) even when $\beta^* = 0$.

Lemma 2. *Without loss of optimality,*

$$\sum_{i\in\mathcal{I}^o} (\bar{y}_i - y_i^o)\mu_i^* = 1. \tag{17}$$

Proof. If $\beta^* > 0$, then (17) follows from (10). On the other hand, if $\beta^* = 0$, then suppose

$$\phi^* \equiv \sum_{i\in\mathcal{I}^o} (\bar{y}_i - y_i^o)\mu_i^* > 1.$$

Consider an alternative solution,

$$\tilde{\boldsymbol{\mu}} \leftarrow \boldsymbol{\mu}^*/\phi^*,$$
$$\tilde{\theta} \leftarrow \theta^*/\phi^*.$$

This solution satisfies (5b), because $\phi^* > 1$ implies, for all $j \in \mathcal{J}$,

$$\sum_{i \in \mathcal{I}^o \cap \mathcal{I}^j} y_i^j \tilde{\mu}_i - \tilde{\theta} = (1/\phi^*) \left(\sum_{i \in \mathcal{I}^o \cap \mathcal{I}^j} y_i^j \mu_i^* - \theta^* \right) \leq 0.$$

It also satisfies (5c), because

$$\sum_{i \in \mathcal{I}^o} (\overline{y}_i - y_i^o) \tilde{\mu}_i = (1/\phi^*) \sum_{i \in \mathcal{I}^o} (\overline{y}_i - y_i^o) \mu_i^* = 1.$$

Furthermore, $\tilde{\boldsymbol{\mu}} \in C^o$ because $1/\phi^* > 0$ and C^o is a cone. Finally, the objective function (5a) equals

$$\tilde{\theta} - \sum_{i \in \mathcal{I}^o} y_i^o \tilde{\mu}_i = (1/\phi^*) \left(\theta^* - \sum_{i \in \mathcal{I}^o} y_i^o \mu_i^* \right) = 0$$

from (12) because $\beta^* = 0$. Therefore, $(\tilde{\theta}, \tilde{\boldsymbol{\mu}})$ is an alternative optimal solution to $(\mathbf{D}^o)$. □

Thus, (17) says that there always exists an optimal solution for which the difference in unscaled scores between the ideal hospital and hospital o equals 1. Furthermore, we have the following.

Lemma 3. *In an optimal solution,* $\beta^* \leq 1$.

Proof. Suppose $\beta^* > 1$. Then, by complementary slackness (10),

$$\sum_{i \in \mathcal{I}^o} (\overline{y}_i - y_i^o) \mu_i^* = 1$$

implies that there exists an $i \in \mathcal{I}^o$ such that $\overline{y}_i - y_i^o > 0$ and $\mu_i^* > 0$. Consequently, (11) implies

$$\sum_{i \in \mathcal{I}^o} \mu_i^* \cdot \left(\sum_{\{j \in \mathcal{J}:\, i \in \mathcal{I}^j\}} y_i^j \lambda_j^* \right) = \sum_{i \in \mathcal{I}^o} \mu_i^* \cdot \left(y_i^o + \beta^* \cdot (\overline{y}_i - y_i^o) \right) > \sum_{i \in \mathcal{I}^o} \overline{y}_i \mu_i^*.$$

Thus, there must exist an $i \in \mathcal{I}^o$ such that

$$\overline{y}_i < \sum_{\{j \in \mathcal{J}: i \in \mathcal{I}^j\}} y_i^j \lambda_j^* \leq \overline{y}_i \sum_{\{j \in \mathcal{J}: i \in \mathcal{I}^j\}} \lambda_j^*.$$

If $\overline{y}_i = 0$, then this reads $0 < 0$. If $\overline{y}_i > 0$, then this implies

$$\sum_{\{j \in \mathcal{J}: i \in \mathcal{I}^j\}} \lambda_j^* > 1,$$

and if $\overline{y}_i < 0$, then the strict inequality is reversed, in either case contradicting (7c) and $\boldsymbol{\lambda} \geq \mathbf{0}$ in (7d). □

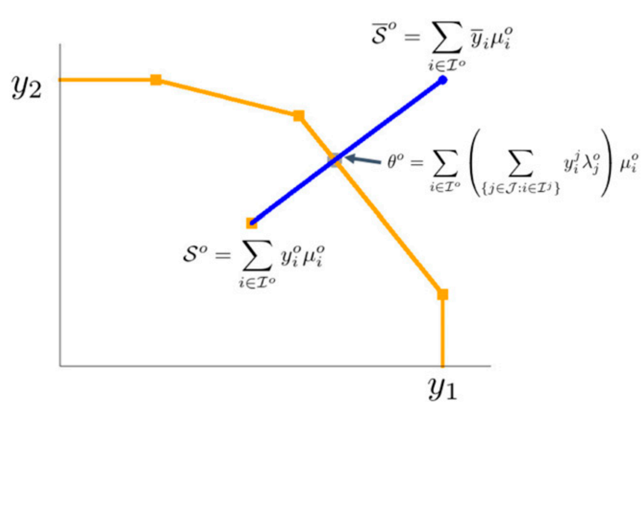
So we can interpret β^* as the distance to the virtual hospital as a fraction of the distance to the ideal hospital. Hence, if $\beta^* = 0$, then hospital o must be in the group of maximally scoring hospitals (14). The normalization (5c), that is, (17), establishes a scale such that differences in scores represent distances, as in (12). From the above interpretation of θ^*, (12) says that at optimality, the difference in scores between the virtual hospital and hospital o equals the distance β^*.

Definition 1. The *technical efficiency* of a hospital $o \in \mathcal{J}$ equals $T^o \equiv 1 - \beta^o$, with $0 \leq T^o \leq 1$. A hospital $o \in \mathcal{J}$ is said to be on the *efficient frontier* if $\beta^o = 0$ or, equivalently, $T^o = 1$.

Indeed, a corollary to a result we show in Appendix A states that the virtual hospital $\langle \sum_{\{j \in \mathcal{J}: i \in \mathcal{I}^j\}} y_i^j \lambda_j^* \rangle_{\forall i \in \mathcal{I}^o}$ is on the efficient frontier. From the discussion above, this implies that all hospitals $j \in \mathcal{J}$ such that $\lambda_j^* > 0$ are on the efficient frontier. These hospitals are called *reference* hospitals, and the set of such hospitals $\{j \in \mathcal{J} : \lambda_j^o > 0\}$ is called the *reference set* for hospital o.

To summarize, Figure 5 represents the optimization performed by the primal–dual pair $(\mathbf{P}^o)$ and $(\mathbf{D}^o)$, with two performance measure dimensions but generalizing to many. The picture depicts four hospitals, three of which are on the efficient frontier connecting the outer points. The hospital o being evaluated is in the interior, with unscaled score $\mathcal{S}^o = \sum_{i \in \mathcal{I}^o} y_i^o \mu_i^o$. The line segment originating at hospital o extends to the ideal hospital with performance $\{\overline{y}_i\}_{\forall i \in \mathcal{I}^o}$ and unscaled score $\overline{\mathcal{S}}^o \equiv \sum_{i \in \mathcal{I}^o} \overline{y}_i \mu_i^o$. The constraints (5c) normalize the difference in these scores to be $\overline{\mathcal{S}}^o - \mathcal{S}^o = 1$, using Lemma 2. This line segment intersects the efficient frontier at the point $\langle \sum_{\{j \in \mathcal{J}: i \in \mathcal{I}^j\}} y_i^j \lambda_j^o \rangle_{\forall i \in \mathcal{I}^o}$, which has unscaled score θ^o, shown in the figure, from (15). Thus, we can see that $(\mathbf{D}^o)$ finds the weights $\boldsymbol{\mu}^o \in C^o$ that score hospital o as close as possible to the virtual hospital, in fact all reference hospitals, on the efficient frontier in the direction of the ideal hospital, potentially offset by $\mathbf{s}^*$ to account for conic constraints on $\boldsymbol{\mu}^*$ (not shown in Figure 5). This leads to the following fundamental property of our approach.

Figure 5. Stage 1 Scores Hospital o as Close as Possible to the Virtual Hospital on the Efficient Frontier in the Direction of the Ideally Performing Hospital, Assuming Slack Offset $\mathbf{s}^o = \mathbf{0}$



Property 1 (Minimizing Distance to the Efficient Frontier). *Each hospital $o \in \mathcal{J}$ is evaluated using hospital-specific weights $\boldsymbol{\mu}^o \in C^o$ that minimize the difference between the unscaled score of maximally scoring hospitals under μ^o and hospital o's unscaled score $\mathcal{S}^o$, subject to the normalization $\overline{\mathcal{S}}^o - \mathcal{S}^o = 1$.*

The discussion above shows how $(\mathbf{D}^o)$ achieves this property.

Furthermore, technical efficiency is defined as

$$T^o = \frac{\overline{\mathcal{S}}^o - \theta^o}{\overline{\mathcal{S}}^o - \mathcal{S}^o},$$

or just $\overline{\mathcal{S}}^o - \theta^o$, because the denominator is normalized to 1. The next result shows that, under simple conditions, a hospital that scores the maximum along any single dimension has 100% efficiency, even if this hospital scores poorest among all hospitals across all other measures. Thus, although the technical efficiency of a hospital provides information, it is not appropriate for scoring hospital performance in our context.

Proposition 1. *Suppose $C^o_\star = \mathbb{R}_+^{|\mathcal{I}^o|}$, and for some $i \in \mathcal{I}^o$, $y_i^o = \overline{y}_i > 0$ and $y_i^j < \overline{y}_i$ for all $j \in \mathcal{J} \setminus \{o\}$. Then, $\beta^o = 0$.*

Proof. Consider $(\mathbf{D}^o)$ and $(\mathbf{P}^o)$ for hospital o. For the $i \in \mathcal{I}^o$ under consideration, feasibility in (7b) reads

$$\overline{y}_i = y_i^o + \beta^* \cdot (\overline{y}_i - y_i^o) \le \sum_{\{j \in \mathcal{J}: i \in \mathcal{I}^j\}} y_i^j \lambda_j^*.$$

Suppose $\lambda_{j'}^* > 0$ for some $j' \in \mathcal{J} \setminus \{o\}$ such that $i \in \mathcal{I}^{j'}$. Then, using $\overline{y}_i > 0$,

$$\begin{aligned}\overline{y}_i \le \sum_{\{j \in \mathcal{J}: i \in \mathcal{I}^j\}} y_i^j \lambda_j^* &= y_i^{j'} \lambda_{j'}^* + \sum_{\{j \in \mathcal{J}: j \neq j', i \in \mathcal{I}^j\}} y_i^j \lambda_j^* \\ &\le y_i^{j'} \lambda_{j'}^* + \left(1 - \lambda_{j'}^*\right)\overline{y}_i \\ &< \overline{y}_i,\end{aligned}$$

because (7c) and (7d) imply $\sum_{\{j \in \mathcal{J}: i \in \mathcal{I}^j\}} \lambda_j^* \le 1$, a contradiction. Thus, if $\lambda_j^* > 0$ for some $j \in \mathcal{J}$, then $y_i^j = \overline{y}_i$. Because $y_i^j < \overline{y}_i$ for all $j \in \mathcal{J} \setminus \{o\}$, we must have $\lambda_o^* = 1$. From complementary slackness (9), $\lambda_o^* > 0$ and (12) imply

$$\beta^* = \theta^* - \sum_{i \in \mathcal{I}^o} y_i^o \mu_i^* = 0. \quad \square$$

When $C^o_\star \neq \mathbb{R}_+^{|\mathcal{I}^o|}$, this proof fails because the sign of s_i^* may be negative.

3.2. Stage 2: Maximizing the Hospital Score

To motivate the need for a second stage, consider the following.

Proposition 2. *If $\beta^* = 0$ and $\sum_{i \in \mathcal{I}^o} y_i^o \mu_i^* \neq 0$, then there exist alternative optimal solutions $(\tilde{\theta}, \tilde{\boldsymbol{\mu}})$ to $(\mathbf{D}^o)$ such that $\sum_{i \in \mathcal{I}^o} y_i^o \tilde{\mu}_i$ is unbounded.*

Proof. Consider any scalar $\alpha \ge 1$, and proposed solution $\tilde{\boldsymbol{\mu}} = \alpha \boldsymbol{\mu}^*$ and $\tilde{\theta} = \alpha\theta^*$. Then, (5b) becomes, for all $j \in \mathcal{J}$,

$$\sum_{i \in \mathcal{I}^o \cap \mathcal{I}^j} y_i^j \tilde{\mu}_i - \tilde{\theta} = \alpha \cdot \left(\sum_{i \in \mathcal{I}^o \cap \mathcal{I}^j} y_i^j \mu_i^* - \theta^*\right) \le 0,$$

because $\alpha \ge 0$. Also, (5c) becomes

$$\begin{aligned}&\sum_{i \in \mathcal{I}^o} (\overline{y}_i - y_i^o)\tilde{\mu}_i - 1 \\ &\quad = \alpha \cdot \sum_{i \in \mathcal{I}^o} (\overline{y}_i - y_i^o)\mu_i^* - 1 \\ &\quad = \sum_{i \in \mathcal{I}^o} (\overline{y}_i - y_i^o)\mu_i^* - 1 + (\alpha - 1) \cdot \sum_{i \in \mathcal{I}^o} (\overline{y}_i - y_i^o)\mu_i^* \\ &\quad \ge \sum_{i \in \mathcal{I}^o} (\overline{y}_i - y_i^o)\mu_i^* - 1 \\ &\quad \ge 0,\end{aligned}$$

where the second to last inequality follows from $\alpha \ge 1$. Finally, because $\boldsymbol{\mu}^* \in C^o$, we must have $\alpha\boldsymbol{\mu}^* \in C^o$. However, the objective value is

$$\tilde{\theta} - \sum_{i \in \mathcal{I}^o} y_i^o \tilde{\mu}_i = \alpha \cdot \left(\theta^* - \sum_{i \in \mathcal{I}^o} y_i^o \mu_i^*\right) = 0,$$

because $\beta^* = 0$ using (12). Thus, $(\tilde{\theta}, \tilde{\boldsymbol{\mu}})$ is an alternative optimal solution for any $\alpha \ge 1$. Now let $\alpha \to \infty$, and use the assumption that $\sum_{i \in \mathcal{I}^o} y_i^o \mu_i^* \neq 0$. $\square$

The proof constructs alternative optimal solutions to $(\mathbf{D}^o)$ by rescaling $\boldsymbol{\mu}^*$ by a scalar $\alpha \ge 1$. Because this scaling factor appears in both the numerator and denominator of S^o, as displayed in (6), it does not impact the score. However, we next give an example for which $\beta^* > 0$ but yet neither the unscaled score $\mathcal{S}^o$ nor scaled score S^o is unique.

Example 1 (Stage 1 Score Can Be Nonunique). Consider four hospitals, each with scores on two different measures, as shown in Table 1. Hospital 1 lies directly on the path from hospital o to $\overline{\boldsymbol{y}} = \langle 4, 5\rangle$, that is, along the vector $\langle 3, 3\rangle$ originating from hospital o. This simplifies the reference set of hospitals to include only hospital 1.

Table 1. Data for the Example Showing Nonuniqueness of the Stage 1 Score

Hospital	Measure 1	Measure 2
o	1	2
1	3	4
2	4	0
3	0	5

Also, hospital o is located at $\langle 1,2\rangle$, and thus the origin is not on this line leading to $\overline{\boldsymbol{y}}$. Consider the cone $C^o = \mathbb{R}^2_+$, that is, the nonnegative orthant.

