Research Article

Statistics in Medicine

Received 7 November 2011, Accepted 27 November 2012 Published online 26 December 2012 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.5710

# Physician performance assessment using a composite quality index

**Kaibo Liu,[a] Shabnam Jain[b] and Jianjun Shi[a*†]**

**Assessing physician performance is important for the purposes of measuring and improving quality of service and reducing healthcare delivery costs. In recent years, physician performance scorecards have been used to provide feedback on individual measures; however, one key challenge is how to develop a composite quality index that combines multiple measures for overall physician performance evaluation. A controversy arises over establishing appropriate weights to combine indicators in multiple dimensions, and cannot be easily resolved. In this study, we proposed a generic unsupervised learning approach to develop a single composite index for physician performance assessment by using non-negative principal component analysis. We developed a new algorithm named iterative quadratic programming to solve the numerical issue in the non-negative principal component analysis approach. We conducted real case studies to demonstrate the performance of the proposed method. We provided interpretations from both statistical and clinical perspectives to evaluate the developed composite ranking score in practice. In addition, we implemented the root cause assessment techniques to explain physician performance for improvement purposes. Copyright © 2012 John Wiley & Sons, Ltd.**

**Keywords:** non-negative principal component analysis (NPCA); physician performance assessment; composite quality index; iterative quadratic programming (IQP) algorithm for NPCA

## 1. Introduction

The healthcare delivery system is highly complex because it is nonlinear and dynamic and often has uncertainty about a single best approach. Since the report from the Institute of Medicine in 2001 highlighting significant problems in the quality and delivery of healthcare, there have been extensive efforts to improve healthcare quality [1]. A more recently recognized problem is the wide variation in health service delivery and spending that exists in medical practices. Practice variation, when not explained by illness-related factors such as acuity, is often due to physician preferences and practice styles. Such variation can be difficult to recognize and often does not improve outcomes but does add to healthcare costs and complexity [2, 3].

An important approach to reducing variation in physician practice is the evaluation of physician performance in multiple domains relative to their peers [4]. Performance scorecards are sometimes used to provide feedback after adjustment for severity and acuity [5]. However, most existing scorecards, while comprehensive and providing physicians with multiple dimensions of their performance, fall short in providing a single meaningful metric (composite quality index) of overall performance. Such a composite index should combine various individual measures to give physicians a realistic idea of their overall practice. Unfortunately, no natural statistical scale exists on which to combine these different measures of practice. The method that is commonly used to derive a composite index is called the "percentile rank approach"; this method constructs a single composite index by taking a weighted average of the physicians' percentile rank for each measure during the reporting period. Although this approach is simple and transparent for physicians, it has several limitations, including the following: (i) it is often difficult to choose a reasonable weighted coefficient to combine the physicians' percentile

[a] *H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, U.S.A.*
[b] *Department of Pediatrics, Emory University and Children's Healthcare of Atlanta, Atlanta, GA 30329, U.S.A.*
**Correspondence to: Jianjun Shi, H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, U.S.A.*
[†] *E-mail: Jianjun.shi@isye.gatech.edu*

rank that can be easily interpreted and justified; (ii) it is not a good practice to make physician performance comparisons based on an index without acknowledging and quantifying the uncertainty inherent in the comparison; and (iii) percentile rank information may be misleading because it masks the absolute differences between the performances of different physicians.

The main issue in construction of a composite index is in establishing appropriate weights to combine sub-indicators from multiple dimensions [6]. Although many efforts have been made in the development of a composite index, experience shows that disputes over the appropriate method of specifying weights cannot be easily resolved. Cox *et al.* [7] summarized the difficulties, which are commonly encountered when proposing weights to combine indicators to a single measure, and concluded that many published weighting schemes are either arbitrary (e.g., based upon too complex multivariate methods) or inexplicable (e.g., have a little social meaning) [7, 8].

Depending on whether a single response variable can be determined, statistical approaches to composite index development can be classified into two categories: supervised learning and unsupervised learning. The supervised learning approach assumes that there is a set of input variables, which can have some influence on one or more outputs [9]. Many regression-based methodologies have been developed to construct composite index in supervised learning case. Porter and Stern used regression analysis and quantitative modeling to develop a composite innovation index for assessing the strengths of national innovation systems [10]. Timbie *et al*. [11] fitted a generalized linear model using patients' observed quality outcomes to derive mortality weights. The regression-based model can not only quantify the relative effect of each input variable on the response variable, but also be used to predict the future outcome based on inputs. However, this method remains controversial that if the concepts to be measured could be represented by a single outcome variable then there would be no need to develop a composite indicator [8]. Furthermore, the single pre-determined response variable dominates the developed composite score [10–12]. Thus, the regression-based approach is appealing if a single response variable is desired (e.g., achieving the lowest possible return rates after an emergency room visit). In most healthcare situations, the interest is in identifying physicians who provide overall high-quality care that takes into account multiple variables of practice with varying levels of significance. To achieve this goal, the following study adopts the unsupervised learning approach.

Unsupervised learning assumes that there is no response variable and all the variables can be considered as inputs [9]. In unsupervised learning, one popular methodology is principal component analysis (PCA), which aims at detecting the maximum variability in the data through a linear combination of variables by a set of weights [13, 14]. These weights are called component loadings or factor loadings, and the constructed variables after linear combinations are called principal components. PCA can be performed by eigenvalue decomposition of the sample covariance matrix of the standardized data set. Lee *et al.* [15] applied the PCA method to explore the structure of the Korean Primary Care Assessment Tool items, which were used to assess the performance of primary care services in South Korea based on the patient's perspective. Filmer and Pritchett [16] used PCA to construct an index from asset ownership indicators to assess wealth. Coste *et al*. [17] summarized that PCA is commonly used to identify "latent" factors that underlie observed variables when constructing composite health measures.

Although the traditional PCA method is theoretically sound, it is usually limited by interpretations when applied to construct a composite index. For the purpose of physician performance assessment, the desired direction (e.g., more tests or less tests), in which physician performance would be preferred, is already known. For example, in an emergency department (ED) setting, a physician is considered lower performing if he or she orders more lab/imaging tests than his or her peers, or has a higher number of 72-h revisit cases, or his or her patients have a longer length of stay in the ED, conditional on other performance being the same. However, because the traditional PCA method is performed by eigenvalue decomposition the signs of the obtained weights can be contradictory to the clinical/pre-study intuition. To address this problem, we used a non-negative principal component analysis (NPCA) method where all weights are non-negative in this study. So far, NPCA has not been used for either physician assessment or healthcare-related assessment applications.

The following example further demonstrates the necessity of using the NPCA approach for performance assessment applications. Assume that only two dimensional performance measures "revisit 72 hours" and "labs" are taken into consideration for performance assessment, in which the detailed variable meanings are summarized in Table I. These two variables are likely to be negatively correlated in clinical practice. The data cluster of physician performance is enclosed by the ellipse as shown in Figure 1. Only the patient with mid-level (urgent) triage acuity is considered on the basis of a standardized triage system. The data of three physicians A, B and C are collected and used for the performance

**Table I.** Variable descriptions and data types.

| Symbol | Description | Data types |
|---|---|---|
| Study group | Category of type of visit the patient fits into based on presenting complaint | Discrete (1–4) |
| Attending physician | Physician ID based on attending provider | Discrete (1–98) |
| Disposition attending physician | Physician ID logged for the "Time to Disposition" (only for patients who had their disposition decided by a second provider). | Discrete (1–97) |
| Md to exit minutes | Length of emergency department stay in minutes. | Nominal |
| Revisit 72 hours | Patient returns within 72 h of a previous visit for the same condition. | Binary |
| Admitted to hospital | Based on exit code. Admitted if exit code is admission, ICU. | Binary |
| Labs (lab tests) | Count of number of tests performed for the following: Basic Metabolic panel, comprehensive metabolic panel, complete blood count with differential count, blood culture, C-reactive protein. | Discrete (0–6) |
| Abdominal/pelvic CT scan | Did patient receive CT scan of the abdomen or pelvis? | Binary |
| Head CT scan | Count of number of CT scan of the head? | Discrete (0–2) |
| Chest X-ray | Count of number of X-ray of the chest? | Discrete (0–2) |
| Abdominal X-ray | Count of number of X-ray of the abdomen? | Discrete (0–3) |
| Intravenous antibiotics | Did patient receive one of the commonly used intravenous antibiotics during the ED visit? | Binary |
| Intravenous antiemetic | Did patient receive a costly intravenous anti-vomiting medication during the ED visit? | Binary |
| Intravenous fluids | Did patient receive one of the commonly used intravenous fluids during the ED visit? | Binary |

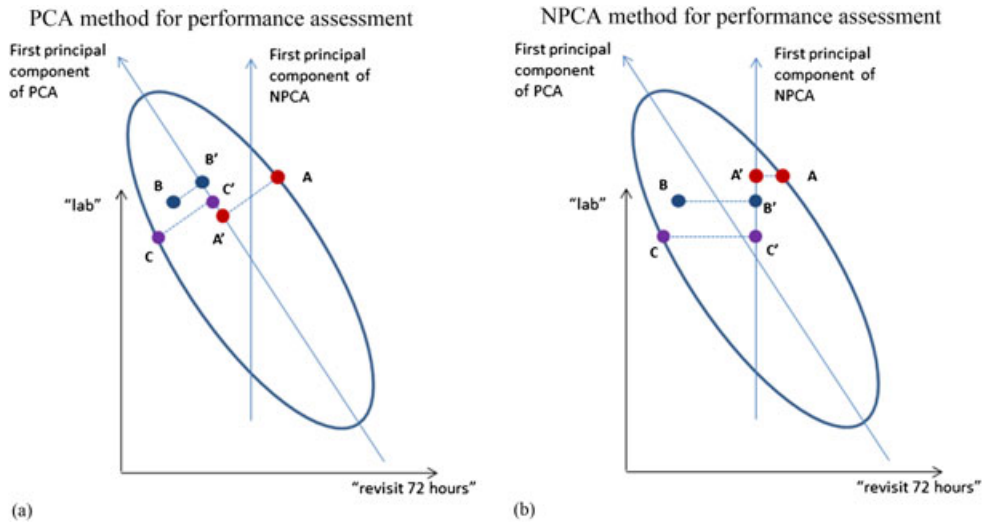


**Figure 1.** An illustrative example to show the necessity of using the NPCA approach for performance ranking. (a) PCA method and (b) NPCA method for performance assessment. The ellipse represents the data cluster of physician performance. The solid line along the major axis of the ellipse represents the first principal component of PCA. The vertical solid line represents the first principal component of NPCA. The dashed lines represent the projections of the data points A–C onto the first principal components of PCA and NPCA.

ranking. Physician A is considered as the lowest-performing physician because physician A spends the most resources and has the highest number of revisits within 72 h. Similarly, physician C is considered as the highest-performing physician because physician C spends the least resources and has the lowest number of revisits within 72 h. Therefore, the ranking scores among these three physicians should be C>B>A. However, if the PCA method is used for performance assessment, the ranking scores can be obtained by projecting the data points A, B and C onto the "first principal component of PCA" axis, which are shown as A', B' and C' in Figure 1(a). As a result, the PCA approach leads to a conclusion that physician C is the median rank physician, which is contradictory to the clinical interpretations.

