

RESEARCH

Open Access



Patient regional index: a new way to rank clinical specialties based on outpatient clinics big data

Xiaoling Peng¹, Moyuan Huang¹, Xinyang Li¹, Tianyi Zhou¹, Guiping Lin² and Xiaoguang Wang^{2*}

Abstract

Background Many existing healthcare ranking systems are notably intricate. The standards for peer review and evaluation often differ across specialties, leading to contradictory results among various ranking systems. There is a significant need for a comprehensible and consistent mode of specialty assessment.

Methods This quantitative study aimed to assess the influence of clinical specialties on the regional distribution of patient origins based on 10,097,795 outpatient records of a large comprehensive hospital in South China. We proposed the patient regional index (PRI), a novel metric to quantify the regional influence of hospital specialties, using the principle of representative points of a statistical distribution. Additionally, a two-dimensional measure was constructed to gauge the significance of hospital specialties by integrating the PRI and outpatient volume.

Results We calculated the PRI for each of the 16 specialties of interest over eight consecutive years. The longitudinal changes in the PRI accurately captured the impact of the 2017 Chinese healthcare reforms and the 2020 COVID-19 pandemic on hospital specialties. At last, the two-dimensional assessment model we devised effectively illustrates the distinct characteristics across hospital specialties.

Conclusion We propose a novel, straightforward, and interpretable index for quantifying the influence of hospital specialties. This index, built on outpatient data, requires only the patients' origin, thereby facilitating its widespread adoption and comparison across specialties of varying backgrounds. This data-driven method offers a patient-centric view of specialty influence, diverging from the traditional reliance on expert opinions. As such, it serves as a valuable augmentation to existing ranking systems.

Keywords Specialty influence, Patient regional index, Outpatient big data, Representative points of statistical distributions, Two-dimensional assessment model

Introduction

In the realm of healthcare, a 'specialty' denotes a distinct branch of medicine that is dedicated to the study and treatment of a specific category of diseases or conditions. The 'clinical specialty ability' encapsulates the capacity and growth potential of medical institutions to deliver specialized healthcare services. The quantitative appraisal of specialty ability is crucial for the distribution of medical resources, industry regulation, and hospital administration. Importantly, objective and rational assessments of specialty ability equip patients with

*Correspondence:

Xiaoguang Wang
wxiaog@mail.sysu.edu.cn

¹ Guangdong Provincial Key Laboratory of Interdisciplinary Research and Application for Data Science, BNU-HKBU United International College, Zhuhai, China

² Sun Yat-sen Memorial Hospital, Sun Yat-sen University, No. 107, Yanjiang West Road, Yuexiu District, Guangzhou, China



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

valuable information, aiding them in making informed decisions regarding their choice of medical institutions.

Over the past few decades, numerous organizations have ranked or rated hospitals and specialties within their respective countries. These rankings serve dual purposes: guiding patients in selecting hospitals or medical centers, and providing a scientific foundation for the evolution of specialties in government-run hospitals. One of the most renowned ranking systems is the 'Best Hospitals Honor Roll', introduced by the U.S. News & World Report in 1990 [1]. This system, designed to assist patients in identifying superior medical centers and physicians across the United States, was established based on patient outcomes, patient experiences, medical technology, specialty reputation (derived from physician surveys), and other health-related indicators. However, the inherent complexity of this ranking system often leads to conflicting standings amongst various institutions [2, 3]. In China, the most authoritative ranking system is the 'Chinese Hospital Specialist Reputation Rankings', issued by the Institute of Hospital Management at Fudan University. The evaluations conducted under this system primarily depend on expert ratings and research capability, and their credibility hinges on the authority and professionalism of the experts involved [4–6]. Despite its intricate computation, this system also neglects the intended audience of these rankings and lacks sufficiently objective evaluation standards [6, 7].

Recently, eight healthcare experts reviewed four major hospital rating systems and identified the following significant issues [8]:

1. Comprehensiveness and representativeness of the data. For example, most rating systems use administrative data collected for billing rather than clinical purposes, and incomplete data can lead to bias in the assessments.
2. Reliability of the data. Rating systems generate their own data through surveys and they are not able to assess the validity and reliability of the data independently.
3. Methods for integrating and weighting composite metrics vary. Different rating systems use different methods to calculate composite indicators, which causes overall scores or ratings to vary widely. In some cases, the choice of weights even depends on stakeholder perceptions.
4. Distorted evaluation of small hospitals. Small hospitals are difficult to assess fairly with the usual performance estimation methods because of their low capacity.
5. Lack of uniform peer-review. Although each rating system uses expert panels to some extent, the expertise of the panels is uncertain and their evaluations are heavily influenced by the subjective thinking of the experts.

From the discussion above, it's evident that while large and complex rating systems may seem comprehensive, their complexity in data collection and computation often hinders their effectiveness. Considering the diversity among different medical systems, it is challenging for these rating systems to draw convincing and consistent conclusions. Therefore, the ranking of hospitals and specialties should be straightforward and quantifiable. Utilizing assessment metrics that emphasize relevant patient-centered information can enhance patient acceptance of these evaluation metrics and better define the concept of patient-centered quality of medical care [9]. For instance, Cram suggested the use of patient-centered objective indicators, such as Readmission Reduction, to measure hospital quality [10].