It is easy to check that $\beta^* = 2/3$ and $\lambda_1^* = 1$ is a feasible solution to ($\mathbf{P}^o$), with $\lambda_j^* = 0$ for all other $j \neq 1$. Indeed, (7b) becomes (written in vector form)

$$\boldsymbol{y}^o + \beta^* \cdot (\overline{y} - \boldsymbol{y}^o) = \langle 1,2\rangle + (2/3)[\langle 4,5\rangle - \langle 1,2\rangle] = \langle 3,4\rangle = \boldsymbol{y}^1.$$

For there to be no duality gap, (12) requires

$$2/3 = \theta^* - (\mu_1^* + 2\mu_2^*). \tag{18}$$

From complementary slackness (9), $\lambda_1^* > 0$ implies

$$\theta^* = 3\mu_1^* + 4\mu_2^*, \tag{19}$$

and from complementary slackness (10), $\beta^* > 0$ implies

$$1 = \sum_{i=1}^{2} (\overline{y}_i - y_i^o)\mu_i^* = 3\mu_1^* + 3\mu_2^*. \tag{20}$$

Furthermore, feasibility in (5b) requires

$$4\mu_1^* \le \theta^* \tag{21}$$

and

$$5\mu_2^* \le \theta^*. \tag{22}$$

It is easy to check that one of (18)–(20) is linear dependent on the other two. This leaves two equations and three unknowns. Sequentially applying (21) and (22) with equality yields two solutions (among an infinite set):

- **Solution 1.** $\mu_1^* = 4/15$, $\mu_2^* = 1/15$, and $\theta^* = 16/15$. This solution scores hospital o as

$$S^o = \frac{\mu_1^* + 2\mu_2^*}{\mu_1^* + \mu_2^*} = 6/5.$$

- **Solution 2.** $\mu_1^* = 1/12$, $\mu_2^* = 1/4$, and $\theta^* = 5/4$. This solution gives hospital o a score of $S^o = 7/4$.

Thus, the score of hospital o is not uniquely determined by Stage 1, and the second score is as large as possible.

To maximize the scaled score, consider the following program with a nonlinear objective function:

$$\max_{\theta\in\mathbb{R},\boldsymbol{\mu}\in\mathbb{R}^{|\mathcal{I}^o|},\Gamma\in\mathbb{R}} \sum_{i\in\mathcal{I}^o} y_i^o \cdot (\mu_i/\Gamma) \tag{23a}$$
$$\text{s.t.} \quad \theta - \sum_{i\in\mathcal{I}^o} y_i^o \mu_i = \Delta^o, \tag{23b}$$
$$\sum_{i\in\mathcal{I}^o} \mu_i = \Gamma, \tag{23c}$$
$$\text{(5b), (5c), (5d)}. \tag{23d}$$

The first constraint, (23b), forces the technical efficiency to equal the optimized efficiency from ($\mathbf{D}^o$). The constraint (23c) merely handles accounting so that the appropriate scaling can be used in the objective function, (23a). Therefore, this program maximizes the hospital's score among feasible scores that minimize the distance to the efficient frontier.

Although this program has a nonlinear objective function, we can transform variables to make it linear:

$$\rho_i = \mu_i/\Gamma \quad \forall i \in \mathcal{I}^o, \tag{24}$$
$$\eta = \theta/\Gamma, \tag{25}$$
$$\phi = 1/\Gamma. \tag{26}$$

This yields

$$\text{(Stage 2)} \quad \max_{\eta\in\mathbb{R},\boldsymbol{\rho}\in\mathbb{R}^{|\mathcal{I}^o|},\phi\in\mathbb{R}} \sum_{i\in\mathcal{I}^o} y_i^o \rho_i \tag{27a}$$
$$\text{s.t.} \quad \sum_{i\in\mathcal{I}^o\cap\mathcal{I}^j} y_i^j \rho_i \le \eta \quad \forall j \in \mathcal{J}, \tag{27b}$$
$$\sum_{i\in\mathcal{I}^o} (\overline{y}_i - y_i^o)\rho_i \ge \phi, \tag{27c}$$
$$\eta - \sum_{i\in\mathcal{I}^o} y_i^o \rho_i = \Delta^o \phi, \tag{27d}$$
$$\phi \ge 0, \tag{27e}$$
$$\sum_{i\in\mathcal{I}^o} \rho_i = 1, \tag{27f}$$
$$\boldsymbol{\rho} \in C^o. \tag{27g}$$

The objective (27a) maximizes the scaled score $\mathcal{S}^o$ as a convex combination of measures, where the ρ_i's sum to one from (27f), which follows directly from (23c), and satisfy the conic constraints (27g), including nonnegativity. Because $\boldsymbol{\mu} \ge 0$ implies $\Gamma \ge 0$ from (23c), we obtain $\phi \ge 0$. Constraints (27d) fix the distance to the efficient frontier, Δ^o, calculated in Stage 1 and derived from (23b). The remaining constraints, (27b) and (27c), follow naturally by variable substitution in (5b) and (5c), respectively.

Given an optimal solution $(\eta^*, \boldsymbol{\rho}^*, \phi^*)$, reversing the mapping (24)–(26) produces an (alternative) optimal solution to ($\mathbf{D}^o$) that satisfies the following property.

Property 2 (Maximal Score). *Hospital $o \in \mathcal{J}$ is evaluated with respect to weights $\boldsymbol{\mu}^o \in C^o$ that maximize its score S^o among all weights that satisfy Property 1.*

In what follows, unless explicitly stated otherwise, we assume that S^o represents this maximal score, that is, Stage 2 is always executed.

4. Properties of Optimal Hospital Scores

With the general methodology laid out, we now specialize to our hospital setting. We propose a collection of constraints to construct the convex cone C^o. One of our aims is to make the conic definition consistent with other aspects of the CMS hospital star ratings (Venkatesh et al. 2017b), so that our method can be easily substituted for the latent variable model and overall hospital summary score calculations, that is, steps 3 and 4 of the CMS methodology.

In Section 4.1, we describe how our approach accounts for hospitals reporting different subsets of measures, which differs from the CMS approach. Then, in Section 4.2, we provide cone constraints that model how the CMS weights different measure groups, including how the CMS adjusts these weights when groups are missing. We then provide new conic constraints that capture the idea that measures covering more people should be weighted more heavily, and lower bounds that ensure all measures are given positive weight.

After proving that the resulting cone is well defined, in Section 4.3, we develop a number of structural properties satisfied by optimal hospital scores. Most important among these is our main theorem showing that a hospital's score improves when it improves its performance. We also show how optimal scores change under affine data transformations and provide a dominance result showing that better-performing hospitals score higher.

4.1. Treatment of Missing Metrics

An important property, already satisfied by $(\mathbf{D}^o)$, is that the methodology should be flexible enough to eliminate performance measures from consideration when evaluating a hospital. Sometimes hospitals do not collect certain metrics, hospitals have data quality issues, or computed metrics are found to be statistically insignificant.

Property 3 (Missing Metrics). *Each hospital $o \in \mathcal{J}$ is evaluated using weights that redistribute across the subset of measures available and relative to other hospitals only on the subset of measures in common that are available.*

In contrast, the CMS methodology also considers only the measures available, but still estimates the latent variables using the same loadings in common across all hospitals in (4), that is, does not redistribute weight. When measures are missing within groups for a hospital o being evaluated, $(\mathbf{D}^o)$ eliminates the weight variable μ_i, that is, the model eliminates weights μ_i for any $i \in \mathcal{I} \setminus \mathcal{I}^o$. Furthermore, constraints (5b) consider only the subset of measures in common between hospital $j \in \mathcal{J}$ and the hospital o being evaluated, that is, measures $i \in \mathcal{I}^o \cap \mathcal{I}^j$. This is equivalent to assuming that $y_i^j = 0$ for any such missing measure $i \in \mathcal{I}^o \setminus \mathcal{I}^j$, which, in our implementation discussed later, will represent the national average; more generally, any imputed value could be used or the set of comparison hospitals $j \in \mathcal{J}$ restricted. Also, (5c) ignores measures not reported by hospital o.

4.2. Conic Constraints

We now develop a complete specification of the convex cone C^o.

4.2.1. Partitioning into Measure Groups.

One function of the conic constraints is to ensure that measure weights $\boldsymbol{\mu}$ partition into groups such that, when summed within each group, they satisfy predetermined totals. In particular, measures $\mathcal{I}$ are partitioned into K groups $\mathcal{G}_k \subseteq \mathcal{I}$ $\forall k \in \mathcal{K} \equiv \{1, 2, \ldots, K\}$, with group k weighted by $0 \le w_k \le 1$, where $\sum_{k \in \mathcal{K}} w_k = 1$. In practice, the CMS hospital star ratings currently have $K = 7$ groups, each weighted according to Table 2. Within each of these groups are many individual measures. Presumably, although stakeholders may agree on the weights that different categories should have, agreement on the relative importance of individual measures is more difficult to achieve. Recently, Rumball-Smith et al. (2018) proposed a user portal in which users can specify their own category weights. Our approach can easily accommodate this, using a web interface that re-solves $(\mathbf{D}^o)$ and $(\mathbf{P}^o)$ for a subset of hospitals selected by the user. We can also accommodate the case when there is only one group, in which case, our methodology will determine individual measure weights.

According to the CMS hospital star ratings methodology (Venkatesh et al. 2017b, p. 18), "if a hospital reports no measures for a given group, CMS considers that group to be 'missing.'" In this case, step 4 of the CMS methodology reallocates group weights among the remaining groups. For each hospital $o \in \mathcal{J}$, let

$$\mathcal{K}^o = \{k \in \mathcal{K} : |\mathcal{I}^o \cap \mathcal{G}_k| > 0\}$$

denote the set of groups for which hospital o has at least one measure reported. Then we reallocate the weights according to

$$w_k^o \equiv \frac{w_k}{\sum_{k' \in \mathcal{K}^o} w_{k'}} \forall o \in \mathcal{J}, k \in \mathcal{K}^o.$$

Although we mimic the CMS methodology in terms of how group weights are handled, we go a step further in requiring that the individual measure weights provide for a convex combination in (6).

Property 4 (Convexity Subject to Partition into Groups). *Each hospital is scored using a convex combination of individual measure weights, which partition into different groups, each having a predetermined total weight that is adjusted to accommodate missing groups.*

Table 2. CMS Hospital Star Ratings Groups and Their Weights as of December 2017

Group	Standard weight (%)
Mortality	22
Safety of Care	22
Readmission	22
Patient Experience	22
Effectiveness of Care	4
Timeliness of Care	4
Efficient Use of Medical Imaging	4

Let $\Gamma \geq 0$ be any scalar allowed to be free. Then, we can enforce these properties by adding the constraint

$$\sum_{i \in \mathcal{I}^o \cap \mathcal{G}_k} \mu_i = w_k^o \Gamma \quad \forall k \in \mathcal{K}^o$$

to the cone C^o for hospital o.

4.2.2. Other Conic Constraints. The next two sets of constraints have no analogue in the CMS methodology. Across all hospitals, for each $i \in \mathcal{I}$, let $N_i = \sum_{j \in \mathcal{J}} N_i^j$ denote the total number of patients impacted by measure i. Assume $N_i \leq N_{i+1}$ for all $i = 1, 2, \ldots, I-1$.

Property 5 (Patient Representation)**.** *Within a given measure group $k \in \mathcal{K}$, and given two measures $i_1, i_2 \in \mathcal{G}_k$ with $N_{i_1} < N_{i_2}$,*

$$\mu_{i_2} \geq \frac{N_{i_2}}{N_{i_1}} \mu_{i_1},$$

that is, the weight of measure i_2 exceeds or equals i_1 by a factor representing the relative number of patients impacted by i_2 versus i_1.

The idea is that, within a given measure group $k \in \mathcal{K}$, if a measure impacts three times as many patients as another measure, then the weight of the former should be at least three times the weight of the latter. The optimal weight can exceed three times (in this example) if it is feasible with respect to the cone C^o. We use national patient counts, rather than the hospital's local counts, because it favors comparability across hospitals. The current CMS approach ignores this concept.

To express these constraints formally, for each group of measures $k \in \mathcal{K}^o$, let $I_k^o = |\mathcal{I}^o \cap \mathcal{G}_k|$ denote the number of available measures in the group for hospital o. Next, let

$$i_k^{[1]}, i_k^{[2]}, \ldots, i_k^{[I_k^o]} \in \mathcal{I}^o \cap \mathcal{G}_k$$

denote the ordered elements of $\mathcal{G}_k$ such that

$$N_{i_k^{[m]}} \leq N_{i_k^{[m+1]}} \quad \forall m = 1, \ldots, I_k^o - 1.$$

(We suppress the dependence on hospital o in the notation $i_k^{[m]}$, as it will always be clear from context.) Then, in C^o, we have constraints

$$N_{i_k^{[m]}} \cdot \mu_{i_k^{[m+1]}} \geq N_{i_k^{[m+1]}} \cdot \mu_{i_k^{[m]}} \quad \forall k \in \mathcal{K}^o, m = 1, \ldots, I_k^o - 1. \tag{28}$$

Because of the ordering, we need constraints only for adjacent measures.

Last, we impose lower bounds on the measure weights to ensure that hospital performance for all reported measures is included in a hospital's score. Constraints (28) imply that adding a single positive lower bound constraint for the first measure in each group (with respect to the ordering above) forces all weights in the group to obey the same positive lower bound. We choose this positive lower bound conservatively, but other options are possible.

Let $k(i) \in \mathcal{K}$ denote the group that contains measure $i \in \mathcal{I}$, that is, $i \in \mathcal{G}_{k(i)}$. For each hospital $j \in \mathcal{J}$, let

$$f_i^j = \frac{N_i^j}{\sum_{i' \in \mathcal{I}^j \cap \mathcal{G}_{k(i)}} N_{i'}^j} \forall i \in \mathcal{I}^j$$

represent the fraction of total patients at hospital j represented in measures for group $k(i)$ that are included in measure i. Similarly, let

$$\bar{f}_i^j = \frac{N_i}{\sum_{i' \in \mathcal{I}^j \cap \mathcal{G}_{k(i)}} N_{i'}} \forall i \in \mathcal{I}^j$$

denote the analogous fraction except nationally. Then denote the positive lower bound associated with μ_i to be

$$\underline{\mu}_i^o \equiv \min_{i' \in \mathcal{I}^o \cap \mathcal{G}_{k(i)}} \min\left\{f_{i'}^o, \bar{f}_{i'}^o\right\} \cdot w_{k(i)} \quad \forall i \in \mathcal{I}^o.$$

The expression is the smallest of smallest fractions, multiplied by the group weight $w_{k(i)}$ to ensure the group weight scalings are satisfied. The inner minimum computes the smaller of the local or national fraction of patients represented. The outer minimum finds the smallest such fraction of patients across all the measures in the group $k(i)$ associated with measure i. These minimums combine to give a most conservative lower bound, which gives the optimization as much leeway as possible to optimize weights. The lower bound itself just represents the lowest fraction of patients impacted by any given measure in the group $k(i)$ of measures associated with measure i.

Although we are focusing on a particular lower bound, the general property at play is the following.

Property 6 (Measure Weights Are Lower Bounded)**.** *In hospital o's score, each individual measure $i \in \mathcal{I}^o$ has a weight that is lower bounded by a flexible constant that may vary across hospitals and measures.*

To add these lower bounds as conic constraints, simply add

$$\mu_i \geq \underline{\mu}_i^o \cdot \Gamma \quad \forall i \in \mathcal{I}^o,$$

where Γ is as defined above and represents the total $\boldsymbol{\mu}$ weight. From the arguments above, we actually need only to add these constraints for measure $i_k^{[1]}$ for each group $k \in \mathcal{K}^o$. However, we express the constraints as in the last display so as to make the formulation more generic for any choice of $\underline{\mu}_i^o$.

4.2.3. Summary of Conic Constraints. In summary, our definition of the convex cone C^o is as follows:

$$C^o = \Big\{ \boldsymbol{\mu} \in \mathbb{R}^{|\mathcal{I}^o|} : \exists \Gamma \in \mathbb{R} \text{ for which}$$

$$\sum_{i \in \mathcal{I}^o \bigcap \mathcal{G}_k} \mu_i = w_k^o \Gamma \quad \forall k \in \mathcal{K}^o, \tag{29}$$

$$N_{i_k^{[m]}} \cdot \mu_{i_k^{[m+1]}} \geq N_{i_k^{[m+1]}} \cdot \mu_{i_k^{[m]}} \quad \forall k \in \mathcal{K}^o, m = 1, \ldots, I_k^o - 1, \tag{30}$$

$$\mu_i \geq \underline{\mu}_i^o \cdot \Gamma \quad \forall i \in \mathcal{I}^o, \tag{31}$$

$$\Gamma \geq 0 \Big\}. \tag{32}$$

The next two lemmas ensure that this definition meets the basic requirements for strong duality to hold; see Ben-Tal and Nemirovski (2001).