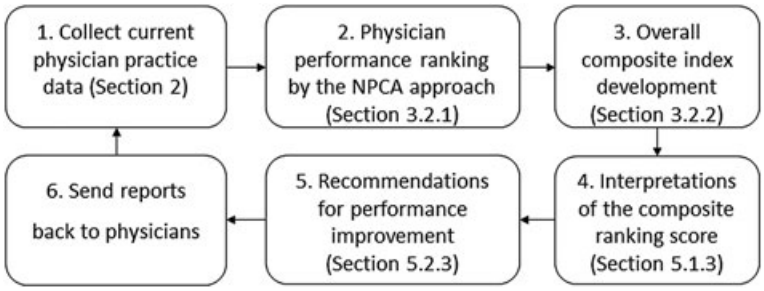


**Figure 2.** Flow chart for physician performance assessment and improvement.

On the contrary, if the NPCA approach is used for the performance assessment, the ranking scores can be obtained by projecting the data points A, B and C onto the "first principal component of NPCA" axis as shown in Figure 1(b). Because the more the resources are used or the higher the number of revisit cases are, the lower the performance should be. A negative sign is introduced after calculating the first principal component as the ranking score in the NPCA approach. In this way, the ranking scores are C>B>A. Thus, this example elaborates the necessity of using the NPCA method to achieve a meaningful performance ranking result, which is consistent with clinical interpretations.

The objective of this paper is to propose a generic approach to develop an effective composite index to identify high-performing physicians on multiple dimensions. An effective composite index should have the following four characteristics: distinguishable, interpretable, obtainable and controllable. Our goal is to derive a composite index that will have these four characteristics. An overview of the proposed strategy is elaborated in Figure 2. This flow chart in Figure 2 will be implemented in practice for performance assessment and improvement purposes.

The rest of the paper is organized as follows: Section 2 introduces the plan for data set collection in this study. Section 3 presents a generic methodology to develop a composite index for physician performance assessment by the NPCA approach. Section 4 develops an iterative quadratic programming (IQP) algorithm to address the numerical issue in the NPCA approach. Section 5 performs a case study based on the real data set discussed in Section 2 to illustrate the proposed methodology. We provided interpretations from both statistical and clinical perspectives to evaluate the developed composite index in a practical sense. Next, we implemented root cause assessment techniques to facilitate the interpretations of physician performance and target improvement dimensions. In addition, we also conducted an uncertainty analysis to understand the impacts of data uncertainty on the constructed performance composite score. Finally, Section 6 draws a conclusion.

## 2. Plan for data set collection

### *2.1. Data set description*

In this study, we collected cross-sectional data on physician practice patterns in the EDs of two pediatric hospitals. We included a subset of all ED patients with four of the most common pediatric conditions (acute gastroenteritis "AGE," "Fever," head injury "HI" and respiratory illness "RI") for analysis. We used a standardized triage system, called the Emergency Severity Index, in the EDs for triage assessment to assign acuity [18]. In this study, risk adjustment is achieved by considering only the mid-level (urgent) triage acuity within the aforementioned four conditions, which account for about one third of the total annual ED visits. These mid-acuity patients also account for the highest variation in practice. The study data set includes a total of 26,366 patient records over a 15-month period. On the basis of the standardized triage system, we assume that there are only small differences among patients within the same condition and triage acuity level. In addition, because a patient presents to be seen without appointment when an emergency condition happens and patients are seen in order of acuity and arrival by the first available physician, physicians are considered to be assigned to patients at random.

### *2.2. Study variables*

In this study, we identified and collected an indicator variable, "revisit 72 hours" defined as return to the ED within 72 h for the same condition. Although mortality is often used as an outcome indicator in many

clinical settings [11, 19, 20], in a pediatric ED setting, mortality rate is extremely low and therefore is not a useful indicator for the quality of care. On the other hand, unplanned return to the ED within 72 h of a visit, a marker of potentially unmet patient care needs in the first visit, is a commonly used quality metric in EDs. Furthermore, there is potentially a tradeoff between resources use at the first visit and a subsequent return visit within 72 h. In other words, return visits can be monitored as a balancing measure for initiatives aimed at reducing resources use. In addition, "md to exit minutes", which is measured as the time interval (in minutes) between a physician selecting a patient to be seen and the patient exiting from the ED, is used as a measure of the ED length of stay that can be influenced by physician practice decisions. The ED length of stay also represents a major component of ED patient satisfaction ratings. Finally, an important variable controlled by physicians in ED practice is the resources used for providing care—such resources include lab tests (e.g., blood counts and blood chemistries), radiographic studies (e.g., X-rays and CT scans), as well as therapeutic interventions (e.g., intravenous antibiotics and fluids). Additionally, a decision to admit the patient to the hospital uses the hospital bed as a major resource. Significant variation has been shown in the use of resources in the ED even after severity adjustment for case mix [21]. The detailed variable descriptions and data types are shown in Table I.

### *2.3. Data set exclusion criteria*

A single provider sees and dispositions most patients in the ED setting. However, at times, a physician may initiate the patient's treatment plan and then another physician may follow up and decide disposition of the patient. Because the number of cases seen by more than one physician in the data set is relatively small (791 cases accounting for 3% of the total sample), we chose to exclude those records where "attending physician" was different from "disposition-attending physician." Little and Rubin [22] have discussed and analyzed this approach. After removing these records as noted, 25,575 patient records seen by 97 physicians were included in the study.

In order to achieve a reliable assessment for physician performance in each study group, during each reporting period, only physicians with a minimum required number of patients, which was chosen to be 10 in this paper, were included in the following data analysis. Thus, we had a final data set of 25,316 patient records.

## 3. Problem formulation and approach

### *3.1. State-of-the-art on NPCA*

In this study, we adopted the NPCA approach for physician performance assessment because of the reasons mentioned in Section 1. So far, there are no published studies that discuss how to apply the NPCA approach for composite index development.

Non-negativity is a desirable property in many areas, such as economics, biology and healthcare. In the literature of composite index development, it is a standard practice to perform rotation of factor loadings that are derived from PCA to achieve non-negativity property [23]. "Varimax rotation" [24] is the most popular rotation method by far. However, because "varimax rotation" tries to maximize the variance of factor loadings, there is no guarantee that the variance of physician practice is also maximized after performing the rotation. On the contrary, imposing the non-negativity constraints directly on PCA can not only maintain the property of explaining the most variability of data sets, but also remove partial cancellations in linear combinations and make data representation that consists of only additive components [25]. Thus, this property is convenient for interpretation of the constructed composite index as a weighted sum of the contributions from each individual sub-indicator.

Zass and Shashua [26], who considered both the non-negativity constraints and sparseness of the factor loadings, first proposed the concept of NPCA. The sparseness constraint is to control the number of nonzero elements in the component loadings. The sparser the component loadings are, the more the number of zero elements, and vice versa. However, imposing non-negativity constraints on PCA usually leads to sparse representation of loading vectors [25], and there is no theoretical reason why sparseness should be in factor loadings. Therefore, sparseness constraints are not considered in this study.

In recent years, several algorithms have been proposed and intended to solve the NPCA problem. An algorithm called non-negative sparse PCA (NSPCA) was first developed in [26] and aimed to solve non-negativity and sparseness at the same time. However, this algorithm requires a good initial estimation of factor loadings before searching starts and only guarantees a local optimal solution. Duong and

Duong [27] proposed another NSPCA method by relaxing several constraints and using re-weighted $L_1$ minimization technique. Sigg and Buhmann [28] presented an algorithm named emPCA, which is based on expectation maximization for probabilistic PCA. The emPCA is intended to solve the problem with less computational complexities. Lipovetsky [29] studied the similarities between PCA and singular value decomposition to obtain non-negative loadings. Recently, Han [25] considered a NPCA-based support vector machine algorithm for high-performance proteomic pattern discovery. Although many methodologies have been developed so far, they are either aimed at solving the NPCA problem via relaxing several constraints or ended up with a local optimal solution. In this paper, we propose a new method which is called IQP algorithm to globally solve the NPCA problem. This new method overcomes the aforementioned limitations and reaches the optimal solutions of NPCA.

### 3.2. *Problem formulation by the NPCA approach*

For compactness, variable notations and meanings are defined as follows:

$I$: the total number of study groups
$i$: the index of the study group
$K_i$: the total number of physicians participating in study group $i$
$k_i$: the index of the physician in study group $i$
$J_i$: the total number of patients belonging to study group $i$
$j_i$: the index of the patient in study group $i$
$n_{k_i}$: the total number of patients treated by physician $k$ in study group $i$

*3.2.1. Formulation for physician performance ranking in each study group.* In each study group $i$, let $\boldsymbol{x_{i1}} \ldots \boldsymbol{x_{iK_i}} \in R^d$ be column vectors of a zero mean collection of data points, arranged as the rows of the matrix $\boldsymbol{X_i} \in R^{K_i \times d}$. Let $\boldsymbol{\beta_{i1}} \ldots \boldsymbol{\beta_{ih}} \in R^d$ be the desired factor loadings, arranged as the columns of the matrix $\boldsymbol{\beta_i} \in R^{d \times h}$ and $h$ be the number of interested principal components. In $\boldsymbol{X_i} \in R^{K_i \times d}$, each row represents the mean performance of each physician, and the dimension $d$ refers to the number of performance measurements that are taken into consideration. Without loss of generality, we assume the data set has been standardized as in the general procedure of PCA, because the PCA method is sensitive to the scaling of variables. Mathematically, the NPCA approach aims at solving the following optimization problem [26]:

$$\max_{\boldsymbol{\beta_i}} \quad Z_i = \frac{1}{2} \|\boldsymbol{X_i \beta_i}\|_F^2 \quad s.t. \quad \boldsymbol{\beta_i^T \beta_i} = \boldsymbol{I}, \quad \boldsymbol{\beta_i} \geqslant \boldsymbol{0} \tag{1}$$

where $\|\boldsymbol{X_i \beta_i}\|_F^2 = \sum_{l=1}^{K_i} \sum_{j=1}^{h} a_{lj}^2$ is the square of Frobenius norm and $a_{lj}$ is the element of matrix $\boldsymbol{X_i \beta_i}$ in the $l$th row and the $j$th column, $\boldsymbol{I}$ is an identity matrix and $Z_i$ is the value of the objective function.