Lemma 4. *C^o is a closed, pointed, polyhedral (convex) cone.*

Proof. If $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \in C^o$, then $\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2 \in C^o$ by setting $\Gamma = \Gamma_1 + \Gamma_2$. If $\boldsymbol{\mu} \in C^o$ and $a \geq 0$ is constant, then $a\boldsymbol{\mu} \in C^o$ by setting $\Gamma \leftarrow a\Gamma$. If $\boldsymbol{\mu} \in C^o$ and $-\boldsymbol{\mu} \in C^o$, then $\boldsymbol{\mu} = \mathbf{0}$ follows from $\boldsymbol{\mu} \geq \underline{\boldsymbol{\mu}}^o \Gamma \geq \mathbf{0}$ and $-\boldsymbol{\mu} \geq \underline{\boldsymbol{\mu}}^o \Gamma \geq \mathbf{0}$ implying $\boldsymbol{\mu} = \mathbf{0}$. Also, C^o is polyhedral and hence convex. □

For a general conic program, C^o needs a nonempty interior (Ben-Tal and Nemirovski 2001) and may contain other nonlinear constraints. However, because in the end we are working with a linear program, it suffices to show that C^o has a strictly positive element, which ensures feasibility from Lemma 1. Trivially, $\mathbf{0} \in C^o$.

Lemma 5. *C^o has a strictly positive element.*

Proof. We construct a $\boldsymbol{\mu} \in C^o$ such that (29) and (30) are tight. Choose any $\Gamma > 0$, thus satisfying (32). To simplify notation, let $\mathcal{N}_k^o \equiv \sum_{m=1}^{I_k^o} N_{i_k[m]}$ denote the total number of patients nationally impacted by measures in group k reported by hospital o. Set

$$\mu_{i_k^{[1]}} = \frac{N_{i_k^{[1]}}}{\mathcal{N}_k^o} \cdot w_k \Gamma > 0.$$

Then, (30) implies

$$\mu_{i_k^{[2]}} \geq \frac{N_{i_k^{[2]}}}{N_{i_k^{[1]}}} \cdot \mu_{i_k^{[1]}} = \frac{N_{i_k^{[2]}}}{\mathcal{N}_k^o} \cdot w_k \Gamma,$$

again positive. Suppose this last display is met with equality; then, recursively, we have

$$\mu_{i_k^{[m]}} = \frac{N_{i_k^{[m]}}}{\mathcal{N}_k^o} \cdot w_k \Gamma \quad \forall m = 1, \ldots, I_k^o, \tag{33}$$

where each is positive. Thus,

$$\sum_{m=1}^{I_k^o} \mu_{i_k^{[m]}} = \frac{\sum_{m=1}^{I_k^o} N_{i_k^{[m]}}}{\mathcal{N}_k^o} \cdot w_k \Gamma = w_k \Gamma,$$

so that constraints (29) are satisfied. Because

$$\underline{\mu}_{i_k^{[m]}}^o = \min_{i' \in \mathcal{I}^o \cap \mathcal{G}_k} \min\{f_{i'}^o, \bar{f}_{i'}^o\} \cdot w_k \leq \frac{N_{i_k^{[m]}}}{\mathcal{N}_k^o} \cdot w_k,$$

using the upper bound $f_{i_k^{[m]}}^o$ on the double minimum, the lower bound constraint (31) is satisfied, that is,

$$\mu_{i_k^{[m]}} \geq \underline{\mu}_{i_k^{[m]}}^o \Gamma. \quad \square$$

This proof reveals in (33) the largest lower bounds $\underline{\mu}_i^o$ that can be employed in constraints (31) while maintaining feasibility.

4.3. Additional Properties of Optima

Having specified the convex cone C^o, we now explore various properties satisfied by optimal hospital scores.

4.3.1. Affine Data Transformations. As discussed in Portela et al. (2004), where it is implicitly assumed that $C^o = \mathbb{R}_+^{|\mathcal{I}^o|}$ and $\mathcal{I}^j = \mathcal{I}\ \forall j \in \mathcal{J}$, an important property of $(\mathbf{P}^o)$ is that optimal solutions are invariant to multiplications and shifts of the input data. This makes the formulation applicable to settings, such as ours, having a heterogeneous collection of inputs. We first extend this result to the conic setting and show how dual optimal solutions are impacted by data transformations. This will allow us to assess the impact of data transformations on hospital scores.

Proposition 3 (Affine Data Transformation)**.** *Assume $\mathcal{I}^j = \mathcal{I}$ for all $j \in \mathcal{J}$. Suppose the input data $\{y_i^j\}_{\forall j \in \mathcal{J}, i \in \mathcal{I}^j}$ are multiplied and shifted, so that*

$$\tilde{y}_i^j \leftarrow y_i^j \gamma_i + \alpha_i \quad \forall j \in \mathcal{J}, i \in \mathcal{I}^j,$$

where $\gamma_i > 0$ and $\alpha_i \in \mathbb{R}$ for all $i \in \mathcal{I}^o$. Then, an optimal primal–dual solution pair $(\beta^o, \boldsymbol{\lambda}^o, \boldsymbol{s}^o), (\theta^o, \boldsymbol{\mu}^o)$ before data transformation maps into an optimal solution

$$\tilde{\beta}^o \leftarrow \beta^o,$$
$$\tilde{\lambda}^o \leftarrow \lambda^o,$$
$$\tilde{s}_i^o \leftarrow \gamma_i s_i^o \quad \forall i \in \mathcal{I}^o,$$
$$\tilde{\theta}^o \leftarrow \theta^o + \sum_{i \in \mathcal{I}^o} (\alpha_i / \gamma_i) \mu_i^o,$$
$$\tilde{\mu}_i^o \leftarrow \mu_i^o / \gamma_i \quad \forall i \in \mathcal{I}^o$$

under the transformed data.

Proof. First, we check that the proposed optimal solution is a feasible solution to $(\mathbf{P}^o)$ after data transformation. For all $i \in \mathcal{I}^o$,

$$\bar{\tilde{y}}_i \leftarrow \max_{\{j \in \mathcal{J} : i \in \mathcal{I}^j\}} \left(y_i^j \gamma_i + \alpha_i \right) = \bar{y}_i \gamma_i + \alpha_i.$$

Therefore, under the proposed optimal solution, the left-hand side of (7b) for any given $i \in \mathcal{I}^o$ becomes

$$\begin{aligned}&\left(y_i^o\gamma_i + \alpha_i\right) + \beta^o \cdot \left(\left(\overline{y}_i\gamma_i + \alpha_i\right) - \left(y_i^o\gamma_i + \alpha_i\right)\right) + \gamma_i s_i^o \\ &\quad = \left(y_i^o\gamma_i + \alpha_i\right) + \gamma_i\beta^o \cdot \left(\overline{y}_i - y_i^o\right) + \gamma_i s_i^o,\end{aligned}$$

whereas the right-hand side becomes

$$\begin{aligned}&\sum_{\{j\in\mathcal{J}: i\in\mathcal{I}^j\}} \left(y_i^j\gamma_i + \alpha_i\right)\lambda_j^o \\ &\quad = \gamma_i \sum_{\{j\in\mathcal{J}: i\in\mathcal{I}^j\}} y_i^j\lambda_j^o + \alpha_i \sum_{\{j\in\mathcal{J}: i\in\mathcal{I}^j\}} \lambda_j^o \\ &\quad = \gamma_i \sum_{\{j\in\mathcal{J}: i\in\mathcal{I}^j\}} y_i^j\lambda_j^o + \alpha_i\end{aligned}$$

from (7c), because $\mathcal{I}^j = \mathcal{I}\ \forall j \in \mathcal{J}$ implies

$$\sum_{\{j\in\mathcal{J}: i\in\mathcal{I}^j\}} \lambda_j^o = \sum_{j\in\mathcal{J}} \lambda_j^o = 1.$$

Combining these together, collecting terms, and dividing through by $\gamma_i > 0$, feasibility requires

$$y_i^o + \beta^o \cdot \left(\overline{y}_i - y_i^o\right) + s_i^o = \sum_{\{j\in\mathcal{J}: i\in\mathcal{I}^j\}} y_i^j\lambda_j^o.$$

This equation is satisfied because $(\beta^o, \boldsymbol{\lambda}^o)$ is an optimal, and hence feasible, solution before data transformation. Furthermore, because $C_\star^o$ is a cone and $\gamma_i > o$ for all i, we have that $\mathbf{s}^o \in C_\star^o$ implies $\gamma\mathbf{s}^o \in C_\star^o$. Thus, the transformed solution is a feasible solution to $(\mathbf{P}^o)$ after the data are transformed.

Next, we show that the proposed dual solution is a feasible solution to $(\mathbf{D}^o)$. Indeed, after data transformation, the left-hand side of (7c) becomes

$$\begin{aligned}\sum_{i\in\mathcal{I}^o}\left(\left(\overline{y}_i\gamma_i + \alpha_i\right) - \left(y_i^o\gamma_i + \alpha_i\right)\right)\tilde{\mu}_i^o &= \sum_{i\in\mathcal{I}^o}\gamma_i\left(\overline{y}_i - y_i^o\right)\tilde{\mu}_i^o \\ &= \sum_{i\in\mathcal{I}^o}\left(\overline{y}_i - y_i^o\right)\mu_i^o \\ &\geq 1.\end{aligned}$$

To verify (7b), for any given $j \in \mathcal{J}$, we obtain

$$\begin{aligned}\sum_{i\in\mathcal{I}^o}\tilde{y}_i^j\tilde{\mu}_i^o - \tilde{\theta}^o &= \sum_{i\in\mathcal{I}^o}\left(y_i^j\gamma_i + \alpha_i\right)\mu_i^o/\gamma_i \\ &\quad - \left(\theta^o + \sum_{i\in\mathcal{I}^o}(\alpha_i/\gamma_i)\mu_i^o\right) \\ &= \sum_{i\in\mathcal{I}^o} y_i^j\mu_i^o - \theta^o \\ &\leq 0.\end{aligned}$$

Also, $\gamma_i > 0\ \forall i \in \mathcal{I}^o$ implies $\tilde{\boldsymbol{\mu}}^o \in C^o$ because C^o is a cone.

Last, from the last derivation above, when $j = o$, the objective function

$$\tilde{\theta}^o - \sum_{i\in\mathcal{I}^o}\tilde{y}_i^o\tilde{\mu}_i^o = \theta^o - \sum_{i\in\mathcal{I}^o} y_i^o\mu_i^o = \beta^o = \tilde{\beta}^o.$$

Hence, there is no duality gap, and so the proposed transformed solution is an optimal primal–dual pair. □

Corollary 1. *Under the data transformation of Proposition 3, the new score for hospital $o \in \mathcal{J}$ equals*

$$\tilde{S}^o \leftarrow \frac{S^o\Gamma^o + \sum_{i\in\mathcal{I}^o}\alpha_i\left(\frac{\mu_i^o}{\gamma_i}\right)}{\sum_{i\in\mathcal{I}^o}\left(\mu_i^o/\gamma_i\right)}. \tag{34}$$

In particular, if there exist constants α and γ such that $\alpha_i = \alpha$ and $\gamma_i = \gamma > 0$ for all $i \in \mathcal{I}^o$, then

$$\tilde{S}^o \leftarrow \gamma S^o + \alpha.$$

Proof. Plugging the optimal solution after transformation into the score expression (6), we obtain

$$\begin{aligned}\tilde{S}^o &\leftarrow \frac{\sum_{i\in\mathcal{I}^o}\tilde{y}_i^o\tilde{\mu}_i^o}{\sum_{i\in\mathcal{I}^o}\tilde{\mu}_i^o} \\ &= \frac{\sum_{i\in\mathcal{I}^o}\left(y_i^o\gamma_i + \alpha_i\right)\left(\mu_i^o/\gamma_i\right)}{\sum_{i\in\mathcal{I}^o}\left(\mu_i^o/\gamma_i\right)} \\ &= \frac{\sum_{i\in\mathcal{I}^o} y_i^o\mu_i^o + \sum_{i\in\mathcal{I}^o}\alpha_i\left(\mu_i^o/\gamma_i\right)}{\sum_{i\in\mathcal{I}^o}\left(\mu_i^o/\gamma_i\right)},\end{aligned}$$

which equals (34). Now, setting $\alpha_i = \alpha$ and $\gamma_i = \gamma$ $\forall i \in \mathcal{I}^o$,

$$\begin{aligned}\tilde{S}^o &\leftarrow \frac{S^o\Gamma^o + (\alpha/\gamma)\sum_{i\in\mathcal{I}^o}\mu_i^o}{(1/\gamma)\sum_{i\in\mathcal{I}^o}\mu_i^o} \\ &= \frac{S^o\Gamma^o + (\alpha/\gamma)\Gamma^o}{(1/\gamma)\Gamma^o} \\ &= \gamma S^o + \alpha. \quad \square\end{aligned}$$

This analysis shows that whereas efficiency is invariant with respect to affine data transformations, the scores in general are not.

Property 7 (Scores Depend on Data Scales). *Hospital scores and rankings depend on the scales used for performance metrics.*

When all measures are transformed by the same shift (α) and multiplier (γ), the scores for all hospitals are shifted to $\gamma S^o + \alpha$, and so hospital rankings with respect to S^o are preserved. In general, however, an affine data transformation can change the hospital rank order. Thus, in applying our approach, to avoid bias due to measurement scales, it is important to take all measures to be on the same scale. Henceforth, we will assume that all measure data have been standardized.

Assumption 2 (Standardized Data)**.** *All measures have a mean of zero and a standard deviation of one.*

This is achieved by setting $\gamma_i = (1/\sigma_i)$ and $\alpha_i = (\mathsf{E}[y_i]/\sigma_i)$ $\forall i \in \mathcal{I}^o$, where $\mathsf{E}[y_i]$ and σ_i are the mean and standard deviation, respectively, of the underlying data. This assumption gives hospital scores (6) a meaningful interpretation as a convex combination of z-scores. Thus, a hospital score equal to zero reflects an average hospital, and other scores reflect standard deviations from the mean. However, this is only a heuristic interpretation because the collection of hospital scores we obtain need not exhibit a standard deviation of one. This could be enforced as a postprocessing step.

4.3.2. Stability and Dominance Properties. The next property is critical because it guarantees hospitals that improve will be rewarded.

Property 8 (Improving Hospitals Improve Their Scores)**.** *Independently of other hospitals, holding $\overline{\boldsymbol{y}}$ and C^o constant, if hospital $o \in \mathcal{I}$ does not perform worse along any measure, then its score S^o does not decrease.*

Although this may seem like an obvious property that should hold, as shown earlier in Section 2.3, the current CMS latent variable model approach can violate this property. In our setting, because the measure weights are not fixed a priori, and instead computed separately for each hospital, this property reassures that hospital incentives are not distorted by the underlying computations. The assumption that $\overline{\boldsymbol{y}}$ be held constant is likely to be true, or nearly true, with winsorized data, or in practice, $\overline{\boldsymbol{y}}$ could be set equal to a fixed ideal target (say, the winsorization cutoff). Another assumption is that the cone C^o does not change, that is, the nationally aggregated denominators, or their ratios, remain constant. Hospitals that report measures are likely to continue reporting them, so although national totals are expected to change over time, one would expect dramatic changes in their ratios to be rare. The next theorem formally states this result.

Theorem 1. *Let $(\theta^*, \boldsymbol{\mu}^*)$ be an optimal solution to $(\mathbf{D}^o)$ inducing maximal hospital score S^o, and without loss of optimality, assume $\sum_{i\in\mathcal{I}^o}(\overline{y}_i - y_i^o)\mu_i^* = 1$. Suppose,* for all *$j \in \mathcal{I} \setminus \{o\}$, hospital j's performance changes from y_i^j to $\hat{y}_i^j$ $\forall i \in \mathcal{I}^j$, and hospital o's performance improves to*

$$\hat{y}_i^o = y_i^o + \alpha_i < \overline{y}_i \quad \forall i \in \mathcal{I}^o,$$

for $\alpha_i \geq 0 \, \forall i \in \mathcal{I}^o$, assuming $\overline{\boldsymbol{y}}$ does not change under $\hat{\boldsymbol{y}}$. Then, after resolving $(\mathbf{D}^o)$ with new data $\hat{y}$, the new score $\hat{S}^o$ satisfies

$$\hat{S}^o \geq S^o.$$

This result holds for any nonnegative change α_i $\forall i \in \mathcal{I}^o$ in a hospital's measure inputs to the model, but this may represent either relative or absolute improvement, depending on how the model is implemented. If Assumption 2 is applied after the improvement period, so that the performance data $\hat{\boldsymbol{y}}$ represent z-scores relative to the updated means and variances of the raw performance data (not represented here), then improvement in measures is relative to all other hospitals. This is how the CMS star ratings are currently implemented. Without any such transformation, improvement in measures would be in absolute, rather than relative, terms. In any case, the theorem is oblivious to whether, or how, the raw performance data are transformed to create model inputs.