In the traditional PCA approach, the constraint $\boldsymbol{\beta_i^T \beta_i} = \boldsymbol{I}$ ensures all factor loadings are orthogonal to each other. However, in the NPCA approach, these two constraints $\boldsymbol{\beta_i^T \beta_i} = \boldsymbol{I}$ and $\boldsymbol{\beta_i} \geqslant \boldsymbol{0}$ indicate that each row of $\boldsymbol{\beta_i}$ contains at most one nonzero element. Because the first principal component accounts for the maximum variability of the data set and our focus is on developing a single composite index for physician performance assessment, we confine our interest of dimension of $\boldsymbol{\beta_i}$ to be $d \times 1$. In other words, we will find and use the first principal component for physician performance evaluation. Accordingly, after ignoring constant terms in the objective function, the problem is reformulated as

$$\max_{\boldsymbol{\beta_{i1}}} \quad Z_i = \boldsymbol{\beta_{i1}^T X_i^T X_i \beta_{i1}} \quad s.t. \quad \boldsymbol{\beta_{i1}^T \beta_{i1}} = 1, \quad \boldsymbol{\beta_{i1}} \geqslant \boldsymbol{0} \tag{2}$$

where $\boldsymbol{\beta_{i1}}$ is the first component loading. Denote the optimal solution of equation (2) as ${}^{(2)}\boldsymbol{\beta_{i1}^*}$ and the corresponding optimal value of the objective function as ${}^{(2)}Z_i^*$. Clearly, so far the problem cannot be solved numerically by quadratic programming because of the nonlinear constraint $\boldsymbol{\beta_{i1}^T \beta_{i1}} = 1$.

*3.2.2. Formulation for overall composite index development.* We first adopt the NPCA approach to rank physician performance in each study group $i$, where the ranking score $\boldsymbol{RS_i} \in R^{K_i \times 1}$ can be obtained by calculating the negative value of the first principal component, $-\boldsymbol{X_i^{(2)} \beta_{i1}^*}$. Each row of $\boldsymbol{RS_i}$ represents the ranking score $RS_{k_i}$ for each physician $k$ in each study group $i$. Because the higher the resources spent, the poorer the performance should be, and thus we manually add a negative sign before $\boldsymbol{X_i^{(2)} \beta_{i1}^*}$

in order to be consistent with our intuition and convenient for interpretations. In this way, the physician with the highest ranking score is the best-performing physician. The overall composite score $RS_k^o$ for each physician $k$ can be obtained by taking a weighted average of the ranking score $RS_{k_i}$ across different study groups, where the weight for $RS_{k_i}$, $W_{k_i}$, is proportional to the number of patients treated by each physician $k$ in each study group $i$, and thus $W_{k_i} = n_{k_i} / \sum_{i=1}^{I} n_{k_i}$.

## 4. Iterative quadratic programming (IQP) algorithm

To address the numerical problem in Section 3.2, we relax equation (2) by adding a penalized parameter $\alpha > 0$ with a box constraint:

$$\max_{\boldsymbol{\beta_{i1}}} \quad Z_i = \boldsymbol{\beta_{i1}^T X_i^T X_i \beta_{i1}} - \alpha \left( \boldsymbol{\beta_{i1}^T \beta_{i1}} - 1 \right) \quad s.t. \quad \mathbf{1} \geqslant \boldsymbol{\beta_{i1}} \geqslant \mathbf{0} \tag{3}$$

This penalized parameter $\alpha$ is intended to control the unit norm of the component loading [26, 30]. The linear constraint $\mathbf{1} \geqslant \boldsymbol{\beta_{i1}} \geqslant \mathbf{0}$ resulted from the constraints $\left( \boldsymbol{\beta_{i1}^T \beta_{i1}} = 1, \boldsymbol{\beta_{i1}} \geqslant \mathbf{0} \right)$ in equation (2), which indicate each element of $\boldsymbol{\beta_{i1}}$ should be less than 1. Furthermore, equation (1) can be reformulated as follows:

$$\min_{\boldsymbol{\beta_{i1}}} \quad Z_i = -\boldsymbol{\beta_{i1}^T} \left( \boldsymbol{X_i^T X_i} - \alpha \boldsymbol{I} \right) \boldsymbol{\beta_{i1}} \quad s.t. \quad \mathbf{1} \geqslant \boldsymbol{\beta_{i1}} \geqslant \mathbf{0} \tag{4}$$

Denote ${}^{(4)}_{\alpha} Z_i^*$ as the optimal value by solving equation (4) at given $\alpha$ value and ${}^{(4)}_{\alpha} \boldsymbol{\beta}_{\boldsymbol{i1}}^*$ as the corresponding optimal solution. As $\alpha$ in equation (4) increases from 0 to a large number, $\boldsymbol{X_i^T X_i} - \alpha \boldsymbol{I}$ will gradually become a negative definite matrix (e.g., when $\alpha >$ the largest eigenvalue of $\boldsymbol{X_i^T X_i}$). In this case, equation (4) will achieve minimum at $\boldsymbol{\beta_{i1}} = \mathbf{0}$ because the optimal value ${}^{(4)}_{\alpha} Z_i^*$ is upper bounded by 0.

*Proposition 1*
For any *discretization step-size* parameter $P > 0$, there is a "boundary point" $\alpha_B$ such that ${}^{(4)}_{\alpha} Z_i^* < 0$ when $\alpha = \alpha_B$, and ${}^{(4)}_{\alpha} Z_i^* = 0$ for $\forall \alpha \geqslant \alpha_B + P$. In addition, ${}^{(4)}_{\alpha_B} \boldsymbol{\beta}_{\boldsymbol{i1}}^* / norm \left( {}^{(4)}_{\alpha_B} \boldsymbol{\beta}_{\boldsymbol{i1}}^* \right)$ will converge to ${}^{(2)} \boldsymbol{\beta}_{\boldsymbol{i1}}^*$ as $P$ goes to 0.

*Proof*
See Appendix A. □

In proposition 1, $P$ is the *discretization step-size* parameter, which represents the step-size of the discretized $\alpha$ value and controls the accuracy of the approximation of ${}^{(4)}_{\alpha_B} \boldsymbol{\beta}_{\boldsymbol{i1}}^* / norm \left( {}^{(4)}_{\alpha_B} \boldsymbol{\beta}_{\boldsymbol{i1}}^* \right)$ to ${}^{(2)} \boldsymbol{\beta}_{\boldsymbol{i1}}^*$. Although $\alpha$ is a continuous variable, we can only use a "close" discrete solution to approximate the optimal value of $\alpha$ because of the finite number of evaluations when searching for $\alpha_B$. The smaller the $P$ value, the better the approximation will be. Proposition 1 indicates that if we increase $\alpha$ to be as large as possible and under the constraint that there always exists a nonzero solution in equation (4), the normalized optimal solution ${}^{(4)} \boldsymbol{\beta}_{\boldsymbol{i1}}^* / norm \left( {}^{(4)} \boldsymbol{\beta}_{\boldsymbol{i1}}^* \right)$ will converge to the optimal solution ${}^{(2)} \boldsymbol{\beta}_{\boldsymbol{i1}}^*$. Thus, instead of solving equation (2) directly, solving equation (4) at $\alpha = \alpha_B$ provides an alternative choice. Now, the problem arises from two aspects: how to solve the quadratic programming in equation (4) at given $\alpha$ and how to find the "boundary point" $\alpha_B$.

When $\alpha$ is between the smallest and the largest eigenvalue of $\boldsymbol{X_i^T X_i}$, equation (4) becomes a nonconvex quadratic programming problem with box constraints. Various methods have been proposed to find the global optimal solution of this optimization problem. De Angelis *et al.* [33] gave a nice literature review on globally solving quadratic programming with box constraints. One of the best effective methods proposed so far is the semi-definite-based finite branch-and-bound algorithm by developing a semi-definite programming relaxation that can better exploit branching information, while still maintaining a compact size [34]. Based on the finite branch-and-bound scheme, Chen and Burer recently proposed a new method for globally solving nonconvex quadratic programming [35], in which relaxations are derived from completely positive and doubly non-negative programs as specified by Burer [36]. Empirical studies have shown that this method outperforms the one proposed by Burer and

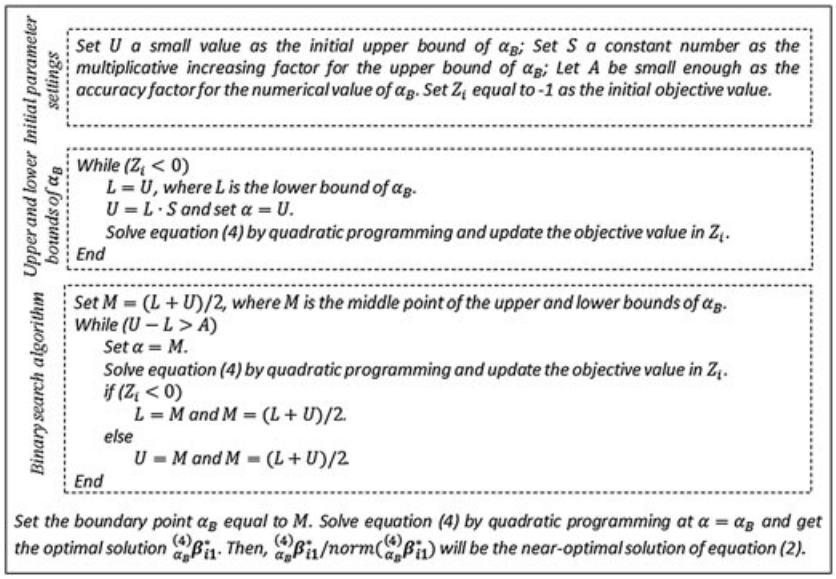


**Figure 3.** Iterative quadratic programming (IQP) algorithm for the NPCA approach.

Vandenbussche [34] in terms of computation time. If the dimension of decision variables is relatively small, some deterministic global optimization algorithms can also be applied to obtain the global optimal solution (e.g., the BARON package [37]). In this research, we implement the method proposed by Chen and Burer [35] to find the global optimal solution of the quadratic programming at given $\alpha$ value in equation (4). The code is publicly available online [38].

Figure 3 demonstrates how the IQP algorithm searches for the boundary point $\alpha_B$. Intuitively, the IQP algorithm sequentially narrows down the lower and upper limits of the boundary point under the constraint that there always exists a nonzero solution in equation (4). The search stops when the difference between these two limits is insignificant. The procedure is inspired by the binary search algorithm. The proposed IQP algorithm has the following advantage: (i) The optimal solution can be obtained automatically without identifying an initial estimation of the factor loading. (ii) The component loading derived by the IQP algorithm in Figure 3 can be proved to converge to the global optimum for equation (2) as $P$ goes to 0 by proposition 1.

When the ranking scores of two physicians are the same, we may also consider using other principal components to further distinguish their performance. It is worth noting that the proposed method can be modified to sequentially derive other factor loadings. For example, let us assume that we have already obtained the first factor loading $\boldsymbol{f_{i1}}$ by the IQP algorithm in Figure 3. Record the index of the nonzero element of $\boldsymbol{f_{i1}}$ into $\boldsymbol{H_{i1}}$. Because the second component loading is supposed to be orthogonal to the first one, equation (4) can be modified as follows:

$$\min_{\boldsymbol{\beta_{i2}}} \quad Z_i = -\boldsymbol{\beta}_{\boldsymbol{i2}}^{\boldsymbol{T}}\left(\boldsymbol{X}_{\boldsymbol{i}}^{\boldsymbol{T}}\boldsymbol{X_i} - \alpha\boldsymbol{I}\right)\boldsymbol{\beta_{i2}} \quad s.t. \quad \mathbf{1} \geqslant \boldsymbol{\beta_{i2}} \geqslant \mathbf{0} \text{ and } \boldsymbol{e}_{\boldsymbol{h}}^{\boldsymbol{T}}\boldsymbol{\beta_{i2}} = 0, h \in \boldsymbol{H_{i1}}, \qquad (5)$$

where $\boldsymbol{e_h} = [0 \ldots 1 \ldots 0]^T$ with only the $h$th element equal to 1. Then, the second factor loading can be acquired by implementing the IQP algorithm in Figure 3. Similarly, other component loadings can be derived sequentially.