To prove this and the next property, we will make use of the following result.

Lemma 6. *Given a feasible solution $(\theta', \boldsymbol{\mu}')$ to $(\mathbf{D}^o)$, there exists an optimal solution $(\theta^*, \boldsymbol{\mu}^*)$ that satisfies*

$$\sum_{i\in\mathcal{I}^o} y_i^o \mu_i^* \geq \sum_{i\in\mathcal{I}^o} y_i^o \mu_i' \tag{35}$$

and

$$S^{*o} \equiv \frac{\sum_{i\in\mathcal{I}^o} y_i^o \mu_i^*}{\Gamma^*} \geq \frac{\sum_{i\in\mathcal{I}^o} y_i^o \mu_i'}{\Gamma'} \equiv S'^o, \tag{36}$$

where $\Gamma^* = \sum_{i\in\mathcal{I}^o} \mu_i^*$ and $\Gamma' = \sum_{i\in\mathcal{I}^o} \mu_i'$.

Proof. Let $\Delta' = \theta' - \sum_{i\in\mathcal{I}^o} y_i^o \mu_i'$. Consider

$$\begin{aligned} f(\Delta) = \max_{\theta\in\mathbb{R},\boldsymbol{\mu}\in\mathbb{R}^{|\mathcal{I}^o|}} \; & \sum_{i\in\mathcal{I}^o} y_i^o \mu_i \\ \text{s.t.} \; & \sum_{i\in\mathcal{I}^o\cap\mathcal{I}^j} y_i^j \mu_i \leq \theta \quad \forall j \in \mathcal{I}, \\ & \sum_{i\in\mathcal{I}^o} (\overline{y}_i - y_i^o)\mu_i \geq 1, \\ & \boldsymbol{\mu} \in C^o, \\ & \theta - \sum_{i\in\mathcal{I}^o} y_i^o \mu_i \leq \Delta. \end{aligned}$$

For $\Delta = \Delta'$, $(\theta', \boldsymbol{\mu}')$ is a feasible solution. Furthermore, $\beta^* \leq \Delta'$, because $(\mathbf{D}^o)$ minimizes. Because decreasing Δ increases the lower bound on the optimal objective value through the last constraint above, $f(\Delta)$ is non-decreasing as Δ decreases. Thus,

$$\sum_{i\in\mathcal{I}^o} y_i^o \mu_i' \leq f(\Delta') \leq f(\beta^*) = \sum_{i\in\mathcal{I}^o} y_i^o \mu_i^*,$$

which yields (35).

To verify (36), consider a parametric version of the Stage 2 program (27), where the constraint (27d) is written as an inequality:

$$\begin{aligned}
\mathbf{g}(\boldsymbol{\Delta}) = \max_{\eta\in\mathbb{R},\rho\in\mathbb{R}^{|\mathcal{I}^o|},\phi\in\mathbb{R}} & \sum_{i\in\mathcal{I}^o} y_i^o\rho_i \\
\text{s.t.} \quad & \sum_{i\in\mathcal{I}^o\cap\mathcal{I}^j} y_i^j\rho_i \le \eta \quad \forall j\in\mathcal{J}, \\
& \sum_{i\in\mathcal{I}^o}(\overline{y}_i - y_i^o)\rho_i \ge \phi, \\
& \phi \ge 0, \\
& \sum_{i\in\mathcal{I}^o}\rho_i = 1, \\
& \boldsymbol{\rho}\in C^o, \\
& \eta - \sum_{i\in\mathcal{I}^o} y_i^o\rho_i \le \Delta\phi.
\end{aligned} \tag{37}$$

The mapping (24)–(26) applied to $(\theta', \boldsymbol{\mu}')$ with $\Gamma' \equiv \sum_{i\in\mathcal{I}^o}\mu_i'$ produces a feasible solution to this program with $\Delta = \Delta'$. Because $\phi \ge 0$, $g(\Delta)$ is non-decreasing as Δ decreases. Thus, we have

$$\begin{aligned}
\frac{\sum_{i\in\mathcal{I}^o} y_i^o\mu_i'}{\Gamma'} &= \sum_{i\in\mathcal{I}^o} y_i^o\rho_i' \le g(\Delta') \le g(\beta^*) \\
&= \sum_{i\in\mathcal{I}^o} y_i^o\rho_i^* = \frac{\sum_{i\in\mathcal{I}^o} y_i^o\mu_i^*}{\Gamma^*},
\end{aligned}$$

which yields (36). □

This result ensures an improving (unscaled and scaled) score starting from a feasible solution to $(\mathbf{D}^o)$.

We can now prove Theorem 1.

Proof of Theorem 1. Because $\boldsymbol{\mu} = 0$ is infeasible in (5c), we must have $\boldsymbol{\mu}^* \ne 0$. Therefore, $\boldsymbol{\mu}^* \ge 0$ and $0 \le \alpha_i < \overline{y}_i - y_i^o\ \forall i\in\mathcal{I}^o$ implies

$$0 \le \sum_{i\in\mathcal{I}^o}\alpha_i\mu_i^* < \sum_{i\in\mathcal{I}^o}(\overline{y}_i - y_i^o)\mu_i^* = 1.$$

To transform $(\theta^*, \boldsymbol{\mu}^*)$ into a feasible solution under the new data $\hat{\boldsymbol{y}}$, let

$$\gamma \equiv \frac{1}{1 - \sum_{i\in\mathcal{I}^o}\alpha_i\mu_i^*}$$

and consider

$$\mu_i' = \gamma\mu_i^* \quad \forall i\in\mathcal{I}^o.$$

Because $0<\gamma<\infty$, $\boldsymbol{\mu}^*\in C^o$ implies $\boldsymbol{\mu}'\in C^o$. Furthermore,

$$\begin{aligned}
\sum_{i\in\mathcal{I}^o}(\overline{y}_i - \hat{y}_i^o)\mu_i' &= \gamma\sum_{i\in\mathcal{I}^o}(\overline{y}_i - y_i^o)\mu_i^* - \gamma\sum_{i\in\mathcal{I}^o}\alpha_i\mu_i^* \\
&= \gamma - \gamma\sum_{i\in\mathcal{I}^o}\alpha_i\mu_i^* \\
&= 1.
\end{aligned}$$

Next, let

$$\theta' = \max_{j\in\mathcal{J}}\sum_{i\in\mathcal{I}^o\cap\mathcal{I}^j}\hat{y}_i^j\mu_i'.$$

Thus, by construction, $(\theta', \boldsymbol{\mu}')$ is a feasible solution to $(\mathbf{D}^o)$ under new data $\hat{\boldsymbol{y}}$. Using $\Gamma' \equiv \sum_{i\in\mathcal{I}^o}\mu_i' = \gamma\Gamma^*$, the associated score is

$$\begin{aligned}
S'^o \equiv \frac{\sum_{i\in\mathcal{I}^o}\hat{y}_i^o\mu_i'}{\Gamma'} &= \frac{\sum_{i\in\mathcal{I}^o}\hat{y}_i^o\mu_i^*}{\Gamma^*} \\
&= \frac{\sum_{i\in\mathcal{I}^o} y_i^o\mu_i^*}{\Gamma^*} + \frac{\sum_{i\in\mathcal{I}^o}\alpha_i\mu_i^*}{\Gamma^*} \\
&= S^o + \frac{\sum_{i\in\mathcal{I}^o}\alpha_i\mu_i^*}{\Gamma^*} \\
&\ge S^o.
\end{aligned}$$

Now apply Lemma 6, showing that there exists an optimal score $\hat{S}^o$ at least equal to S'^o. □

In Appendix A, we specialize to the case in which improvement occurs in the direction between the ideal hospital and hospital o. Not only can hospital o improve in this direction, but hospitals on the efficient frontier, and even the ideal hospital $\overline{\boldsymbol{y}}$, may improve in this direction as well. In this special case, though not realistic, we can derive the new score and efficiency for hospital o in closed form. Interestingly, the new score does not depend on the magnitude of improvement for reference hospitals or the ideal hospital.

The next property that we show indicates how the scores of comparable hospitals relate to each other.

Property 9 (Dominance). *A hospital that performs at least as well as another comparable hospital across all metrics scores at least as high.*

Although this property may seem obvious, the current CMS latent variable model approach may violate it if loadings are allowed to be negative. In our setting, it is not obvious that this property would hold because each hospital is given its own set of measure weights, and the resulting two convex combinations may place weights differently. The comparability assumption requires that the two hospitals have the same set of measures available and that the optimal measure weights for the weaker hospital are feasible for the stronger hospital with respect to the convex cone C^o. Formally, we have the following result.

Proposition 4. *Consider $o, o' \in \mathcal{J}$ such that $\mathcal{I}^o = \mathcal{I}^{o'}$, $y_i^o \ge y_i^{o'}\ \forall i\in\mathcal{I}^o$, $y_i^{o'} < \overline{y}_i\ \forall i\in\mathcal{I}^o$, and $\boldsymbol{\mu}^{o'}\in C^o$. Then, there exists an optimal solution $(\theta^o, \boldsymbol{\mu}^o)$ to $(\mathbf{D}^o)$ such that*

$$S^o \ge S^{o'},$$

where $S^{o'}$ is the maximal score for hospital o'.

Proof. Suppose $y_i^o = y_i^{o'} + \alpha_i$ for $\alpha_i \geq 0\ \forall i \in \mathcal{I}^o$. We construct a feasible solution to ($\mathbf{D}^o$) from an optimal solution $(\theta^{o'}, \boldsymbol{\mu}^{o'})$ to ($\mathbf{D}^{\mathbf{o}'}$). Let

$$\gamma \equiv \frac{1}{1 - \sum_{i\in\mathcal{I}^o} \alpha_i \mu_i^{o'}},$$

and consider

$$\mu_i' = \gamma \mu_i^{o'} \quad \forall i \in \mathcal{I}^o.$$

Following the same logic as in the proof of Theorem 1, we conclude that $\boldsymbol{\mu}^{o'} \in C^o$ implies $\boldsymbol{\mu}' \in C^o$. To verify (5c), observe

$$\begin{aligned}\sum_{i\in\mathcal{I}^o} (\bar{y}_i - y_i^o)\mu_i' &= \gamma \cdot \left(\sum_{i\in\mathcal{I}^o} (\bar{y}_i - y_i^{o'})\mu_i^{o'} - \sum_{i\in\mathcal{I}^o} \alpha_i \mu_i^{o'}\right) \\ &\geq \gamma \cdot \left(1 - \sum_{i\in\mathcal{I}^o} \alpha_i \mu_i^{o'}\right) = 1.\end{aligned}$$

Setting $\theta' = \gamma\theta^{o'}$, we have, for all $j \in \mathcal{J}$,

$$\theta' - \sum_{i\in\mathcal{I}^o\cap\mathcal{I}^j} y_i^j \mu_i' = \gamma \cdot \left(\theta^{o'} - \sum_{i\in\mathcal{I}^o\cap\mathcal{I}^j} y_i^j \mu_i^{o'}\right) \geq 0.$$

Thus, $(\theta', \boldsymbol{\mu}')$ is a feasible solution to ($\mathbf{D}^o$). Letting

$$\Gamma' \equiv \sum_{i\in\mathcal{I}^o} \mu_i' = \gamma \sum_{i\in\mathcal{I}^o} \mu_i^{o'} = \gamma\Gamma^{o'},$$

where $\Gamma^{o'} \equiv \sum_{i\in\mathcal{I}^o} \mu_i^{o'}$, the score S'^o associated with this solution is

$$\begin{aligned}S'^o \equiv \frac{\sum_{i\in\mathcal{I}^o} y_i^o \mu_i'}{\Gamma'} &= \frac{\sum_{i\in\mathcal{I}^o} y_i^o \mu_i^{o'}}{\Gamma^{o'}} \\ &= \frac{\sum_{i\in\mathcal{I}^o} y_i^{o'} \mu_i^{o'}}{\Gamma^{o'}} + \frac{\sum_{i\in\mathcal{I}^o} \alpha_i \mu_i^{o'}}{\Gamma^{o'}} \\ &= S^{o'} + \frac{\sum_{i\in\mathcal{I}^o} \alpha_i \mu_i^{o'}}{\Gamma^{o'}} \\ &\geq S^{o'}.\end{aligned}$$

Now invoke Lemma 6 with feasible solution $(\theta', \boldsymbol{\mu}')$ to ($\mathbf{D}^o$) to conclude that there exists an optimal solution $(\theta^*, \boldsymbol{\mu}^*)$ to ($\mathbf{D}^o$) such that the associated score $S^{*o} \geq S'^o$. Thus, $S^{*o} \geq S^{o'}$, and we can set $S^o = S^{*o}$. □

5. Numerical Results

We now compare the output of our proposed efficient frontier approach with the output of the CMS's latent variable model. In Appendix C, we provide details of our numerical setup, including information about data sources and the methodology used by the CMS. As described there, to make a fair comparison, we implement the CMS approach as closely as possible, substituting our efficient frontier approach for the LVM. We do not compare hospitals based on number of stars, for which the CMS uses a *k*-means clustering algorithm. Instead, we compare them based on the quintiles their overall scores fall into, with 1 representing the lowest-performing quintile and 5 representing the highest. Because quintiles are equal-sized groups, rather than variable-sized groups as given by clustering, this allows us to more fairly compare how hospitals rank between the two methods.

Much of our analysis reports on how performance compares and contrasts across hospitals of different sizes. We use the same partitioning of hospital sizes described in Shwartz et al. (2008): “large” hospitals have 400 or more beds, “medium” hospitals have fewer than 400 but at least 100 beds, and “small” hospitals have fewer than 100 beds. This partition pertains to acute care hospitals. As a fourth category, we consider “critical access” hospitals, which are rural hospitals having no more than 25 beds that qualify for funding under the Balanced Budget Act of 1997. They can be thought of as extra small hospitals. To determine hospital size for acute care hospitals, we used the impact file published by the CMS as part of the fiscal year 2018 IPPS (Inpatient Prospective Payment System) Final Rule.[3]

Our empirical results provide a number of insights:

1. In aggregate, the efficient frontier approach and the LVM approach generate similar results. Both produce bell-shaped curves for overall hospital scores, and they exhibit similar trends in group scores across hospital sizes. Similar percentages of hospitals change quintiles from year to year.
2. Larger hospitals tend to rank lower than smaller hospitals under both approaches. The reasons for this appear to be independent of the LVM or efficient frontier approach.
3. Many hospitals rank in different quintiles under the efficient frontier approach than the LVM approach, with approximately half of critical access hospitals changing by at least one quintile up or down. The differences are largely driven by the efficient frontier approach (a) generating more dispersed measure weights, (b) shifting group weight from missing measures onto measures a hospital reports (Property 3), and (c) forcing measure weights to meet the patient representation constraints (Property 5).

In Section 5.1, we discuss how hospital and group summary scores compare in aggregate across hospitals. Then, in Section 5.2, we discuss how individual measures change across hospital sizes and are weighted differently by the two approaches. Last, in Section 5.3, we discuss how hospital rankings change from year to year and across the two approaches.

5.1. Hospital and Group Summary Scores

Figure 6 compares the overall distribution of hospital summary scores produced by the CMS latent variable model approach versus our proposed efficient frontier approach. The CMS distribution has greater variance with symmetrically thicker tails. The CMS approach computes hospital scores to be a convex combination of group scores, each of which the LVM requires to be normally distributed with mean zero and variance one. Thus, hospital summary scores are normally distributed (although here we display only published hospitals). In contrast, we compute hospital scores as a convex combination of measures standardized to have mean zero and variance one, but not necessarily normally distributed. Nonetheless our approach produces a bell-shaped distribution with mean near zero. Thus, although hospital summary scores are not directly comparable across the methods, they reside on roughly the same scale.