## 5. Case studies

In this section, a real data set described in Section 2 is used to demonstrate the proposed methodology for physician performance assessment by using the NPCA approach. Detailed analyses and interpretations from both statistical and clinical perspectives are also provided. Finally, recommendations for physician performance improvement by implementing the root cause assessment techniques are presented, which are intended to be combined with the developed composite index to better serve the healthcare system.

### 5.1. *Physician performance assessment by the NPCA approach*

*5.1.1. Implementation and evaluation of the IQP algorithm for the NPCA approach.* As presented in Section 2.3, the study includes 25,316 records of patient visits in four different study groups seen by 97 physicians. Eleven performance measurements, also called sub-indicators ("md to exit minutes," "revisit 72 hours," "admitted to hospital," "labs," "abdominal/pelvic CT scan," "head CT scan," "chest X-ray," "abdominal X-ray," "intravenous antibiotics," "intravenous fluids" and "intravenous antiemetic") in Table I are taken into consideration. Because not all physicians see patients in each study group, in our study, $K_1 = 75$, $K_2 = 83$, $K_3 = 69$ and $K_4 = 90$. As discussed in Section 3.2.2, we will first adopt the NPCA approach to rank physician performance in each study group and then obtain an overall composite score by taking a weighted average of the ranking score across different study groups. For each study group $i$, recall that each row of $\boldsymbol{X_i} \in R^{K_i \times 11}$ in equation (4) represents the mean performance of each physician. In addition, we are only interested in deriving the first principal component for ranking purpose.

In the case study, the initial parameter settings in Figure 3 for each study group $i$ are given as $U = 1$, $S = 10$, $A = 10^{-6}$ and $Z_i = -1$. The effectiveness of our developed IQP algorithm is compared with the NSPCA method proposed by Zass and Shashua [26]. Furthermore, the PCA method is also adopted to evaluate these two methods. The NSPCA approach proposed by Zass and Shashua [26] depends on several parameters, especially requiring a good starting point. However, the paper [26] does not provide a generic guideline on how to set the initial starting point. Therefore, we use the positive part of the first factor loadings from PCA as the initial estimation. A MatLab implementation of the NSPCA method is available on Ron Zass's webpage [39]. Table II summarizes the detailed factor loadings of these three methods. As expected, the standard PCA method will always achieve the best result in terms of explaining the most variability of the data set because factor loadings are not subjected to the non-negativity constraints.

On the basis of the comparison result in Table II, the following conclusions can be drawn:

(1) The IQP algorithm does not require the specifications of initial values, whereas in the NSPCA method, a poor choice of the initial values may lead to sub-optimal results as shown in Table II. We have also changed the initial points of the NSPCA method several times, and the empirical results have shown that the final solution obtained from NSPCA is very sensitive to the initialization, which is a common problem in high dimensional optimization.
(2) The IQP algorithm can find the global optimal solution whereas the NSPCA method may end up in a local optimal solution. Thus, the results from the IQP algorithm are more reliable than the results from the NSPCA method. In this particular data set, a large difference between the methods is not seen, but due to the aforementioned reasons, there may be a bigger difference in other data sets. In general, the IQP algorithm is preferred as the global solver than the NSPCA method as a local solver.
(3) In study group Fever, where the first factor loadings of the PCA method are all positive, the IQP algorithm achieves the same result as the PCA method, which further validates our proposed method. Because the starting point is set as the positive part of the first factor loadings from PCA, the NSPCA method stops searching at the starting point.
(4) The IQP algorithm has a better performance than the NSPCA method when more discrepancies in the non-negativity constraints are found in the first component loadings of the PCA method. For example, in study group AGE, because only the coefficient of "revisit 72 hours" violates the non-negativity constraints and the magnitude of this coefficient is relatively small (which indicates this variable is insignificant in explaining the variability of the data set), the result of the IQP algorithm is comparable with the NSPCA method (i.e., the difference in the "Explained Variance" is only showed in the fifth digits after the decimal point). On the contrary, because the magnitudes of the negative coefficient of "revisit 72 hours" in study group HI and "head CT scan" in study group RI are relative large, the IQP algorithm explains more variability in these two study groups than the NSPCA method.
(5) The IQP algorithm for the NPCA approach can also lead to sparse representation of factor loadings. The negative coefficients obtained by the PCA method will be automatically reduced to 0 in the IQP algorithm. Thus, more weights will be assigned to other coefficients in order to maintain the unit norm of the factor loadings. In this situation, the IQP algorithm can decide the best distribution of weights to explain the most variability.

**Table II.** First component loadings of the PCA, the NSPCA and the IQP algorithm.

| | PCA | | | | NSPCA | | | | IQP | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Study group | AGE (1) | Fever (2) | HI (3) | RI (4) | AGE (1) | Fever (2) | HI (3) | RI (4) | AGE (1) | Fever (2) | HI (3) | RI (4) |
| Md to exit minutes | 0.4328 | 0.4211 | 0.4661 | 0.4754 | 0.4328 | 0.4211 | 0.5133 | 0.4901 | 0.4333 | 0.4211 | 0.4672 | 0.4768 |
| Revisit 72 hours | −0.0097 | 0.0807 | −0.2150 | 0.2793 | 0 | 0.0807 | 0 | 0.2793 | 0.0000 | 0.0807 | 0.0000 | 0.2773 |
| Admitted to hospital | 0.3444 | 0.2978 | 0.3175 | 0.4341 | 0.3445 | 0.2978 | 0.3176 | 0.4341 | 0.3444 | 0.2978 | 0.3217 | 0.4345 |
| Labs | 0.3841 | 0.3685 | 0.3679 | 0.2660 | 0.3841 | 0.3685 | 0.3679 | 0.2660 | 0.3843 | 0.3685 | 0.3832 | 0.2815 |
| Abdominal/pelvic CT scan | 0.2960 | 0.2923 | 0 | 0.0438 | 0.2960 | 0.2923 | 0 | 0.0438 | 0.2953 | 0.2923 | 0.0000 | 0.0423 |
| Head CT scan | 0.1887 | 0.2027 | 0.2718 | -0.1190 | 0.1887 | 0.2027 | 0.2718 | 0 | 0.1883 | 0.2027 | 0.2834 | 0.0000 |
| Chest X-ray | 0.1076 | 0.2247 | 0.1604 | 0.2690 | 0.1076 | 0.2247 | 0.1604 | 0.2690 | 0.1076 | 0.2247 | 0.1570 | 0.2673 |
| Abdominal X-ray | 0.0730 | 0.1284 | 0.0457 | 0.0601 | 0.0730 | 0.1284 | 0.0453 | 0.0601 | 0.0732 | 0.1284 | 0.0308 | 0.0702 |
| Intravenous antibiotics | 0.2496 | 0.3773 | 0 | 0.4043 | 0.2496 | 0.3773 | 0 | 0.4043 | 0.2492 | 0.3773 | 0.0000 | 0.4108 |
| Intravenous fluids | 0.4807 | 0.3988 | 0.5250 | 0.3797 | 0.4807 | 0.3988 | 0.5250 | 0.3797 | 0.4812 | 0.3988 | 0.5434 | 0.3792 |
| Intravenous antiemetic | 0.3360 | 0.3111 | 0.3508 | 0.1928 | 0.3360 | 0.3111 | 0.3508 | 0.1928 | 0.3358 | 0.3111 | 0.3609 | 0.1925 |
| **Explained variance** | 3.5632 | 4.2790 | 2.4640 | 3.0179 | 3.5630 | 4.2790 | 2.3880 | 2.9884 | 3.5630 | 4.2790 | 2.3938 | 2.9897 |

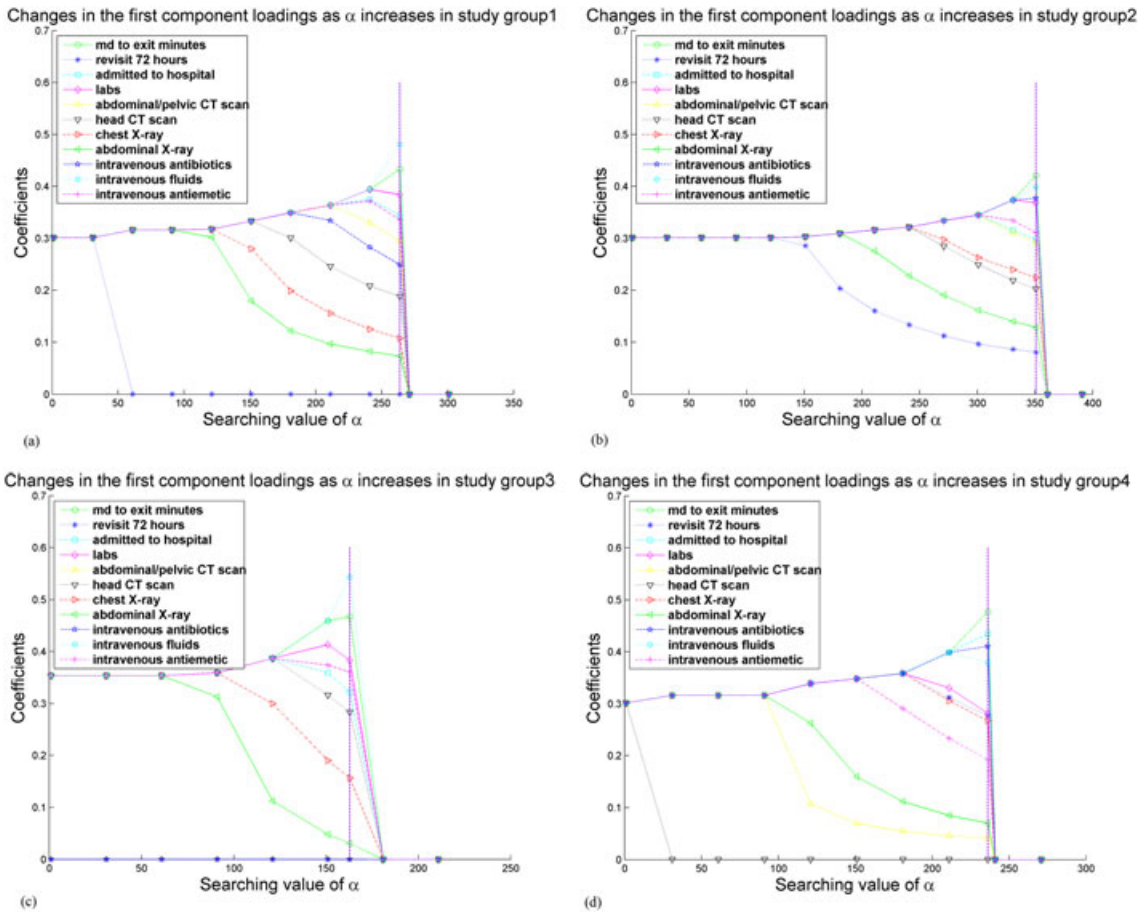


**Figure 4.** Changes in the profiles of the first component loadings as $\alpha$ increases in study groups (a) 1, (b) 2, (c) 3 and (d) 4.

Figure 4 demonstrates how the profiles of the first component loadings vary as the penalized parameter $\alpha$ in Section 4 increases. As expected, when $\alpha$ is very small, all the coefficients tend to reach the bounds of the inequality constraint $\mathbf{1} \geqslant \boldsymbol{\beta}_{i1} \geqslant \mathbf{0}$ in equation (4). Thus, after normalization, the first component loadings achieve the same value (except a few in Figure 4(c) equal to 0). As $\alpha$ becomes larger, coefficients are gradually changed to compensate for the increased penalty on the norm of the coefficients. At the boundary point $\alpha = \alpha_B$ (the intersection of the horizontal axis and the vertical line), coefficients vary in a way that explains the most variability of the data set. Once across the boundary point $\alpha_B$, all coefficients are reduced to 0.

*5.1.2. Interpretations of the component loadings of the IQP algorithm.* The values in the IQP algorithm columns in Table II reveal the differences in the weights of the first component loadings.

- The weights of "md to exit minutes," "admitted to hospital" and "labs" in the four study groups are comparably and consistently large, which indicate that the patient length of stay, the number of admission cases and the number of lab tests vary much across physicians.
- Except for the study group RI, "revisit 72 hours" is insignificant in distinguishing performance among physicians. Thus, there is an opportunity for physicians to improve performance by optimizing resources use while maintaining low return rate.
- Table II also validates some expected clinical principles regarding use of common diagnostic/imaging tests and therapeutic interventions for these four different conditions. For example, "abdominal/pelvic CT scan" is often used for patients presenting with AGE or Fever, but is seldom ordered for patients in study groups HI and RI. In Table II, the weights of "abdominal/pelvic CT scan" in study groups AGE and Fever are relatively large and therefore explain much performance variability in these two study groups, but they are nearly 0 in study groups HI and RI.

Similarly, "head CT scan" and "chest X-ray" are more commonly used tests for patients in study groups HI and RI, respectively. In our approach, the coefficient of "head CT scan" in study group HI and the coefficient of "chest X-ray" in study group RI are the largest among all study groups and are thus consistent with the clinical principles for these conditions.

*5.1.3. Construction and interpretations of the composite ranking score.* After obtaining the results in Table II, the ranking score in each study group can be calculated by the negative value of the first principal component as explained in Section 3.2.2. The physician with the highest ranking score, $-\boldsymbol{X_i}^{(2)}\boldsymbol{\beta}^*_{\boldsymbol{i1}}$, is the best-performing physician in study group $i$. In addition, an overall composite index can be acquired by taking a weighted sum of the ranking score across different study groups, where the weight is proportional to the number of patients treated by each physician in each study group.

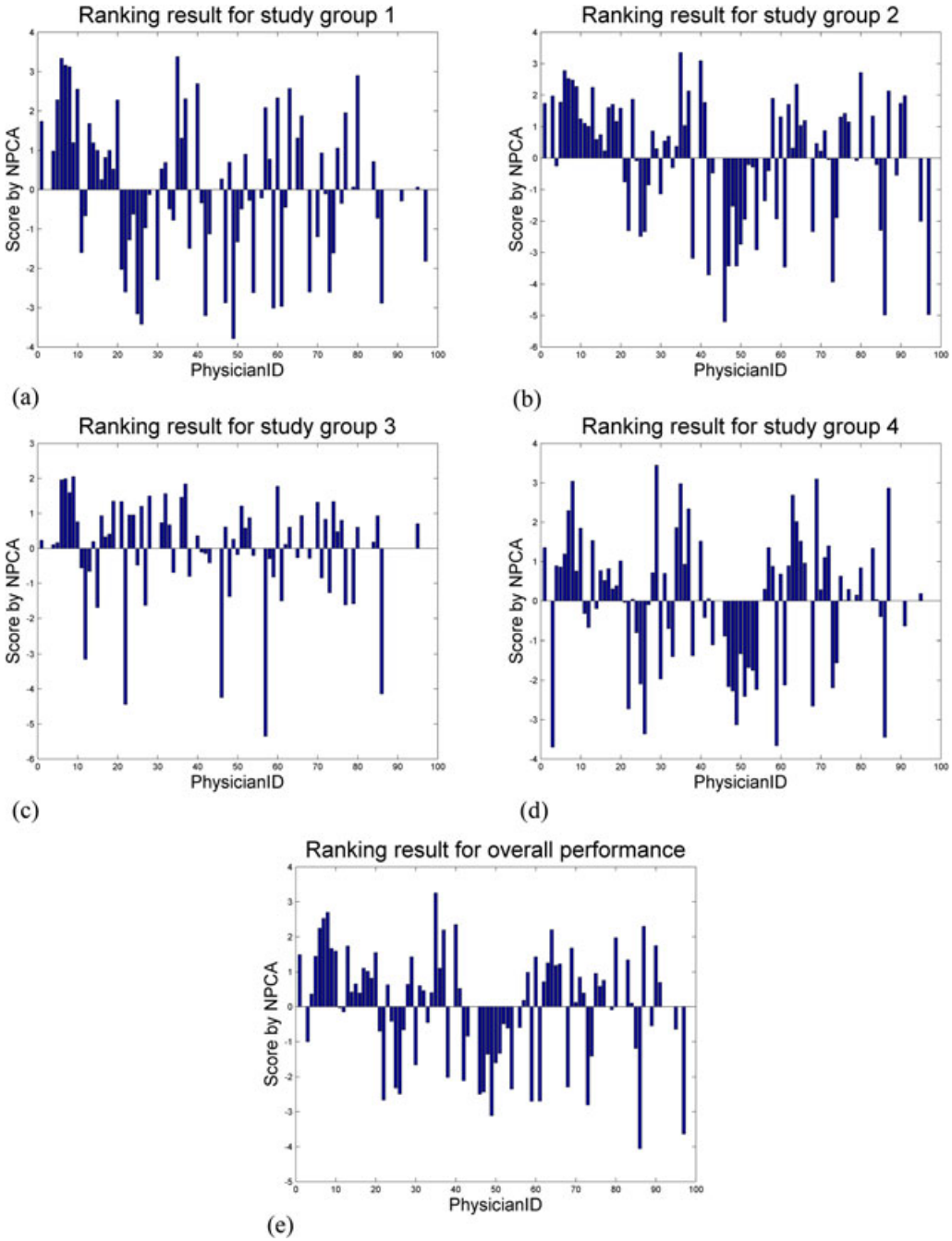


**Figure 5.** Ranking score by NPCA for physician performance assessment for study groups (a) 1, (b) 2, (c) 3 and (d) 4 and for (e) overall performance.

Figure 5(a)–(d) shows the ranking score of each physician in each study group. Each bar represents the score of one physician. Figure 5(e) draws the composite score in overall performance. The higher the score, the better the performance of the physician is. For certain physicians, the bar value is missing in certain study groups, because this physician does not participate in treating this study group. Thus, no ranking score is calculated in this case.

Table III provides the standardized mean performance and ranking scores for physicians 35, 86 and 34, who obtained the highest, lowest and median rank according to the overall composite score, respectively. A negative sign represents fewer resources used than the mean performance of all physicians, whereas a positive sign means more resources used than the mean performance of all physicians. "Resources used" includes performance in all 11 sub-indicators.

- Physician 35 participated in study groups AGE, Fever and RI, where most of the sub-indicators are associated with negative sign and large magnitude. Thus, physician 35 obtained very high scores and top ranks in these three study groups.
- Physician 86 had a longer ED length of stay and ordered more "labs," "intravenous fluids" and "intravenous antiemetic" than physicians did overall in all study groups. In addition, physician 86 consistently used more "intravenous antibiotics" than physicians did overall in study groups AGE, Fever and RI. As mentioned in Section 5.1.3, "head CT scan" and "chest X-ray" are more frequently used tests for patients in study groups HI and RI, respectively; physician 86 also used more resources in these study groups. Because physician 86 used many more resources in each study group, the physician obtained a low rank in each study group and also the lowest overall rank.
- Physician 34 participated in all study groups and obtained an overall median rank. Compared with physician 86 in terms of resources used in study group HI, physician 34 had a shorter length of stay performance and ordered more "labs," less "head CT scan," more "chest X-ray," more "abdominal X-ray," less "intravenous fluids" and less "intravenous antiemetic." Because variables with the two heaviest weights in study group HI are "md to exit minutes" and "intravenous fluids", physician 34 acquired a higher score than physician 86 in this case although more "labs," "chest X-ray" and "abdominal X-ray" were used by physician 34. In addition, although physician 34 ranked below the median in study groups AGE, Fever and HI, the physician used very few resources in study group RI, which resulted in an overall median rank.

The NPCA method takes all variables into consideration to explain the most variability of the data set while subjected to the non-negativity constraints. The first component loadings reveal different significance of variables in distinguishing physician performance. Although this approach is able to point out which dimensions the variability of performance exists in, it cannot be directly used for physician improvement. We believe that the root cause assessment techniques, which are explained in the following section, can be useful to address this problem.

### *5.2. Root cause assessment techniques for explaining physician performance and identifying areas for improvement*

*5.2.1. Root cause assessment for each performance variable.* Although a composite index of performance is able to summarize complex, multi-dimensional realities and facilitate communications, it may also mask failings in some dimensions and increase the difficulty of identifying proper remedial action [23]. In other words, the developed composite index can be used to explain physician performance, but because it cannot pinpoint the areas that a physician needs to improve in, it is not suitable to provide essential information for quality improvement. In order to provide recommendations for physician performance improvement, root cause assessment techniques are adopted in this study.

Because a standardized triage system is applied in this study to achieve risk adjustment, the differences among patients within the same study group are assumed to be minor, as mentioned in Section 2.1. Furthermore, although the data set is over a 15-month period, no interventions for performance improvement such as reports of performance assessment were issued to physicians during this period. Therefore, patients in each study group treated by a given physician can be considered as the repeated performance measurements with the assumption that physician performance does not change over this period.