To compare the score distributions by hospital size and type, in Table 3 we show the percentage of cases in each quintile by approach. As stated earlier, the table shows that larger hospitals tend to rank lower than smaller hospitals under both approaches. This effect is driven by three factors, which act in concert:

1. Larger hospitals tend to be evaluated worse than smaller hospitals on many important measures.
2. Smaller (larger) hospitals tend to report fewer (more) measures, and thus carry fewer (more) chances to drive their overall scores down.
3. All hospitals, regardless of size, are lumped together when ranking hospitals.

Notably, this effect is the exact opposite of what is reported in Shwartz et al. (2008), where it is shown that larger hospitals tend to be ranked in the top deciles. One potential explanation is that the measures used there are all process, not outcome, measures. Outcome measures are risk adjusted to equalize differences in patient characteristics between hospitals. However, Sosunov et al. (2016) show that a common risk-adjustment approach used by the CMS results in higher-volume (larger) hospitals ranking lower, even if they otherwise have equal performance. Hota et al. (2019) discuss other issues with underlying measures that bias them against larger hospitals, including the absence of socioeconomic factors in the CMS's risk adjustment for readmission. These factors exist independently of either the LVM or the efficient frontier approach, but because they affect the inputs, they have a significant impact on the scores and rankings that result.

In Table 4, we show how average group scores compare between the approaches, by hospital size. For our approach, a hospital o's score for group $k \in \mathcal{K}^o$ is simply computed as a convex combination of measures within the group, that is,

$$\sum_{i \in \mathcal{I}^o \cap \mathcal{G}_k} y_i^o \mu_i^o / \Gamma_k^o,$$

where $\Gamma_k^o = \sum_{i \in \mathcal{I}^o \cap \mathcal{G}_k} \mu_i^o$. Similarly to hospital summary scores, group scores are not directly comparable across the methods. However, the proposed approach maintains the same trends by hospital size as the CMS

Figure 6. Distributions of CMS (Gray Bars with Black Border) and Proposed (Yellow) Hospital Scores for December 2017

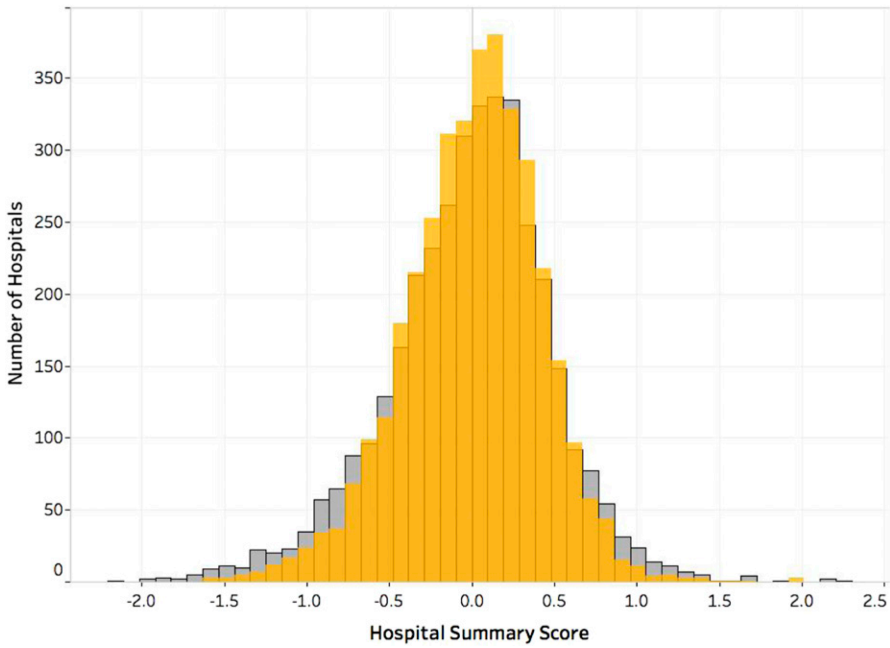


Table 3. Percentage of Cases in Each Quintile by Hospital Size and Approach

	Critical access		Small		Medium		Large	
Quintile	CMS (%)	Proposed (%)	CMS (%)	Proposed (%)	CMS (%)	Proposed (%)	CMS (%)	Proposed (%)
1	3.1	8.1	9.6	9.0	30.1	28.7	32.6	32.1
2	15.7	14.1	19.9	20.0	22.0	21.8	18.9	22.9
3	29.0	20.7	22.7	21.4	15.4	18.8	17.5	18.3
4	30.4	26.7	24.3	24.9	15.4	16.6	10.5	10.8
5	21.8	30.3	23.5	24.7	17.1	14.1	20.5	15.9

approach. For example, scores for Mortality and Efficient Use of Medical Imaging increase with hospital size, where Readmission generally decreases with hospital size. The main instance in which the CMS's group score moves in a different direction than ours is for Effectiveness of Care comparing critical access, small, and medium-sized hospitals, where for the CMS they go up, but for us they go down. Otherwise, our group score averages are directionally equivalent to the CMS group score averages.

5.2. Weights and Performance on Individual Measures

In Tables D.1–D.7 of Appendix D, we provide detailed calculations for each hospital measure by group. The hospital measures listed in Appendix D correspond with those defined in Appendix B, but are expressed with prefix "std_" to designate they are standardized and with underscores "_" instead of dashes "–" to match output from SAS. The tables show how performance on individual measures is evaluated differently depending on hospital size, as discussed in the previous section.

For many groups, the measure weights given by the empirical Bayes estimate (4) concentrate mostly on one measure. This effect is strongest for Readmission and Safety of Care, but also holds true for Timeliness of Care and Efficient Use of Medical Imaging. It is likely an artifact of there being multiple latent factors, but only one is allowed. The measure weights are more spread out for Mortality, Patient Experience, and Effectiveness of Care. For Readmission, nearly all of the weight is placed on the measure std_READM_30_HOSP_WIDE, and for Safety of Care, nearly all of the weight is placed on the measure std_PSI_90_SAFETY. This means that hospitals that do poorly (well) on these measures are significantly penalized (rewarded), even if they do well (poorly) on all other measures in the group. Our approach tends to spread the weights more evenly, mitigating this effect. One exception is Effectiveness of Care, in which our approach concentrates weight on OP-22 (emergency department left without being seen) because it impacts a disproportionate number of patients relative to the other measures.

Another observation is that our measure weights for Patient Experience do not vary. This is because the national aggregated denominator of each measure in this case equals the number of respondents to the patient survey. Hence, it is the same for all measures

Table 4. CMS vs. Proposed Group Scores by Hospital Size (Averages with Standard Errors in Parentheses)

Group	Critical access		Small		Medium		Large	
CMS								
Mortality	−0.092	(0.013)	−0.077	(0.019)	0.022	(0.022)	0.313	(0.060)
Safety of Care	−0.008	(0.003)	0.135	(0.020)	−0.072	(0.027)	−0.237	(0.070)
Readmission	0.061	(0.025)	0.156	(0.028)	−0.103	(0.029)	−0.380	(0.069)
Patient Experience	0.798	(0.034)	0.254	(0.029)	−0.394	(0.022)	−0.264	(0.040)
Effectiveness of Care	−0.097	(0.024)	0.021	(0.024)	0.121	(0.018)	−0.083	(0.044)
Timeliness of Care	0.606	(0.019)	0.361	(0.019)	−0.345	(0.022)	−1.037	(0.054)
Efficient Use of Medical Imaging	−0.033	(0.034)	−0.110	(0.028)	0.058	(0.023)	0.104	(0.048)
Proposed								
Mortality	−0.145	(0.024)	−0.024	(0.021)	0.079	(0.018)	0.265	(0.041)
Safety of Care	−0.011	(0.052)	0.124	(0.017)	−0.041	(0.021)	−0.115	(0.052)
Readmission	0.138	(0.020)	0.169	(0.022)	−0.091	(0.021)	−0.286	(0.050)
Patient Experience	0.727	(0.029)	0.244	(0.024)	−0.342	(0.018)	−0.276	(0.033)
Effectiveness of Care	0.282	(0.028)	0.086	(0.024)	−0.083	(0.020)	−0.438	(0.050)
Timeliness of Care	0.568	(0.017)	0.315	(0.016)	−0.178	(0.017)	−0.733	(0.041)
Efficient Use of Medical Imaging	0.103	(0.029)	−0.022	(0.022)	0.070	(0.013)	0.078	(0.023)

Table 5. Percentage of Hospitals That Transition from an October 2016 Quintile to a December 2017 Quintile Under the Proposed Efficient Frontier Approach

	Proposed quintile, December 2017				
Proposed quintile, October 2016	1	2	3	4	5
1	62.3%	25.5%	9.4%	2.3%	0.6%
2	25.6%	32.9%	22.9%	14.0%	4.7%
3	8.3%	24.4%	32.5%	26.6%	8.2%
4	3.7%	12.6%	24.8%	32.6%	26.4%
5	0.7%	4.1%	9.8%	24.7%	60.7%

because each measure represents an identical number of patients. In contrast, the CMS approach differentiates between the various measures. According to the scree plots of Venkatesh et al. (2017b), Patient Experience is the only group where the principal component explains more than 70% of the variance, so that the assumption of a single latent variable is less dubious.

In general, differences between weights under our approach are largely driven by different denominators between the measures. Hence, we weight std_MORT_30_PN higher than std_MORT_30_HF because pneumonia impacts more people than heart failure. In contrast, the CMS approach reverses this ordering because it assumes 30-day mortality for heart failure is more correlated with the latent quality factor than 30-day mortality for pneumonia. Another example of this behaviour is in Safety of Care, comparing std_PSI_90_SAFETY and std_COMP_HIP_KNEE, whereas the other safety measures impact relatively fewer patients.

5.3. Changes in Hospital Rankings

We now report how hospital rankings are impacted by the different methods and over time. Because the performance and reporting of hospitals change over time, it is expected that rankings will change over time as well. Under our proposed approach, Table 5 displays the percentage of hospitals that transition from a given quintile in October 2016 to a given quintile in December 2017. For comparison, Table 6 shows this for the CMS approach, where hospitals are ranked using hospital summary scores. Both approaches yield similar changes, and no obvious patterns stand out.

For December 2017, Table 7 displays the number of published hospitals in a given CMS quintile that are assigned a given quintile under our approach. Hence, it measures the extent to which our approach revises the CMS rankings. Hospitals often remain in their same quintiles. However, many hospitals are reassigned quintiles, typically moving up or down by only one quintile. Out of a total of 3,692 published hospitals, there are 33 hospitals that move up at least two quintiles, and 33 hospitals that move down at least two quintiles. Among these, there are 4 hospitals that drop three quintiles.

Upon examination, these are instances in which the hospital does not report the measures that are heavily weighted by the CMS approach. In fact, not a single published critical access hospital reports std_PSI_90_SAFETY. However, many report poor performance on one or other measures; for example, they report only std_HAI_6 in the Safety of Care group and do poorly on this measure. As a consequence, the CMS approach assigns a group summary score of nearly zero, because the latent variable model believes this measure is essentially uncorrelated with quality—indeed std_HAI_6 is given a loading factor of only 0.01 (Venkatesh et al. 2017a). On the other hand, our approach shifts all weight onto the reported measure, assuming, in the absence of additional

Table 6. Percentage of Hospitals That Transition from an October 2016 Quintile to a December 2017 Quintile Under CMS's Latent Variable Model Approach

	CMS quintile, December 2017				
CMS quintile, October 2016	1	2	3	4	5
1	66.5%	22.5%	8.6%	1.8%	0.6%
2	22.5%	33.2%	28.2%	11.4%	4.6%
3	9.1%	25.2%	29.3%	25.8%	10.6%
4	3.1%	13.4%	21.4%	37.3%	24.8%
5	0.1%	5.2%	9.7%	24.0%	60.9%

Table 7. Number of Hospitals in Each Quintile Using CMS's Latent Variable Model Approach That Transition to Each Quintile Under the Proposed Efficient Frontier Approach, for December 2017

	Proposed quintile				
CMS quintile	1	2	3	4	5
1	642	95	2	0	0
2	84	499	147	8	0
3	10	132	426	147	23
4	3	11	155	457	112
5	0	1	8	126	604

Table 8. Statistics on Hospital Reporting of Measures and Groups

Hospital size	Count	Avg. # groups (std. dev.)	Avg. # measures (std. dev.)	Min # measures	Max # measures
Critical access	651	5.82 (0.99)	25.97 (8.15)	10	45
Small	1,012	6.77 (0.56)	38.33 (8.52)	13	54
Medium	1,654	6.95 (0.30)	48.18 (6.03)	14	56
Large	371	6.97 (0.24)	52.11 (4.50)	18	57

Note. std. dev., Standard deviation.

information, that poor performance on the measure available reflects a broader performance problem. In our example, because std_HAI_6 is the only measure reported, our approach assumes that the indication of an extremely serious *Clostridium difficile* infection rate suggests a broader safety problem, whereas the CMS approach ignores it. Hence, our proposed approach substantially lowers the rankings of such hospitals. These hospitals tend also to report fewer groups, and so the resulting reweighting of the groups accentuates this effect.

This same dynamic causes hospitals to improve under our proposed approach, by working in the opposite direction. For example, consider again a hospital that reports only the measure std_HAI_6 in the Safety of Care group. If it does exceptionally well on this measure, then our proposed approach would put all of the group weight (22%) onto it, resulting in a high score for this group. In contrast, the CMS approach, as before, would assign a group summary score of nearly zero.

As shown in Table 8, smaller hospitals report many fewer measures (and groups). Thus, this effect impacts all hospital categories, but it is dramatic for critical access hospitals. Figure 7 compares the distributions in number of measures reported in the Safety of Care group between critical access hospitals and large hospitals. They are inverted, with 53.1% of critical access hospitals reporting no measures at all and 46.9% reporting one to four measures to constitute their Safety of Care group, versus 79.8% of large hospitals reporting all eight measures in the Safety of Care group. (The distributions for small and medium hospitals (not shown) have modes of three and seven, respectively.) This sparsity of reporting is likely related to the fact that critical access hospitals are not required to participate in the CMS's Hospital-Acquired Condition Reduction program. Figure 8 shows the distributions of the Safety of Care group score, for both the CMS approach and our approach. The former produces scores of 0 ± 0.3, whereas the latter distributes scores between −3 and 2, an order of magnitude greater dispersion. Figure 9 shows that 312 critical access hospitals, or nearly half, change. Of these, 138 decrease, whereas 174 increase, so that under the CMS approach, relatively more are not receiving credit for good safety performance than not receiving penalties for poor safety performance. In either case, the total number of hospital ratings impacted is substantial.