On the basis of the underlying statistical distribution of each category of resources use, different confidence limits can be constructed to reveal potential improvement dimensions for each physician. The variables "md to exit minutes" and "revisit 72 hours" are influenced by many factors outside of a physician's practice. Therefore, these two indicators can be difficult to control by physicians alone and are not

**Table III.** Standardized mean performance and ranking score for physicians 35, 86, and 34.

| Physician ID | 35 | | | | 86 | | | | 34 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Study group | AGE (1) | Fever (2) | HI (3) | RI (4) | AGE (1) | Fever (2) | HI (3) | RI (4) | AGE (1) | Fever (2) | HI (3) | RI (4) |
| Md to exit minutes | −1.8364 | −2.0577 | — | −1.7474 | 0.6135 | 1.1908 | 1.3552 | 1.1528 | 2.3918 | 0.2515 | −0.1235 | −0.9555 |
| Revisit 72 hours | −0.9140 | 0.1420 | — | −1.0297 | 0.3627 | −0.6198 | −0.4407 | −0.0320 | 2.1124 | −1.2930 | −0.4407 | 0.2867 |
| Admitted to hospital | −1.5900 | −1.3327 | — | −1.5740 | 0.2813 | 0.4358 | −0.4819 | 0.5864 | 0.0734 | 0.5859 | −0.4819 | −0.6969 |
| Labs | −1.2393 | −1.1644 | — | −0.1014 | 1.0634 | 1.1286 | 2.0786 | 0.2075 | −0.1093 | −0.6859 | 2.4441 | 0.6086 |
| Abdominal/pelvic CT scan | −0.6256 | −0.5409 | — | −0.2275 | −0.6256 | 1.2735 | 0 | 1.6191 | −0.6256 | −0.5409 | 0 | −0.2275 |
| Head CT scan | −0.8876 | −0.4700 | — | −0.1872 | 0.7776 | 0.4139 | 1.0493 | −0.1872 | 1.0860 | 1.4478 | −0.7788 | −0.1872 |
| Chest X-ray | 1.2933 | −0.4421 | — | −0.2027 | 1.7840 | 2.8230 | −0.4146 | 2.4046 | −1.6510 | −1.3576 | 2.3623 | −1.0000 |
| Abdominal X-ray | −1.0363 | −0.8580 | — | −0.6091 | 1.6938 | 2.0001 | −0.2465 | −0.0079 | −0.8919 | −0.8580 | 5.9820 | −0.6091 |
| Intravenous antibiotics | 1.0640 | −1.1882 | — | −0.9111 | 3.2283 | 3.7016 | 0 | 2.8121 | 0.1822 | −0.0919 | 0 | −0.9111 |
| Intravenous fluids | −2.4151 | −1.4075 | — | −1.4782 | 0.3276 | 0.6239 | 2.8418 | 1.3547 | 0.0854 | −0.3325 | −0.5064 | −1.4782 |
| Intravenous antiemetic | −1.1127 | −0.6428 | — | −0.5252 | 2.6067 | 3.1959 | 3.0356 | 1.1176 | −0.3113 | 0.4677 | −0.2526 | −0.5252 |
| **Ranking score** | 3.3785 | 3.3535 | — | 2.9735 | −2.8864 | −4.9840 | −4.1391 | −3.4489 | −0.7785 | 0.3783 | −0.6917 | 1.8632 |
| **Rank/total number of physicians** | 1/75 | 1/83 | — | 4/80 | 69/75 | 82/83 | 66/69 | 78/80 | 52/75 | 43/83 | 54/69 | 10/80 |

considered here for the purposes of targeting improvement areas. On the basis of the characteristics of each variable in Table I, "admitted to hospital," "abdominal/pelvic CT scan," "intravenous antibiotics," "intravenous fluids" and "intravenous antiemetic" can be assumed to follow binomial distribution. On the contrary, "labs," "head CT scan," "chest X-ray" and "abdominal X-ray" can be assumed to follow Poisson distribution.

For the variables that follow binomial distribution, we assume that the probability of a patient in study group $i$ to receive each type of resource $g$ is $p_{g_i}$ and patients are independent of each other. The confidence limits can be determined as follows:

$$\mathrm{UCL}_{g_{k_i}} = \bar{p}_{g_i} + 3\sqrt{\frac{\bar{p}_{g_i}\left(1-\bar{p}_{g_i}\right)}{n_{k_i}}}, CL_{g_i} = \bar{p}_{g_i}, LCL_{g_{k_i}} = \bar{p}_{g_i} - 3\sqrt{\frac{\bar{p}_{g_i}\left(1-\bar{p}_{g_i}\right)}{n_{k_i}}}, \tag{6}$$

where $\mathrm{UCL}_{g_{k_i}}$ and $LCL_{g_{k_i}}$ are the upper and lower confidence limits for resource $g$ of physician $k$ in study group $i$, respectively. $CL_{g_i}$ is the center line for resource $g$ in study group $i$ and $\bar{p}_{g_i}$ is an

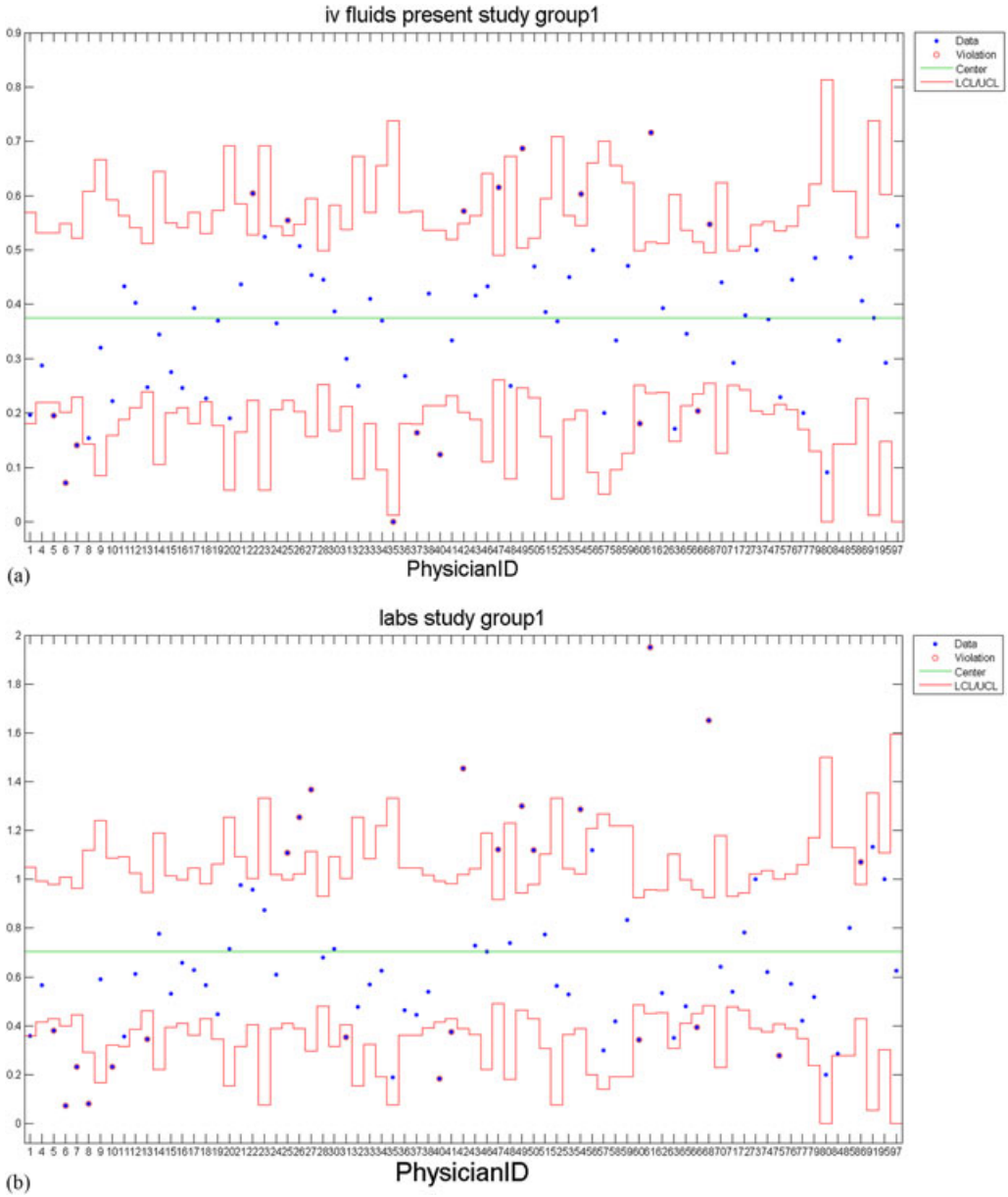


**Figure 6.** Results of root cause assessment for physician performance improvement in (a) IV fluids present and (b) labs study group 1.

estimation of $p_{g_i}$ and equals to "the total number of times that patients in study group $i$ receive resource $g$"/$J_i$. Furthermore, the confidence limits proposed here are different across physicians because of the variation of $n_{k_i}$. If the performance of physician $k$ in study group $i$ is higher than the upper confidence limit, it indicates excessive use of resources. Similarly for the variables that follow Poisson distribution, we assume that the number of times that a patient in study group $i$ receives each type of resource $g$ is $u_{g_i}$ and patients are independent of each other. The confidence limits can be determined as follows:

$$\mathrm{UCL}_{g_{k_i}} = \bar{u}_{g_i} + 3\sqrt{\frac{\bar{u}_{g_i}}{n_{k_i}}}, CL_{g_i} = \bar{u}_{g_i} LCL_{g_{k_i}} = \bar{u}_{g_i} - 3\sqrt{\frac{\bar{u}_{g_i}}{n_{k_i}}}, \tag{7}$$

where $\bar{u}_{g_i}$ is an estimation of $u_{g_i}$ and equals to "the total number of times that patients in study group $i$ receive resource $g$"/$J_i$.

*5.2.2. Implementation of root cause assessment for each performance variable.* We applied the root cause assessment techniques for each performance variable identified in the previous section. Figure 6(a) shows an example of applying root cause assessment analysis for variable "intravenous fluids" in study group AGE, which accounts for the largest coefficient of the binary type resources in the "IQP" column of Table II. Similarly, Figure 6(b) shows one example of applying root cause assessment analysis for "labs" in study group AGE, which accounts for the largest coefficient of the discrete type resources in the "IQP" column of Table II. According to which variable is out of the upper confidence limit, over-used resources can be identified in a straightforward manner for each physician in each study group. With such data available, physicians can adjust their practice to be within these confidence limits and achieve performance improvement.

*5.2.3. Interpretation and recommendation for physician performance improvement.* Although interpretations of the newly developed composite score can be achieved by comparing the mean performance of each variable in Table III across physicians, it is difficult to delineate whether the differences in the performance of two physicians are due to special causes. Root cause assessment analysis can not only be used to identify significant differences in the performance across physicians, but also point out which dimensions of performance each physician should target for improvement efforts. Table IV summarizes the resources use status of each sub-indicator for physicians 35, 86 and 34. "0" represents the variable is within confidence limits, whereas "1" (−1) represents the variable is out of the upper (lower) confidence limit. According to Table IV, interpretations of the composite score can be made as follows:

- No excessive resource use was detected in any study group that physician 35 was involved in. In addition, physician 35 used significantly less "intravenous fluids" than overall physicians did in study group AGE. Thus, physician 35 is ranked first in this study group and also at the top in other study groups.
- Many resources were identified as being used excessively by physician 86. Furthermore, "labs," "intravenous antibiotics" and "intravenous antiemetic" were noted to be above the upper confidence limits in three out of four study groups, which might imply that physician 86 tended to use these resources consistently across different study groups. Thus, physician 86 had a very low rank in all study groups as well as the lowest overall rank.
- Physician 34 used expected amounts of resources in study groups AGE, Fever and RI. However, "labs" and "abdominal X-ray" were identified as resources that were used more often than expected in study group HI; thus, physician 34 had the lowest rank in study group HI.
- Table IV shows that based on resources identified as being used excessively, physicians can adjust their practice so that their performance falls within confidence limits. For example, physician 86 may need to re-evaluate his or her practice on ordering "labs" in the first three study groups to improve his or her performance.

### *5.3. Uncertainty analysis of the composite ranking score derived from the NPCA approach*

As mentioned in Section 2.1, on the basis of the triage system, we assume that there are only small differences among patients within the same condition and triage acuity level. In addition, only physicians with at least 10 patients were included in the foregoing analysis. In this way, the data uncertainty can be reduced. However, it is inevitable that certain physicians attend to more patients than others.

**Table IV.** Resources use status for physicians 35, 86 and 34 ("0" represents the variable is within confidence limits, whereas "1" (−1) represents the variable is out of the upper (lower) confidence limit).

| Physician ID | 35 | | | | 86 | | | | 34 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Study group | AGE (1) | Fever (2) | HI (3) | RI (4) | AGE (1) | Fever (2) | HI (3) | RI (4) | AGE (1) | Fever (2) | HI (3) | RI (4) |
| Admitted to hospital | 0 | 0 | — | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Labs | 0 | 0 | — | 0 | +1 | +1 | +1 | 0 | 0 | 0 | +1 | 0 |
| Abdominal/Pelvic CT scan | 0 | 0 | — | 0 | 0 | +1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Head CT scan | 0 | 0 | — | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Chest X-ray | 0 | 0 | — | 0 | 0 | +1 | 0 | +1 | 0 | 0 | 0 | 0 |
| Abdominal X-ray | 0 | 0 | — | 0 | +1 | +1 | 0 | 0 | 0 | 0 | +1 | 0 |
| Intravenous antibiotics | 0 | 0 | — | 0 | +1 | +1 | 0 | +1 | 0 | 0 | 0 | 0 |
| Intravenous fluids | −1 | 0 | — | 0 | 0 | 0 | +1 | 0 | 0 | 0 | 0 | 0 |
| Intravenous antiemetic | 0 | 0 | — | 0 | +1 | +1 | +1 | 0 | 0 | 0 | 0 | 0 |

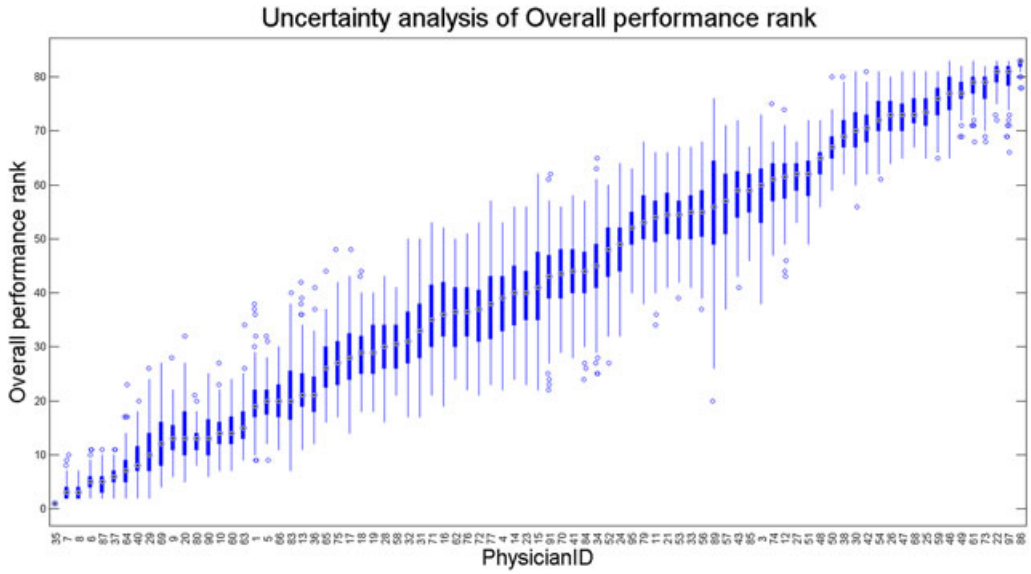


**Figure 7.** Uncertainty analysis of overall performance rank.

To evaluate the effect of the number of patients on the overall performance rank, we employ bootstrap sampling technique in the way that each physician has the same and sufficient number of patients in each study group. We repeat the experiments 100 times, and Figure 7 shows the variation in the overall performance rank.

Figure 7 shows that physicians with either high or low rank in the overall performance have less uncertainties than the physicians in the middle rank. For example, physicians 35 and 86, who obtain the highest rank and the lowest rank according to the overall composite score in Table III, have a very narrow bar length in their overall rank position. Figure 7 also shows that there is significant difference between high-performing (e.g., physician ranks in the first quantile) and low-performing (e.g., physician ranks in the last quantile) physicians. On the basis of Figure 7, we can further compare the performance of any two physicians by looking at the percentage of overlapped bar length. This uncertainty study indicates that the proposed composite index is robust to the sample size uncertainties. It also indicates that the impact of data uncertainty on the final ranking result is different from physician to physician.

## 6. Conclusion

Assessment of existing physician practice, followed by feedback to encourage appropriate changes in practice represents an important opportunity to reduce variation and improve the quality of healthcare. Physician performance scorecards increasingly being used for this purpose are comprehensive to provide feedback on multiple dimensions of practice. However, there is no natural scale on which to combine different measures to construct a composite index for overall physician performance assessment. This paper proposes a general unsupervised learning approach to develop a single composite index for physician performance assessment by the NPCA approach. Although this paper focuses on a particular problem of physician performance evaluation, the developed approach is generic, which can be extended as a guideline for composite index development in other healthcare-related or facility performance assessment applications. In addition, the IQP algorithm developed in this paper can solve the numerical problem in the NPCA approach. The developed methodology performs well on a real data set collected from two pediatric hospitals' EDs. Interpretations from both statistical and clinical perspectives are provided to evaluate the developed composite index in a practical sense. Root cause assessment analysis is applied to help physicians identify and target specific areas to improve performance in terms of resources use; this, combined with the developed composite index, can provide physicians more complete information for performance improvement efforts. Finally, uncertainty analysis is conducted for the constructed composite index derived from the NPCA approach. The developed composite index in this paper satisfies the four criteria for a good composite index, that is, it is distinguishable, interpretable, obtainable and controllable.

## Appendix A. Proofs

### Proof of proposition 1

We first prove that this proposition can be used to acquire the first component loading in PCA. In other words, we consider the following optimization problem:

$$\min_{\boldsymbol{\beta_{i1}}} Z_i = -\boldsymbol{\beta_{i1}^T}\left(\boldsymbol{X_i^T X_i} - \alpha \boldsymbol{I}\right)\boldsymbol{\beta_{i1}} \quad s.t. \quad \mathbf{1} \geqslant \boldsymbol{\beta_{i1}} \geqslant -\mathbf{1} \tag{8}$$

Denote ${}^{(8)}_{\alpha} Z_i^*$ as the optimal value by solving equation (8) at given $\alpha$ value and ${}^{(8)}_{\alpha}\boldsymbol{\beta}^*_{\boldsymbol{i1}}$ as the corresponding optimal solution. Because $\boldsymbol{X_i^T X_i}$ is a symmetric positive semi-definite matrix, and let $\left\{\boldsymbol{q}^{\mathbf{j}}_{\boldsymbol{x_i}}, \lambda^j_{x_i}\right\}$ be the eigenvectors and eigenvalues of $\boldsymbol{X_i^T X_i}$ $\left(\lambda^1_{x_i} \geqslant \ldots \geqslant \lambda^d_{x_i}\right)$, then the objective function can be rewritten as $\min_{\boldsymbol{\beta_{i1}}} Z_i = -\boldsymbol{\beta_{i1}^T Q_{X_i}^T}\left(\boldsymbol{\Lambda_{X_i}} - \alpha \boldsymbol{I}\right)\boldsymbol{Q_{X_i}}\boldsymbol{\beta_{i1}}$, where $\boldsymbol{Q^T_{X_i}} = \left[\boldsymbol{q}^{\mathbf{1}}_{\boldsymbol{x_i}}, \ldots, \boldsymbol{q}^{\mathbf{d}}_{\boldsymbol{x_i}}\right]$ and $\boldsymbol{\Lambda_{X_i}} = \mathrm{diag}\left(\lambda^1_{x_i}, \ldots \lambda^d_{x_i}\right)$. Denote the largest element of $\boldsymbol{q}^{\mathbf{1}}_{\boldsymbol{X_i}}$ as $large\left(\boldsymbol{q}^{\mathbf{1}}_{\boldsymbol{X_i}}\right)$ When $\alpha = \lambda^1_{\boldsymbol{X_i}}$, $\boldsymbol{Q^T_{X_i}}\left(\boldsymbol{\Lambda_{X_i}} - \alpha \boldsymbol{I}\right)\boldsymbol{Q_{X_i}}$ becomes a negative semi-definite matrix. Thus, ${}^{(8)}_{\lambda^1_{\boldsymbol{X_i}}} Z_i^* = 0$ if and only if $\boldsymbol{\beta_{i1}} = \mathbf{0}$ or $\frac{a}{large\left(\boldsymbol{q}^{\mathbf{1}}_{\boldsymbol{X_i}}\right)}\boldsymbol{q}^{\mathbf{1}}_{\boldsymbol{X_i}}$, where $a$ is any nonzero constant number and $-1 \leqslant a \leqslant 1$. Let $P$ be the *discretization step-size* parameter for $\alpha$, and $\alpha_B < \lambda^1_{\boldsymbol{X_i}}$ and $\alpha_B + P \geqslant \lambda^1_{\boldsymbol{X_i}}$. Then, for $\forall \alpha > \alpha_B + P \geqslant \lambda^1_{\boldsymbol{X_i}}$, ${}^{(8)}_{\alpha} Z_i^* = 0$ if and only if ${}^{(8)}\boldsymbol{\beta}^*_{\boldsymbol{i1}} = \mathbf{0}$. Because of the continuity property [31, 32], ${}^{(8)}_{\alpha_B}\boldsymbol{\beta}^*_{\boldsymbol{i1}}$ approaches to $\pm\frac{1}{large\left(\boldsymbol{q}^{\mathbf{1}}_{\boldsymbol{X_i}}\right)}\boldsymbol{q}^{\mathbf{1}}_{\boldsymbol{X_i}}$ as $P$ goes to 0. In other words, for $\forall P$, $\exists \varepsilon$, such that ${}^{(8)}_{\alpha_B} Z_i^* < 0$ if and only if the unit vector $\boldsymbol{\beta_{i1}}$ satisfies $\left|\boldsymbol{\beta_{i1}} - {}^{(8)}_{\alpha_B}\boldsymbol{\beta}^*_{\boldsymbol{i1}}/norm\left({}^{(8)}_{\alpha_B}\boldsymbol{\beta}^*_{\boldsymbol{i1}}\right)\right| \leqslant \varepsilon$ where $\varepsilon$ approaches to 0 as $P$ goes to 0. Therefore, ${}^{(8)}_{\alpha_B}\boldsymbol{\beta}^*_{\boldsymbol{i1}}/norm\left({}^{(8)}_{\alpha_B}\boldsymbol{\beta}^*_{\boldsymbol{i1}}\right)$ will be the near-optimal component loading, which converge to the first component loading in PCA as $P$ goes to 0.