Figure 7. Number of Measures Reported in the Safety of Care Group by Critical Access and Large Hospitals

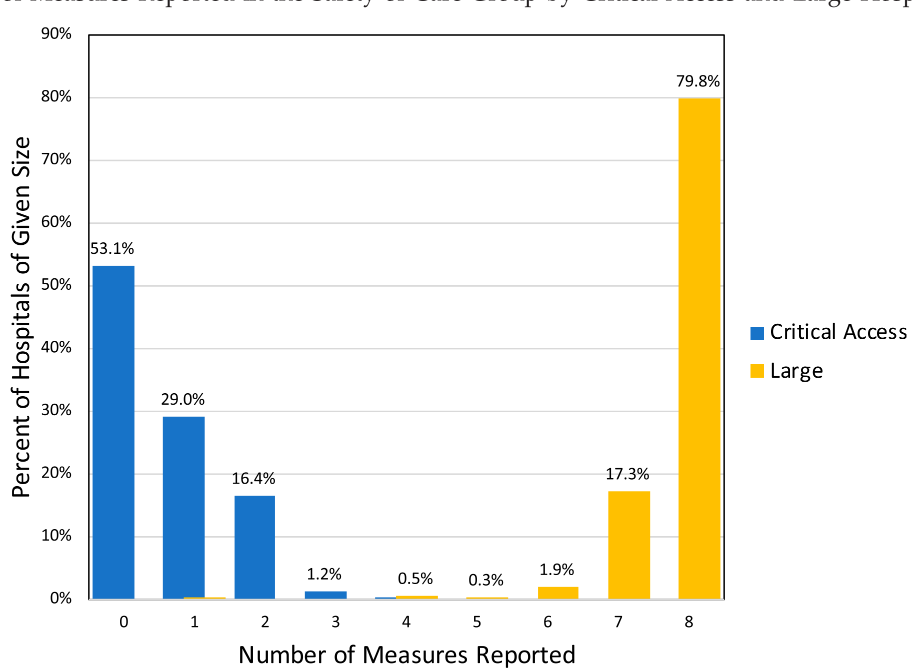


Figure 8. Distribution of Safety of Care Group Scores for Critical Access Hospitals: The CMS vs. Proposed Approach

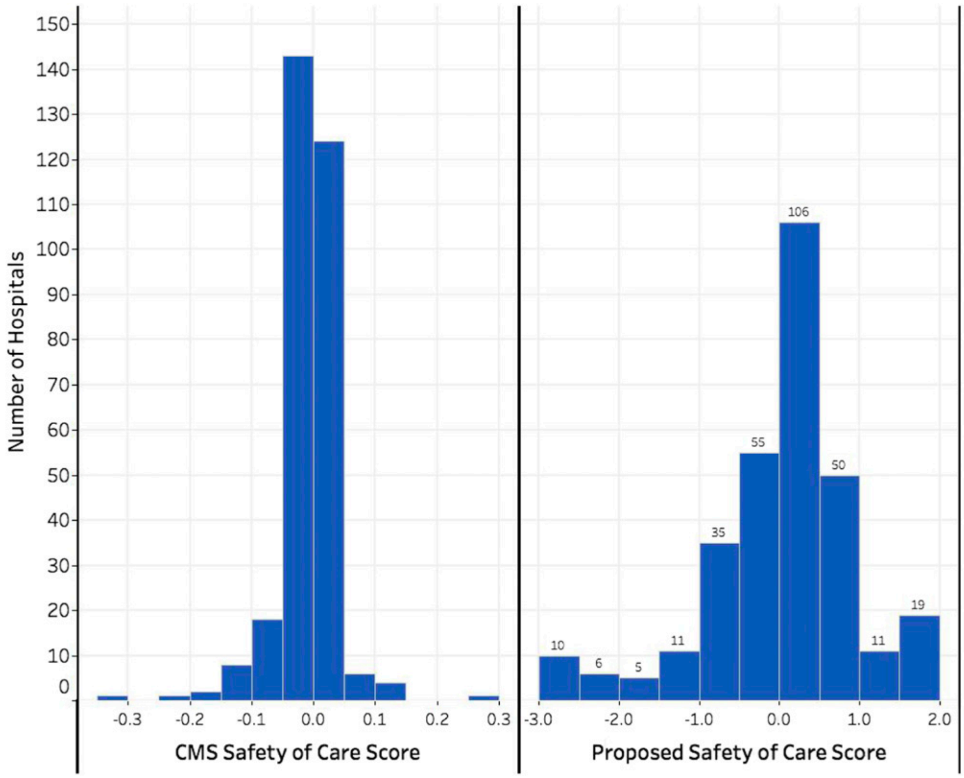
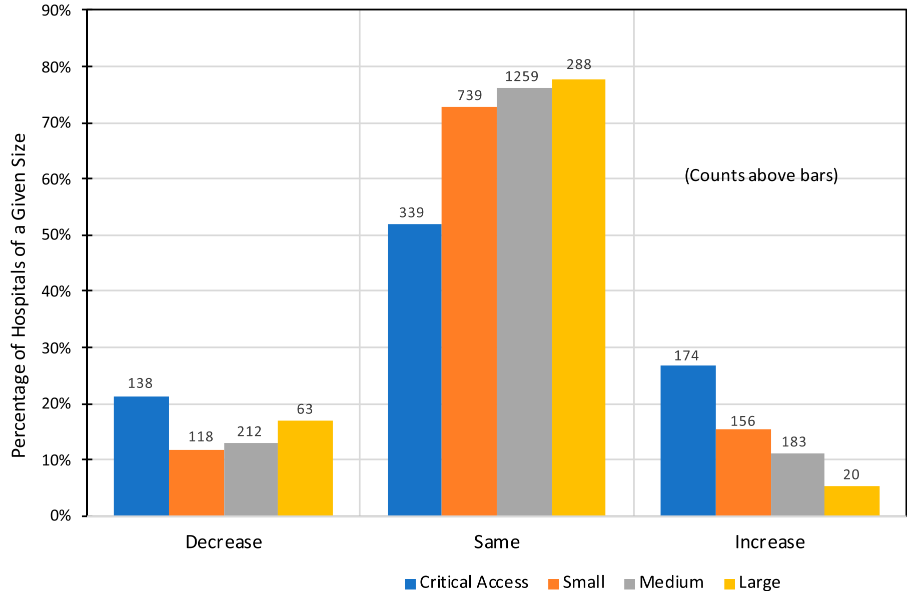


Figure 9 also shows that larger hospitals are more likely to retain their same CMS quintile than smaller hospitals. However, conditional on being revised, larger hospitals are more than three times more likely to decrease their quintile than increase it (63 hospitals versus 20). In contrast, smaller hospitals that revise are 32.2% more likely to increase than to decrease.

A closer look suggests this is an artifact of smaller hospitals crowding out larger hospitals in the rankings, as discussed earlier, rather than an inherent bias in our proposed approach. When recomputing quintiles within each hospital size category separately, rather than globally across all hospitals, the apparent bias disappears, as shown in Figure 10. Again, larger

Figure 9. CMS vs. Proposed Change in Global Quintile by Hospital Size



hospitals are less likely to be revised than smaller hospitals, but the conditional likelihood of decreasing or increasing quintiles is almost completely symmetric. Hence, there is a rationale to partition hospitals into size categories for purposes of ranking, and possibly even calculating, scores, to avoid bias because hospitals of different sizes are not fairly comparable to one another.

6. Conclusions

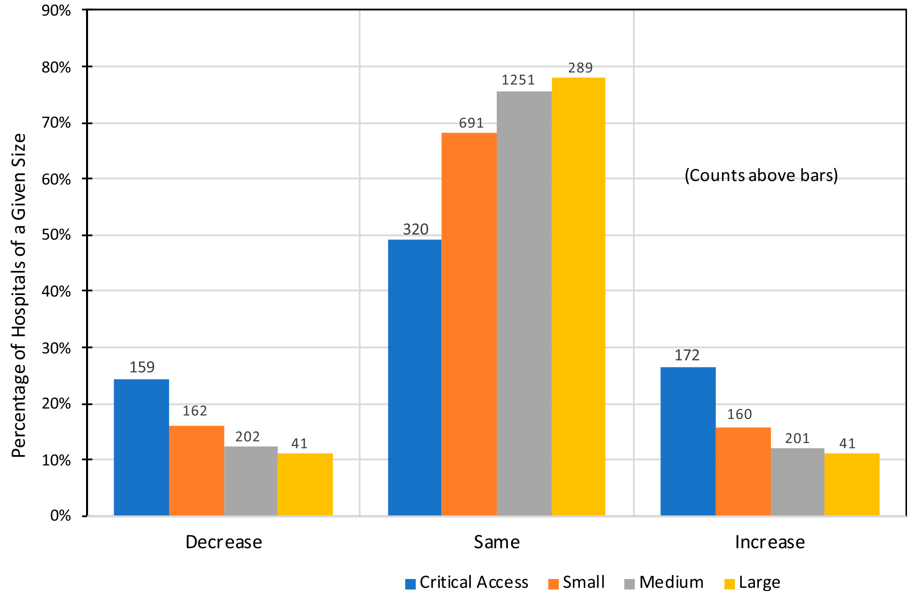
We have developed a new approach for scoring and ranking hospitals that has many benefits over the existing latent variable model used by the CMS. Our approach is more equitable and transparent in giving hospitals credit or penalties for measures ignored by the LVM approach. To provide sound footing for our approach, we have provided a rather complete theory of its inner workings. Our most important theoretical result, in terms of practical implications, states that hospital scores are guaranteed to improve if hospitals improve on all measures relative to all other hospitals, under reasonable conditions that can be expected to be approximately satisfied in practice (Property 8 and Theorem 1). We believe that this result is a basic requirement that hospitals should expect to hold, in the interest of fairness and creating appropriate incentives to improve. This theory stands in contrast to the existing approach, in which weights can change dramatically over time as the result of infinitesimal shifts in correlations, as we show rigorously using simple examples. At the same time, our approach meets the CMS's design requirement that it can autonomously compute weights for measures as they are added and subtracted over time.

Our improvement result requires, as would any such result, that certain aspects of the scoring environment remain stable, and can break down when they are not. For example, a measure for which a hospital performs relatively poorly can always be added, resulting in a lower hospital score. Even fixing the set of measures, in our approach significant shifts in the national volumes of patients impacted by the measures could significantly shift measure weights, making it possible for improving hospitals to decrease their scores. However, infinitesimal shifts will result in only infinitesimal shifts in hospital scores (a result of math programming sensitivity analysis), not dramatic shifts in the scores and measure weights, as we see in the LVM approach with respect to correlations. Thus, our approach enjoys substantially greater stability properties. Furthermore, for hospitals trying to predict their future scores, forecasting patient volumes is a more straightforward exercise than forecasting correlations between measures in next year's national data.

Although theory establishes conditions under which hospital scores are guaranteed to improve, this does not imply that it is easier for hospitals to obtain a higher score under our approach than the CMS's approach. In fact, it may be harder because our approach tends to disperse more weight across measures. Thus, even if a hospital improves along a few measures, they may be more likely to get penalized for degrading performance along the other measures.

Empirically, we find that our approach scores and ranks hospitals in a manner largely consistent with the CMS's approach, except many hospitals increase or decrease their quintiles because of our approach

Figure 10. CMS vs. Proposed Change in Quintile Computed Within Hospital Size Categories



giving hospitals credit or penalties for measures largely ignored by the current LVM approach. Because smaller hospitals tend to report fewer measures, this can distort their scores under the current approach when they do not report measures that CMS weights heavily. This is because it leads to a group score near zero (average), even if they perform favorably or poorly on the measures they do report. As a result of this effect, nearly half of all critical access hospitals increase or decrease by at least one quintile in the rankings under our approach versus the CMS approach. Although it appears that, conditional on changing quintiles, larger hospitals are three times more likely to decrease than increase, with the reverse but weaker effect for smaller hospitals, this bias disappears when hospitals are ranked within their size category. This suggests that larger hospitals are currently being crowded out in the ordinal rankings by smaller hospitals. In general, our results suggest that the current approach of lumping together all hospitals, for both calculation and reporting, should be revised. Furthermore, the rules allowing measures to not be reported should be reviewed and potentially made more strict to help make hospitals more comparable and minimize opportunities for hospitals to game the system.

Our approach does have some limitations, which could be addressed in future research. One concern may be that it does not account for a hospital's sample size in the measures reported. Although true as presented here, many of the measures themselves are already adjusted to account for sample size, known as "shrinkage." To adjust for it again, as the CMS approach currently does using likelihood weights, represents double shrinkage. Nonetheless, our approach can also adjust for hospital sample sizes through the cones C^o, using hospital-specific constraints. For example, rather than use the nationally aggregated denominators, one could use hospital-specific denominators, or some combination of the two. Because our hospital group and summary scores are convex combinations of underlying measures, confidence intervals for them can be obtained by considering a hospital's standard errors and sample sizes for the measures themselves. The bootstrap methods developed by Simar and Wilson (1998) may also be relevant.

Another criticism is that requiring the ratios of weights to exceed their denominator ratios (i.e., Property 5) assumes that measures within a group all have equivalent clinical significance to an individual patient. However, it would be trivial to add an adjustment factor to each measure's denominator to account for this. When measures account for relatively rare events in very large populations, one could consider lower bounding the ratio of two measures with their incidence ratios by using numerators instead of denominators. Our approach is flexible enough to accommodate new constraints and parameters deemed important by policy makers.

Our approach is intended only to fix problems associated with the latent variable model. If there are other problems with the CMS star ratings, they will persist. For example, as cited earlier, concerns have been raised about individual measures not properly adjusting for socioeconomic factors, and the potential for bias against larger hospitals. Ranking large and small hospitals together creates distortion in ordinal rankings. Small hospitals are permitted to report many fewer measures yet still be rated. The *k*-means clustering step changes the cutoffs each time it is run, and therefore, hospitals may not increase in their ratings over time even if they do so in their overall hospital scores. Even problematic is the notion itself of labeling hospitals with one to five stars, when the differences between hospitals may not be that great, especially at the boundaries. Because of these persistent issues, any rankings, scores, or ratings that result from using our approach should continue to be viewed with great skepticism and criticism. We offer improvement in only one specific aspect of the ratings system, which may not be fully appreciated until other elements are improved.

We have developed theory and methodology for the purpose of scoring and ranking hospitals, but in general, they can be applied in many other application settings to score and rank the performance of entities. Although we focused here on the directional distance model, our ideas may extend to other models in the efficient frontier literature to provide a new integrated approach to both efficiency analysis and performance evaluation.

Acknowledgments

This paper would not have been possible without the inspiration and contextual support of Dr. Omar B. Lateef. The author thanks his colleagues Barış Ata and John Birge for comments on an early draft of this manuscript. The author also thanks discussant Serpil Sayın and the participants of the 2019 Healthcare Operations Workshop at Koç University for many helpful questions and comments. The author thanks three anonymous reviewers for many suggestions that significantly improved this manuscript's exposition.

Appendix A. Result Pertaining to Directional Improvement

We specialize Theorem 1 to the particular case in which improvement occurs in the direction oriented between hospital o and the ideal hospital, given by $\overline{y}$. Under given conditions, we can formally adjust the optimal solution from before improvement to an optimal solution after improvement. This allows us to identify the change in score and efficiency in closed form. Among the required conditions is one that ensures the scores of best-performing hospitals, as computed under the preimprovement measure weights, do not change too much.

Proposition A.1 (Directional Perturbation). *Let* $(\beta^*, \boldsymbol{\lambda}^*, \boldsymbol{s}^*)$, $(\theta^*, \boldsymbol{\mu}^*)$ *denote an optimal primal–dual solution to* $(\mathbf{P}^o), (\mathbf{D}^o)$ *such that* $\sum_{i\in\mathcal{I}^o}(\bar{y}_i - y_i^o)\mu_i^* = 1$ *and* S^o *is maximal. Suppose now that each hospital* $j \in \mathcal{J}$ *changes its performance from* y_i^j *to* $\hat{y}_i^j$ $\forall i \in \mathcal{I}^j$. *Let there exist nonnegative constants* $\hat{\beta}, \beta^\lambda, \bar{\beta}$ *satisfying* $\hat{\beta} \le \beta^* + \beta^\lambda$, $\hat{\beta} < 1 + \bar{\beta}$, *such that for all* $i \in \mathcal{I}^o$, *performance improvement at certain hospitals occurs in the direction* $\bar{\boldsymbol{y}} - \boldsymbol{y}^o$, *that is,*

$$\begin{aligned}
\hat{y}_i^o &= y_i^o + \hat{\beta}\cdot(\bar{y}_i - y_i^o),\\
\sum_{\{j\in\mathcal{J}: i\in\mathcal{I}^j\}} \hat{y}_i^j\lambda_j^* &= \sum_{\{j\in\mathcal{J}: i\in\mathcal{I}^j\}} y_i^j\lambda_j^* + \beta^\lambda\cdot(\bar{y}_i - y_i^o),\\
\hat{\bar{y}}_i &= \bar{y}_i + \bar{\beta}\cdot(\bar{y}_i - y_i^o),
\end{aligned} \tag{A.1}$$

and

$$\max_{j\in\mathcal{J}} \sum_{i\in\mathcal{I}^o\cap\mathcal{I}^j} \hat{y}_i^j\mu_i^* \le \max_{j\in\mathcal{J}} \sum_{i\in\mathcal{I}^o\cap\mathcal{I}^j} y_i^j\mu_i^* + \beta^\lambda.$$

Then,

$$\hat{S}^o = S^o + \frac{\hat{\beta}}{\Gamma^*},$$

and

$$\hat{T}^o = \frac{T^o + \bar{\beta} - \beta^\lambda}{1 + \bar{\beta} - \hat{\beta}}.$$

Proof. Let

$$\gamma = \frac{1}{1 + \bar{\beta} - \hat{\beta}} > 0.$$

We will show that the following is an optimal primal–dual solution to $(\mathbf{P}^o), (\mathbf{D}^o)$ under the new data $\hat{y}$:

$$\begin{aligned}
\beta' &= \gamma\cdot\left(\beta^* + \beta^\lambda - \hat{\beta}\right),\\
\lambda_j' &= \lambda_j^* \quad \forall j \in \mathcal{J},\\
s_i' &= s_i^* \quad \forall i \in \mathcal{I}^o,\\
\mu_i' &= \gamma\mu_i^* \quad \forall i \in \mathcal{I},\\
\theta' &= \gamma\cdot\left(\sum_{i\in\mathcal{I}^o}\hat{y}_i^o\mu_i^* + \beta^* + \beta^\lambda - \hat{\beta}\right).
\end{aligned}$$

First, we verify feasibility to $(\mathbf{P}^o)$ with data $\hat{y}$. For each $i \in \mathcal{I}^o$, the left-hand side of (7b) reads

$$\begin{aligned}
&\hat{y}_i^o + \beta'\cdot\left(\hat{\bar{y}}_i - \hat{y}_i^o\right) + s_i'\\
&\quad= y_i^o + \hat{\beta}\cdot\left(\bar{y}_i - y_i^o\right) + \beta'\\
&\qquad\times\left[\left(\bar{y}_i + \bar{\beta}\cdot\left(\bar{y}_i - y_i^o\right)\right) - \left(y_i^o + \hat{\beta}\cdot\left(\bar{y}_i - y_i^o\right)\right)\right] + s_i^*\\
&\quad= y_i^o + \hat{\beta}\cdot\left(\bar{y}_i - y_i^o\right) + \beta'\cdot\left(1 + \bar{\beta} - \hat{\beta}\right)\left(\bar{y}_i - y_i^o\right) + s_i^*\\
&\quad= y_i^o + \left(\beta^* + \beta^\lambda\right)\left(\bar{y}_i - y_i^o\right) + s_i^*,
\end{aligned}$$

plugging in the proposed expression for β' and canceling terms. Subtracting this from (A.1), which equals the right-hand side of (7b), yields 0, because $(\beta^*, \boldsymbol{\lambda}^*, \boldsymbol{s}^*)$ is a feasible solution to $(\mathbf{P}^o)$ with data $\boldsymbol{y}$. All other constraints of $(\mathbf{P}^o)$ are trivially satisfied.

Next we verify feasibility to $(\mathbf{D}^o)$ with data $\hat{\boldsymbol{y}}$. The left-hand side of (5c) becomes

$$\begin{aligned}
&\sum_{i\in\mathcal{I}^o}\left(\hat{\bar{y}}_i - \hat{y}_i^o\right)\mu_i'\\
&\quad= \sum_{i\in\mathcal{I}^o}\left[\left(\bar{y}_i + \bar{\beta}\cdot\left(\bar{y}_i - y_i^o\right)\right) - \left(y_i^o + \hat{\beta}\cdot\left(\bar{y}_i - y_i^o\right)\right)\right]\mu_i'\\
&\quad= \sum_{i\in\mathcal{I}^o}\left(1 + \bar{\beta} - \hat{\beta}\right)\left(\bar{y}_i - y_i^o\right)\mu_i'\\
&\quad= \sum_{i\in\mathcal{I}^o}\left(\bar{y}_i - y_i^o\right)\mu_i^*\\
&\quad= 1,
\end{aligned}$$

using the definition of γ. Also, because $0 < \gamma < \infty$, $\boldsymbol{\mu}^* \in C^o$ implies $\gamma\boldsymbol{\mu}^* \in C^o$. To verify (5b), we have

$$\begin{aligned}
\theta' &= \gamma\cdot\left(\sum_{i\in\mathcal{I}^o}\hat{y}_i^o\mu_i^* + \beta^* + \beta^\lambda - \hat{\beta}\right)\\
&= \gamma\cdot\left(\sum_{i\in\mathcal{I}^o}\left(y_i^o + \hat{\beta}\cdot\left(\bar{y}_i - y_i^o\right)\right)\mu_i^* + \beta^* + \beta^\lambda - \hat{\beta}\right)\\
&= \gamma\cdot\left(\sum_{i\in\mathcal{I}^o} y_i^o\mu_i^* + \beta^* + \beta^\lambda\right)\\
&= \gamma\cdot\left(\sum_{i\in\mathcal{I}^o}\left(y_i^o + \beta^*\cdot\left(\bar{y}_i^o - y_i^o\right)\right)\mu_i^* + \beta^\lambda\right)\\
&= \gamma\cdot\left(\theta^* + \beta^\lambda\right)\\
&\ge \gamma\left(\max_{j\in\mathcal{J}}\sum_{i\in\mathcal{I}^o\cap\mathcal{I}^j} y_i^j\mu_i^* + \beta^\lambda\right)\\
&\ge \gamma\max_{j\in\mathcal{J}}\sum_{i\in\mathcal{I}^o\cap\mathcal{I}^j}\hat{y}_i^j\mu_i^*\\
&\ge \sum_{i\in\mathcal{I}^o\cap\mathcal{I}^j}\hat{y}_i^j\mu_i' \quad \forall j \in \mathcal{J}.
\end{aligned}$$

Last, the objective function of $(\mathbf{D}^o)$ evaluates to

$$\begin{aligned}
\theta' - \sum_{i\in\mathcal{I}^o}\hat{y}_i^o\mu_i' &= \gamma\cdot\left(\sum_{i\in\mathcal{I}^o}\hat{y}_i^o\mu_i^* + \beta^* + \beta^\lambda - \hat{\beta}\right) - \sum_{i\in\mathcal{I}^o}\hat{y}_i^o\mu_i'\\
&= \frac{\beta^* + \beta^\lambda - \hat{\beta}}{1 + \bar{\beta} - \hat{\beta}}\\
&= \beta'.
\end{aligned}$$

Therefore, the proposed primal–dual solution is optimal under new data $\hat{\boldsymbol{y}}$.

Finally, the new score equals

$$\begin{aligned}
\hat{S}^o &= \sum_{i\in\mathcal{I}^o}\hat{y}_i^o\frac{\mu_i'}{\Gamma'}\\
&= \sum_{i\in\mathcal{I}^o}\hat{y}_i^o\frac{\gamma\mu_i^*}{\gamma\Gamma^*}\\
&= \sum_{i\in\mathcal{I}^o}\hat{y}_i^o\frac{\mu_i^*}{\Gamma^*}\\
&= \sum_{i\in\mathcal{I}^o}\left(y_i^o + \hat{\beta}\cdot\left(\bar{y}_i - y_i^o\right)\right)\frac{\mu_i^*}{\Gamma^*}\\
&= S^o + \frac{\hat{\beta}}{\Gamma^*}.
\end{aligned}$$

Furthermore, the proposed expression for $\hat{T}^o = 1 - \beta'$ follows straightforwardly from substituting β' defined above and rearranging terms. □

As a consequence of this result, we can show that the virtual hospital, constructed as a convex combination of reference hospitals, is indeed on the efficient frontier.

Corollary A.1. *In Proposition A.1, if $\hat{\beta} = \beta^*$ and $\beta^\lambda = \overline{\beta} = 0$, then $\hat{T}^o = 1$. In other words,*

$$\hat{y}_i^o = y_i^o + \beta^* \cdot (\overline{y}_i - y_i^o) \quad \forall i \in \mathcal{I}^o$$

is on the efficient frontier. Furthermore, so is the virtual hospital, which is defined by

$$\sum_{\{j \in \mathcal{J} : i \in \mathcal{I}^j\}} y_i^j \lambda_j^* \quad \forall i \in \mathcal{I}^o.$$

Proof. Substituting the given values into the expression in Proposition A.1 for $\hat{T}^o$ yields $\hat{T}^o = 1$. This shows that $\hat{\boldsymbol{y}}$ is on the efficient frontier. To verify that the virtual hospital is on the efficient frontier, note that from (7b),

$$\sum_{\{j \in \mathcal{J} : i \in \mathcal{I}^j\}} y_i^j \lambda_j^* = \hat{y}_i^o + s_i^* \quad \forall i \in \mathcal{I}^o.$$

Thus, we just need to show that $\hat{\boldsymbol{y}}^o + \boldsymbol{s}^*$ is on the efficient frontier. Using the optimal primal–dual pair from the proposition, $(\beta', \boldsymbol{\lambda}', \boldsymbol{s}')$, $(\theta', \boldsymbol{\mu}')$ we see that (5b) changes only for $j = o$. The left-hand side reads

$$\sum_{i \in \mathcal{I}^o} (\hat{y}_i^o + s_i^*) \mu_i' = \sum_{i \in \mathcal{I}^o} \hat{y}_i^o \mu_i' + \gamma \sum_{i \in \mathcal{I}^o} s_i^* \mu_i^* = \sum_{i \in \mathcal{I}^o} \hat{y}_i^o \mu_i' \leq \theta'$$

because of complementary slackness (11). Because the objective function (5a) has the same terms as in the last display, we conclude from $\hat{T}^o = 1$ that it equals 0. The constraint (5c) is satisfied by a similar argument using complementary slackness (11). To verify feasibility to $(\mathbf{P}^o)$, replace $\boldsymbol{s}' = \boldsymbol{s}^*$ with $\boldsymbol{s}' = 0$, which is feasible because $\mathbf{0} \in C_\star^o$. Then constraints (7b) read, with $\beta' = 0$,

$$(\hat{y}_i^o + s_i^*) + s_i' = \sum_{\{j \in \mathcal{J} : i \in \mathcal{I}^j\}} y_i^j \lambda_j^* \quad \forall i \in \mathcal{I}^o,$$

which is satisfied by construction with $\boldsymbol{s}' = \mathbf{0}$. Therefore, by strong duality, we have that $\beta' = 0$ implies that the virtual hospital is on the efficient frontier. □

Appendix B. Descriptions of Measures Included in December 2017 Hospital Star Ratings

Appendix C of Venkatesh et al. (2017a, pp. 22–24) provides descriptions the measures included in the December 2017 hospital star ratings. We reproduce them below for convenience:

Mortality

1. MORT-30-AMI: Acute Myocardial Infarction (AMI) 30-Day Mortality Rate
2. MORT-30-CABG: Coronary Artery Bypass Graft (CABG) 30-Day Mortality Rate
3. MORT-30-COPD: Chronic Obstructive Pulmonary Disease (COPD) 30-Day Mortality Rate
4. MORT-30-HF: Heart Failure (HF) 30-Day Mortality Rate
5. MORT-30-PN: Pneumonia (PN) 30-Day Mortality Rate
6. MORT-30-STK: Acute Ischemic Stroke (STK) 30-Day Mortality Rate
7. PSI-4-SURG-COMP: Death Among Surgical Patients with Serious Treatable Complications

Safety of Care

1. HAI-1: Central-Line Associated Bloodstream Infection (CLABSI)
2. HAI-2: Catheter-Associated Urinary Tract Infection (CAUTI)
3. HAI-3: Surgical Site Infection from colon surgery (SSI-colon)
4. HAI-4: Surgical Site Infection from abdominal hysterectomy (SSI-abdominal hysterectomy)
5. HAI-5: MRSA Bacteremia
6. HAI-6: *Clostridium difficile* (*C. difficile*)
7. COMP-HIP-KNEE: Hospital-Level Risk-Standardized Complication Rate (RSCR) Following Elective Primary Total Hip Arthroplasty (THA) and Total Knee Arthroplasty (TKA)
8. PSI-90-SAFETY: Complication/Patient Safety for Selected Indicators (PSI)

Readmission

1. READM-30-CABG: Coronary Artery Bypass Graft (CABG) 30-Day Readmission Rate
2. READM-30-COPD: Chronic Obstructive Pulmonary Disease (COPD) 30-Day Readmission Rate
3. READM-30-Hip-Knee: Hospital-Level 30-Day All-Cause Risk-Standardized Readmission Rate (RSRR) Following Elective Total Hip Arthroplasty (THA)/Total Knee Arthroplasty (TKA)
4. READM-30-PN: Pneumonia (PN) 30-Day Readmission Rate
5. READM-30-STK: Stroke (STK) 30-Day Readmission Rate
6. READM-30-HOSP-WIDE: HWR Hospital-Wide All-Cause Unplanned Readmission
7. EDAC-30-AMI: Excess Days in Acute Care (EDAC) after hospitalization for Acute Myocardial Infarction (AMI)
8. EDAC-30-HF: Excess Days in Acute Care (EDAC) after hospitalization for Heart Failure (HF)
9. OP-32: Facility 7-Day Risk Standardized Hospital Visit Rate after Outpatient Colonoscopy

Patient Experience

1. H-CLEAN-HSP: Cleanliness of Hospital Environment (Q8)
2. H-COMP-1: Nurse Communication (Q1, Q2, Q3)
3. H-COMP-2: Doctor Communication (Q5, Q6, Q7)
4. H-COMP-3: Responsiveness of Hospital Staff (Q4, Q11)
5. H-COMP-4: Pain Management (Q13, Q14)
6. H-COMP-5: Communication About Medicines (Q16, Q17)
7. H-COMP-6: Discharge Information (Q19, Q20)
8. H-HSP-RATING: Overall Rating of Hospital (Q21)
9. H-QUIET-HSP: Quietness of Hospital Environment (Q9)
10. H-RECMND: Willingness to Recommend Hospital (Q22)
11. H-COMP-7: HCAHPS 3 Item Care Transition Measure (CTM-3)

Effectiveness of Care

1. IMM-2: Influenza Immunization
2. IMM-3/OP-27: Healthcare Personnel Influenza Vaccination

3. OP-4: Aspirin at Arrival
4. OP-22: Emergency Department (ED)-Patient Left Without Being Seen
5. OP-23: Emergency Department (ED)-Head Computed Tomography (CT) or Magnetic Resonance Imaging (MRI) Scan Results for Acute Ischemic Stroke or Hemorrhagic Stroke Who Received Head CT or MRI Scan Interpretation Within 45 Minutes of Arrival
6. PC-01: Elective Delivery
7. VTE-6: Hospital Acquired Potentially-Preventable Venous Thromboembolism
8. OP-29: Endoscopy/Poly Surveillance-Appropriate Follow-up Interval for Normal Colonoscopy in Average Risk Patients
9. OP-30: Endoscopy/Poly Surveillance: Colonoscopy Interval for Patients with a History of Adenomatous Polyps—Avoidance of Inappropriate Use
10. OP-33: External Beam Radiotherapy for Bone Metastases

Timeliness of Care

1. ED-1b: Median Time from Emergency Department (ED) Arrival to ED Departure for Admitted ED Patients
2. ED-2b: Admit Decision Time to Emergency Department (ED) Departure Time for Admitted Patients
3. OP-3b: Median Time to Transfer to Another Facility for Acute Coronary Intervention
4. OP-5: Median Time to Electrocardiography (ECG)
5. OP-18b: Median Time from Emergency Department (ED) Arrival to ED Departure for Discharged ED Patients
6. OP-20: Door to Diagnostic Evaluation by a Qualified Medical Professional
7. OP-21: Emergency Department (ED)-Median Time to Pain Management for Long Bone Fracture

Efficient Use of Medical Imaging

1. OP-8: Magnetic Resonance Imaging (MRI) Lumbar Spine for Low Back Pain
2. OP-10: Abdomen Computed Tomography (CT) Use of Contrast Material
3. OP-11: Thorax Computed Tomography (CT) Use of Contrast Material
4. OP-13: Cardiac Imaging for Preoperative Risk Assessment for Non-Cardiac Low-Risk Surgery
5. OP-14: Simultaneous Use of Brain Computed Tomography (CT) and Sinus CT

Appendix C. Numerical Details and Overview of the CMS Methodology

Whenever hospital star ratings are released publicly, the CMS simultaneously posts the computer code and data on www.qualitynet.org in the form of an SAS package, which completely replicates the published ratings as detailed in the methodology report (Venkatesh et al. 2017b). Far more detailed information is available from the SAS output than the CMS publishes on its Hospital Compare website (www.medicare.gov/hospitalcompare). For example, the SAS output provides hospital-level summary and group scores, whereas the Hospital Compare website provides only the overall star rating (one to five stars) that is derived from those scores.