It follows the same logic when solving the NPCA problem. However, unlike in the PCA approach, where $\alpha_B$ can be determined by $\lambda^1_{\boldsymbol{X_i}} - P < \alpha_B < \lambda^1_{\boldsymbol{X_i}}$, the boundary point $\alpha_B$ cannot be easily obtained because of the smaller regional domains that resulted from the non-negative constraints. Furthermore, the optimal solution in equation (4) will reduce to 0 before $\alpha$ reaches $\lambda^1_{\boldsymbol{X_i}}$. To address this issue, we propose an IQP algorithm in Figure 3 to automatically find the boundary point in the NPCA problem.

Assume we can find the boundary point $\alpha_B$ by the IQP algorithm for any given small $P$ by the IQP algorithm. Next, we want to prove ${}^{(4)}_{\alpha_B}\boldsymbol{\beta}^*_{\boldsymbol{i1}}/norm\left({}^{(4)}_{\alpha_B}\boldsymbol{\beta}^*_{\boldsymbol{i1}}\right)$ converges to the optimal solution for equation (2) as $P$ goes to 0. Similarly, for $\forall P$, $\exists \varepsilon$, such that ${}^{(4)}_{\alpha_B} Z_i^* < 0$ if and only if the unit vector $\boldsymbol{\beta_{i1}}$ satisfies $\left|\boldsymbol{\beta_{i1}} - {}^{(4)}_{\alpha_B}\boldsymbol{\beta}^*_{\boldsymbol{i1}}/norm\left({}^{(4)}_{\alpha_B}\boldsymbol{\beta}^*_{\boldsymbol{i1}}\right)\right| \leqslant \varepsilon$ where $\varepsilon$ approaches to 0 as $P$ goes to 0. It is equivalent to state that ${}^{(4)}_{\alpha_B}\boldsymbol{\beta}^*_{\boldsymbol{i1}}/norm\left({}^{(4)}_{\alpha_B}\boldsymbol{\beta}^*_{\boldsymbol{i1}}\right)$ is the near-optimal solution with unit norm for equation (9). After ignoring the constant term $\alpha$ in equation (9), it proves that ${}^{(4)}_{\alpha_B}\boldsymbol{\beta}^*_{\boldsymbol{i1}}/norm\left({}^{(4)}_{\alpha_B}\boldsymbol{\beta}^*_{\boldsymbol{i1}}\right)$ converges to the optimal solution for equation (2) as $P$ goes to 0.

$$\min_{\boldsymbol{\beta_{i1}}} Z_i = -\boldsymbol{\beta_{i1}^T}\left(\boldsymbol{X_i^T X_i} - \alpha \boldsymbol{I}\right)\boldsymbol{\beta_{i1}} \quad s.t. \quad \boldsymbol{\beta_{i1}^T}\boldsymbol{\beta_{i1}} = 1, \mathbf{1} \geqslant \boldsymbol{\beta_{i1}} \geqslant \mathbf{0} \tag{9}$$

## References

1. Committee on Quality of Health Care in America, Institute of Medicine. *Crossing the Quality Chasm: A New Health System for the 21st Century*. National Academy Press: Washington, D.C., 2001.
2. Gawande AA. The cost conundrum: what a Texas town can teach us about health care. *The New Yorker* 2009; **85**:36–44.
3. Gawande AA, Fisher ES, Gruber J, Rosenthal MB. The cost of health care: highlights from a discussion about economics and reform. *The New England Journal of Medicine* 2009; **361**:1421–1423.
4. Woodard DB, Gelfand AE, Barlow WE, Elmore JG. Performance assessment for radiologists interpreting screening mammography. *Statistics in Medicine* 2007; **26**:1532–1551.
5. Parkerton PH, Smith DG, Belin TR, Feldbau GA. Physician performance assessment: non-equivalence of primary care measures. *Medical Care* 2003; **41**:1034–1047.

6. Coste J, Fermanian J, Venot A. Methodological and statistical problems in the construction of composite measurement scales: a survey of six medical and epidemiological journals. *Statistics in Medicine* 1995; **14**:331–345.
7. Cox DR, Fitzpatrick R, Fletcher AE, Gore SM, Spiegelhalter DJ, Jones DR. Quality-of-life assessment: can we keep it simple? *Journal of the Royal Statistical Society Series A* 2009; **155**:353–393.
8. Saisana M, Tarantola S. State-of-the-art report on current methodologies and practices for composite indicator development. *EUR 20408 EN Report*, European Commission, Joint Research Center, Italy, 2002.
9. Hastie T, Tibshiran R, Friedman JH. *The Elements of Statistical Learning*. Springer: New York, 2001.
10. Porter ME, Stern S. *The New Challenge to America's Prosperity: Findings from the Innovation Index*. Council on Competitiveness Publication Office: Washington, 1999.
11. Timbie JW, Shahian DM, Newhouse JP, Rosenthal MB, Normand ST. Composite measures for hospital quality using quality-adjusted life years. *Statistics in Medicine* 2009; **28**:1238–1254.
12. Martin S, Smith PC. Multiple public service performance indicators: toward an integrated statistical approach. *Journal of Public Administration Research and Theory* 2005; **15**:599–613.
13. Feingold M, Nelson DA, Parfitt AM. Composite index of skeletal mass: principal component analysis of regional bone mineral densities. *Journal of Bone and Mineral Research* 1992; **7**:89–96.
14. D'Agostino RB, Belanger AJ, Markson EW, Kelly-Hayes M, Wolf PA. Development of health risk appraisal functions in the presence of multiple indicators. The Framingham Study Nursing Home Institutionalization Model. *Statistics in Medicine* 1995; **14**:1757–1770.
15. Lee JH, Choi Y, Sung NJ, Kim SY, Chung SH, Kim J, Jeon T, Park HK. Development of the Korean primary care assessment tool—measuring user experience: tests of data quality and measurement performance. *International Journal for Quality in Health Care* 2009; **21**:103–111.
16. Filmer D, Pritchett LH. Estimating wealth effects without expenditure data—or tears: an application to educational enrollments in states of India. *Demography* 2001; **38**(1):115–132.
17. Coste J, Bouée S, Ecosse E, Leplège A, Pouchot J. Methodological issues in determining the dimensionality of composite health measures using principal component analysis: case illustration and suggestions for practice. *Quality of Life Research* 2005; **14**:641–654.
18. Durani Y, Brecher D, Walmsley D, Attia MW, Loiselle JM. The Emergency Severity Index Version 4: reliability in pediatric patients. *Pediatric Emergency Care* 2009; **25**:751–753.
19. Austin PC. A comparison of regression trees, logistic regression, generalized additive models, and multivariate adaptive regression splines for predicting AMI mortality. *Statistics in Medicine* 2007; **26**:2937–2957.
20. Buuren SV, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. *Statistics in Medicine* 1999; **18**:681–694.
21. Jain S, Elon LK, Johnson BA, Frank G, DeGuzman M. Physician practice variation in the pediatric emergency department and its impact on resource use and quality of care. *Pediatric Emergency Care* 2010; **26**:902–908.
22. Little RJA, Rubin DB. *Statistical Analysis with Missing Data*. Wiley: New York, 2002.
23. Nardo M, Saisana M, Saltelli A, Tarantola S, Hoffman A, Giovannini E. Handbook on constructing composite indicators: methodology and user guide. *OECD Statistics Working Paper,* Organization for Economic Cooperation and Development (OECD), Paris, 2005.
24. Kaiser HF. The varimax criterion for analytic rotation in factor analysis. *Psychometrika* 1958; **23**:187–200.
25. Han H. Nonnegative principal component analysis for mass spectral serum profiles and biomarker discovery. *BMC Bioinformatics* 2010; **11**:1–9.
26. Zass R, Shashua A. Nonnegative sparse PCA. *Advances in Neural Information Processing Systems* 2007; **19**:1561–1568.
27. Duong TDX, Duong VN. Non-negative sparse principal component analysis for multidimensional constrained optimization. *Proceeding of the 10th Pacific Rim International Conferences on Artificial Intelligence: Trends in Artificial Intelligence* 2008; **5351**:103–114.
28. Sigg CD, Buhmann JM. Expectation-maximization for sparse and non-negative PCA. *Proceedings of the 25th International Conference on Machine Learning* 2008:960–967.
29. Lipovetsky S. PCA and SVD with nonnegative loadings. *Pattern Recognition* 2009; **42**:68–76.
30. Tibshirani R. *Journal of the Royal Statistical Society B* 1996; **58**:267–288.
31. TAM NN. On continuity properties of the solution map in quadratic programming. *Acta Mathematica Vietnamica* 1999; **24**:47–61.
32. TAM NN. Continuity of the optimal value function in indefinite quadratic programming. *Journal of Global Optimization* 2002; **23**:43–61.
33. De A, P D, Pardalos PM, Toraldo G. Quadratic programming with box constraints. In *Developments in Global Optimization*, Bomze IM, Csendes T, Horst R, Pardalos PM (eds). Kluwer Academic Publishers: Dordrecht, 1997; 73–93.
34. Burer S, Vandenbussche D. Globally solving box-constrained nonconvex quadratic programs with semidefinite-based finite branch-and-bound. *Computational Optimization and Applications* 2007; **43**:181–195.
35. Chen J, Burer S. Globally solving nonconvex quadratic programming problems via completely positive programming. *Mathematical Programming Computation* 2012; **4**:33–52.
36. Burer S. On the copositive representation of binary and continuous nonconvex quadratic programs. *Mathematical Programming* 2009; **120**:479–495.
37. Sahinidis NV. BARON: a general purpose global optimization software package. *Journal of Global Optimization* 1996; **8**:201–205.
38. Burer S, Vandenbussche D. *QuadProgBB code (online)*. Accessed: 20 October 2011. URL: http://dollar.biz.uiowa.edu/~sburer/pmwiki/pmwiki.php{%}3Fn=Main.QuadprogBB{%}3Faction=logout.html.
39. Zass R. Homepage. *NSPCA MatLab code (online)*. Accessed: 19 April 2009. URL http://www.cs.huji.ac.il/~zass.