Venkatesh et al. (2017b, p. 9) provide the following summary of the CMS methodology:

> The measures were first selected based on their relevance and importance as determined through stakeholder and expert feedback. The selected measures were standardized to be consistent in terms of direction and magnitude, with outlying values trimmed (Step 1). In Step 2, the measures were organized into seven groups by measure type. In Step 3, the standardized measures for each group were used to construct a latent variable statistical model that reflected the dimension of quality represented by the measures within the given group. Each of the seven statistical models generated a hospital-specific group score, which is obtained as a prediction of the latent variable. In Step 4, a weight was applied to each group score, and all available groups were averaged to calculate a hospital summary score. In Step 5, the public reporting threshold is applied, requiring hospitals to have a minimum of three measure groups (one of which must be an outcome group) with at least three measure in each of the three groups to be included in the final clustering step. Finally, in Step 6, to assign star ratings, hospital summary scores were organized into five categories using a clustering algorithm.

We ran the SAS packages from October 2016 and December 2017, and used the standardized measure data and denominators from step 1 of Venkatesh et al. (2017b) as input into our proposed models. Larger measure values represent better performance. For comparison, we ran all steps of the CMS methodology, except step 6, merely substituting our methodology for steps 3 and 4. We do not report star ratings based on k-means clustering as described in the methodology's step 6. With k-means clustering, it is possible that a hospital can increase its ordinal rank, but decrease in the cluster it is assigned to. Instead, we partition the hospital-level summary scores in quintiles (i.e. numbered 1 through 5, with 5 representing the top 20%), and we do the same for the hospital scores S^o coming from our model.

For the purpose of computing scores, for both methods, we do so using all hospitals computed together and then report only on hospitals that meet the criterion described in step 5 of Venkatesh et al. (2017b) requiring minimal levels of measure reporting. For December 2017, this criterion reduces the number of hospitals to report on from 4,571 down to 3,692, which we call "published" hospitals.

Appendix D. Hospital Performance and Weights on Individual Measures

In this appendix, we provide detailed calculations for measures across all groups. See Appendix B for a detailed description of each measure. Each Table D.1–D.7 is broken down into "Performance," displaying averages of raw standardized measure values, and "Weights," displaying averages of optimal weights across hospitals. So for a given measure $i \in \mathcal{I}$, we report the all-hospital ("All") average

$$\frac{1}{|\{j \in \mathcal{J} : i \in \mathcal{I}^j\}|} \sum_{\{j \in \mathcal{J} : i \in \mathcal{I}^j\}} \mu_i^o / \Gamma^o.$$

For the different size categories, we average over only hospitals within that category.

For comparison, we compute the empirical Bayes estimate of the latent variable using its posterior distribution in the balanced case using (4). Using these measure weights, we estimated the group scores for each hospital and computed the R^2 against the actual group scores produced by the SAS program, displayed in Table D.8 under the column "Fixed." When we compute measure weights using (4) applied to only the subset of measures that a hospital reports, we obtain the R^2 displayed under the column "Variable" in the table. We do not report these variable measure weights in Tables D.1–D.7, say, averaged over all hospitals, because they differ only slightly from the fixed Bayesian estimates reported. As Table D.8 shows, both of these estimates capture the vast majority of the variability in group scores, often providing a near perfect fit. The variability not captured represents the variation in group scores caused by the weights in the likelihood function. Notably, the Effectiveness of Care category gives the worst fit. This is likely due to the measures having more varied representation across hospitals,

Table D.1. Hospital Performance and Weights (Averages) for Readmission (22%) by Hospital Size

	Weights (%)		Performance				
Measure	CMS approach (Bayes est. × 22%)	Efficient frontier (all)	All	Critical access	Small	Medium	Large
std_EDAC_30_AMI	0.10	0.83	0.005	0.597	0.392	−0.012	−0.246
std_READM_30_CABG	0.04	0.23	0.006	N/A	−0.149	−0.033	0.094
std_READM_30_COPD	0.09	1.49	−0.012	0.099	0.151	−0.083	−0.110
std_EDAC_30_HF	0.12	2.08	−0.092	0.459	0.233	−0.182	−0.502
std_READM_30_HIP_KNEE	0.06	1.72	−0.011	0.080	0.092	−0.059	0.004
std_READM_30_HOSP_WIDE	22.27	12.71	−0.027	0.080	0.171	−0.077	−0.345
std_READM_30_PN	0.13	2.60	−0.071	0.259	0.203	−0.190	−0.255
std_READM_30_STK	0.08	0.86	−0.012	0.309	0.163	0.022	−0.414
std_OP_32	0.01	1.32	−0.002	0.040	−0.064	0.031	0.011

Note. est., Estimate.

Table D.2. Hospital Performance and Weights (Averages) for Mortality (22%) by Hospital Size

	Weights (%)		Performance				
Measure	CMS approach (Bayes est. × 22%)	Efficient frontier (all)	All	Critical access	Small	Medium	Large
std_MORT_30_AMI	2.98	2.21	0.007	−0.209	−0.099	−0.024	0.257
std_MORT_30_CABG	2.20	0.70	0.007	N/A	−0.428	−0.123	0.297
std_MORT_30_COPD	4.99	4.07	0.018	−0.072	−0.002	0.006	0.119
std_MORT_30_HF	7.09	5.71	0.090	−0.392	−0.169	0.147	0.479
std_MORT_30_PN	6.31	9.00	0.033	−0.064	−0.060	0.025	0.306
std_MORT_30_STK	3.04	2.63	0.025	−0.311	−0.055	0.068	−0.031
std_PSI_4_SURG_COMP	1.56	0.62	0.000	N/A	−0.039	0.042	−0.119

Note. est., Estimate.

Table D.3. Hospital Performance and Weights (Averages) for Safety of Care (22%) by Hospital Size

	Weights (%)		Performance				
Measure	CMS approach (Bayes est. × 22%)	Efficient frontier (all)	All	Critical access	Small	Medium	Large
std_COMP_HIP_KNEE	0.03	6.60	−0.008	0.007	−0.025	−0.050	0.200
std_HAI_1	0.01	0.11	0.021	N/A	0.152	0.019	−0.028
std_HAI_2	0.00	0.14	0.016	0.218	0.132	0.019	−0.092
std_HAI_3	0.01	0.05	0.009	−0.219	−0.041	0.076	−0.209
std_HAI_4	0.01	0.01	0.010	0.967	−0.154	0.130	−0.147
std_HAI_5	0.02	0.05	0.016	N/A	−0.007	0.015	0.019
std_HAI_6	0.01	1.14	0.017	−0.130	0.096	−0.007	−0.046
std_PSI_90_SAFETY	23.13	17.13	−0.012	N/A	0.141	−0.059	−0.218

Note. est., Estimate.

Table D.4. Hospital Performance and Weights (Averages) for Patient Experience (22%) by Hospital Size

	Weights (%)		Performance				
Measure	CMS approach (Bayes est. × 22%)	Efficient frontier (all)	All	Critical access	Small	Medium	Large
std_H_CLEAN_HSP_LINEAR	0.99	2.00	−0.185	0.967	0.327	−0.387	−0.552
std_H_COMP_1_LINEAR	4.54	2.00	−0.151	0.780	0.265	−0.357	−0.262
std_H_COMP_2_LINEAR	1.56	2.00	−0.154	0.808	0.329	−0.381	−0.337
std_H_COMP_3_LINEAR	2.24	2.00	−0.200	1.003	0.383	−0.447	−0.543
std_H_COMP_4_LINEAR	2.99	2.00	−0.152	0.743	0.247	−0.341	−0.300
std_H_COMP_5_LINEAR	2.14	2.00	−0.178	0.873	0.269	−0.385	−0.366
std_H_COMP_6_LINEAR	1.07	2.00	−0.102	0.461	0.191	−0.230	−0.257
std_H_COMP_7_LINEAR	3.54	2.00	−0.151	0.642	0.149	−0.333	−0.084
std_H_HSP_RATING_LINEAR	4.05	2.00	−0.138	0.612	0.149	−0.314	−0.061
std_H_QUIET_HSP_LINEAR	0.60	2.00	−0.140	0.575	0.373	−0.351	−0.471
std_H_RECMND_LINEAR	2.20	2.00	−0.104	0.428	0.002	−0.231	0.198

Note. est., Estimate.

Table D.5. Hospital Performance and Weights (Averages) for Timeliness of Care (4%) by Hospital Size

	Weights (%)		Performance				
Measure	CMS approach (Bayes est. × 4%)	Efficient frontier (all)	All	Critical access	Small	Medium	Large
std_ED_1B	3.50	0.85	−0.120	0.614	0.362	−0.238	−0.845
std_ED_2B	0.60	0.81	−0.143	0.683	0.372	−0.291	−0.813
std_OP_3B	0.05	0.00	0.081	−0.209	0.124	0.041	0.081
std_OP_5	0.06	0.09	0.031	0.033	0.045	0.038	−0.108
std_OP_18B	0.38	0.94	−0.214	0.733	0.400	−0.375	−1.119
std_OP_20	0.14	1.35	−0.085	0.368	0.157	−0.132	−0.514
std_OP_21	0.07	0.18	−0.055	0.267	0.122	−0.112	−0.255

Note. est., Estimate.

Table D.6. Hospital Performance and Weights (Averages) for Effectiveness of Care (4%) by Hospital Size

	Weights (%)		Performance				
Measure	CMS approach (Bayes est. × 4%)	Efficient frontier (all)	All	Critical access	Small	Medium	Large
std_IMM_2	1.14	0.14	0.216	−0.222	0.110	0.295	0.149
std_IMM_3_OP_27	0.34	0.47	0.023	0.131	0.039	−0.018	0.159
std_OP_4	0.96	0.02	0.046	0.023	0.070	0.066	−0.258
std_OP_22	0.67	3.72	−0.081	0.483	0.100	−0.098	−0.481
std_OP_23	0.42	0.00	0.027	−0.337	−0.050	0.105	−0.298
std_OP_29	1.18	0.01	0.030	−0.006	−0.052	0.073	0.053
std_OP_30	1.46	0.01	0.027	0.156	0.045	0.032	−0.038
std_PC_01	0.49	0.01	0.046	−0.014	0.004	0.059	0.066
std_VTE_6	0.61	0.00	0.030	N/A	−0.184	0.036	0.027
std_OP_33	0.25	0.00	0.015	−1.116	0.134	−0.049	0.110

Note. est., Estimate.

Table D.7. Hospital Performance and Weights (Averages) for Efficient Use of Medical Imaging (4%) by Hospital Size

	Weights (%)		Performance				
Measure	CMS approach (Bayes est. × 4%)	Efficient frontier (all)	All	Critical access	Small	Medium	Large
std_OP_8	0.03	0.11	0.048	−0.845	−0.225	0.125	0.174
std_OP_10	4.68	1.92	0.041	0.013	−0.078	0.095	0.114
std_OP_11	0.47	1.06	0.054	−0.062	−0.123	0.110	0.248
std_OP_13	0.00	0.40	−0.019	0.308	0.131	−0.024	−0.238
std_OP_14	0.01	1.08	−0.088	0.547	0.206	−0.166	−0.267

Note. est., Estimate.

Table D.8. R^2 by Group and Likelihood Approach for Balanced Empirical Bayes Estimates of CMS Hospital Scores

	R^2	
Group	Fixed	Variable
Mortality	0.944	0.894
Readmission	0.999	0.999
Safety of Care	0.999	0.999
Patient Experience	0.996	0.996
Effectiveness of Care	0.762	0.619
Efficient Use of Medical Imaging	0.914	0.906
Timeliness of Care	0.957	0.976

which corresponds to more variation in the likelihood function weights.

Endnotes

[1] According to Venkatesh et al. (2017b, page 9), the CMS sought an approach that would "accommodate changes in the included measures."

[2] See https://www.modernhealthcare.com/article/20180706/TRANSFORMATION01/180709943/cms-hospital-star-ratings-system-changes-will-be-modest.

[3] See https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/FY2018-IPPS-Final-Rule-Home-Page.html. A few acute care hospitals in the U.S. territories are missing from this list, and hence from our analysis.

References

Adler N, Friedman L, Sinuany-Stern Z (2002) Review of ranking methods in the data envelopment analysis context. *Eur. J. Oper. Res.* 140(2):249–265.

Allen R, Athanassopoulos A, Dyson RG, Thanassoulis E (1997) Weights restrictions and value judgements in data envelopment analysis: Evolution, development and future directions. *Ann. Oper. Res.* 73:13–34.

Ben-Tal A, Nemirovski A (2001) *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications* (Society for Industrial and Applied Mathematics, Philadelphia).

Castelluci M (2018a) CMS changes to hospital star ratings don't address concerns. *Modern Healthcare* (December 3), https://www.modernhealthcare.com/article/20181203/TRANSFORMATION01/181209988/cms-changes-to-hospital-star-ratings-don-t-address-concerns.

Castelluci M (2018b) CMS hospital star ratings were wrong from the start, health system finds. *Modern Healthcare* 48(25):12.

Castelluci M (2019) Hospitals hopeful big changes are coming to the CMS' star ratings. *Modern Healthcare* 49(9):6–7.

Centers for Medicaid and Medicare Services (2017) Hospital quality star ratings public comment verbatim responses. Technical Report, Centers for Medicaid and Medicare Services, Woodlawn, MD.

Chambers RG (1996) Benefit and distance functions. *J. Econom. Theory* 70(2):407–419.

Chung JW, Dahlke AR, Barnard C, DeLancey JO, Merkow RP, Bilimoria KY (2019) The Centers for Medicare and Medicaid Services hospital ratings: Pitfalls of grading on a single curve. *Health Affairs* 38(9):1523–1529.

Cooper WW, Seiford LM, Tone K (2007) *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software* (Springer, New York).

Everitt B (1984) *An Introduction to Latent Variable Models* (Chapman and Hall, London).

Everitt B, Hothorn T (2011) *An Introduction to Applied Multivariate Analysis with R* (Springer, New York).

Hollingsworth B (2008) The measurement of efficiency and productivity of healthcare delivery. *Health Econom.* 17(10):1107–1128.

Hota B, Lateef O, Webb T (2019) Commentary: Four steps would help CMS fix star-rating program's unintended flaws. *Modern Healthcare* (March 1), https://www.modernhealthcare.com/opinion-editorial/commentary-four-steps-would-help-cms-fix-star-rating-programs-unintended-flaws.

Jolliffe IT (2002) *Principal Component Analysis*, 2nd ed. (Springer, New York).

Mardia KV, Kent JT, Bibby JM (1979) *Multivariate Analysis* (Academic Press, London).

Nerenz DR, Hu J, Waterman B, Jordan J (2019) Weighting of measures in the safety of care group of the overall hospital quality star rating program: An alternative approach. *Amer. J. Medical Quality* 35(1):46–51.

Ozcan YA (2008) *Healthcare Benchmarking and Performance Evaluation* (Springer, New York).

Portela MVAS, Thanassoulis E, Simpson G (2004) Negative data in DEA: A directional distance approach applied to bank branches. *J. Oper. Res. Soc.* 55(10):1111–1121.

Rumball-Smith J, Gurvey J, Friedberg M (2018) Personalized hospital ratings—Transparency for the internet age. *New England J. Medicine* 379(9):806–807.

Shwartz M, Ren J, Peköz EA, Wang X, Cohen AB, Restuccia JD (2008) Estimating a composite measure of hospital quality from the Hospital Compare database. *Medical Care* 46(8):778–785.

Simar L, Wilson PW (1998) Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Management Sci.* 44(1):49–61.

Sosunov EA, Egorova NN, Lin HM, McCardle K, Sharma V, Gelijns AC, Moskowitz AJ (2016) The impact of hospital size on CMS hospital profiling. *Medical Care* 54(4):373–379.

Venkatesh AK, Bernheim SM, Qin L, Bao H, Simoes J, Norton E, Wing M, Glennon G, Shah R, Herrin J, et al. (2017a) Overall hospital quality star rating on Hospital Compare: December 2017 updates and specifications report. Technical report, Yale New Haven Health Services Corporation/Center for Outcomes Research and Evaluation, New Haven, CT.

Venkatesh AK, Bernheim SM, Qin L, Bao H, Simoes J, Wing M, Norton E, Glennon G, Shah R, Herrin J, et al. (2017b) Overall hospital quality star rating on Hospital Compare: Methodology report (version 3.0). Technical report, Yale New Haven Health Services Corporation and Center for Outcomes Research and Evaluation, New Haven, CT.

Daniel Adelman is the Charles I. Clough Jr. Professor of Operations Management at The University of Chicago Booth School of Business, where he leads the Healthcare Analytics Laboratory. He conducts methodological research in various areas of application, such as in revenue management and pricing, the electricity smart grid, and healthcare delivery.