Many studies investigating the factors that influence patients' access to medical care have identified both the reputation of the specialty and geographical distance as crucial variables affecting patient choices [11–13]. Typically, patients favor medical institutions with a higher reputation in their specialty and those closer in proximity [14]. Therefore, the scope of a specialty service is tied to its overall societal reputation; the higher the reputation, the broader the geographical origins of its patient base [15]. Specialties with stronger reputations draw patients not only from the immediate vicinity of the medical institution and the city but also from areas outside the city and even the province. The geographical proximity of the patients' origins to the medical institution serves as an objective indicator of the patients' trust in its specialty competencies.

Representative Points (RPs) were proposed as a technique to discretize and approximate a continuous statistical distribution [16] and have since been utilized in a variety of fields, such as information compression and transmission in signal processing [17]. Fang and He have applied RPs for the grouping of Chinese body sizes to develop clothing standards [18]. The most common type of RPs are the mean squared error representative points (MSE-RPs), which aim to minimize the mean squared error in relation to the original distribution. MSE-RPs are accomplished by segmenting the domain of the distribution into distinct intervals, each symbolized by a single point, known as the representative point. A unique characteristic of MSE-RPs is their self-consistency in single-peaked distributions, where the RP for each interval aligns with the expected value for that interval. This

property allows for the application of the Lloyd-Max method [19–21] to calculate RPs. By iteratively adjusting the interval endpoints and the expected values within each interval, MSE-RPs and their corresponding intervals can be obtained from a set of initial points.

Drawing from outpatient clinics big data, the objective of this paper is to introduce a novel index referred to as the patient regional index for assessing specialty influence using the MSE-RPs technology. Unlike conventional ranking methods that rely on diverse indicators and expert evaluations, our data-centric approach is straightforward to execute and comprehend, and it accurately represents the specialty influence from the patient's viewpoint. Consequently, it offers superior consistency and comparability when evaluating the specialty influence across various medical institutions.

Methods

Data selection and preprocessing

The Chinese healthcare system is a government-funded and government-administered system. In China, public general hospitals are typically the main healthcare institutions that provide comprehensive medical services, including various specialties. In this study, we utilized a large, comprehensive hospital in South China to demonstrate the specific calculation process of the PRI.

From all 33 departments of this general hospital, we meticulously chose 16 specialties as the focus of our analysis. The selection of these specific specialties was guided by several factors. Firstly, they represent the hospital's priority areas, reflecting the institution's strategic focus. Secondly, these specialties have a well-established history within the hospital, indicating their enduring relevance. Lastly, these specialties are ubiquitous across most general hospitals, underscoring their widespread prevalence. It's important to note that certain departments, such as the Emergency and Intensive Care Department, were deliberately omitted from our selection due to their unique patient demographics. As a result, the data set was collected from the outpatient registration records spanning 16 departments such as Pediatrics and Urology. Covering the period from 2014 to 2021, the dataset comprised a total of 10,098,024 visit records.

We began data processing by converting all patients' origins in the database into latitude and longitude coordinates, where the patient regional information was extracted from the address and telephone number in the patients' visit records. To achieve this, we utilized the Baidu Open Platform's geocoding Application Programming Interface (API), which enabled us to convert text addresses into their corresponding latitude and longitude coordinates. When we encountered invalid addresses, such as blank fields or unrecognizable entries, we turned

to the phone module in Python to extract area information from the patients' phone numbers. Instances where data lacked both a valid address and a valid phone number were classified as invalid.

We then employed the Python library 'geopy' to compute the distance from each patient's origin to the hospital, using the latitude and longitude coordinates. All coordinates were maintained to four decimal places, adhering to the default WGS-84 model. Upon calculation, 229 distances exceeding 5,000 km were identified. These were excluded from the analysis as they did not align with reality. Ultimately, we were left with 10,097,795 valid distances, which were further employed for estimating the baseline distance distribution and calculating the patient regional index (PRI).

Patient regional index

After segmenting the distances from patients' origins to the healthcare institution into several intervals, the PRI is then constructed by weighting the quantity (or proportion) of patients in each interval, inversely proportionate to the distance.

In this paper, we apply the theory of MSE-RPs to derive an optimal partition for the statistical distribution of distances, which allows us to determine the corresponding weights for each interval. Our initial step involves establishing a baseline distribution for these distances.

Fitting the baseline distance distribution

Considered that the majority of patients come from nearby areas, the likelihood of a patient traveling from a remote location is comparatively low. As such, the distribution of distances should exhibit right-skewness. Therefore, the two-parameter Gamma distribution $Ga(\alpha, \beta)$ serves as an appropriate model to characterize the distribution of distances from patients' residences to the healthcare institution.

Utilizing R 4.0.4, we fitted the baseline distance distribution $Ga(\alpha, \beta)$ using all valid distances from patients' origins to the hospital. The maximum likelihood estimation yielded parameter estimates of $\hat{\alpha}=0.1954$ and $\hat{\beta}=0.0014$. Figure 1 illustrates the distance distribution as a histogram, superimposed with the probability density curve for $Ga(0.1954, 0.0014)$. From Fig. 1 we see that the fitted Gamma distribution effectively captures the right-skewed nature of the distance distribution. This distribution indicates that the majority of visits originate from areas in close proximity to the hospital, with the frequency of visits significantly decreasing as the distance increases. Therefore, $Ga(0.1954, 0.0014)$ will serve as the baseline distance distribution for future partitioning and weighting.

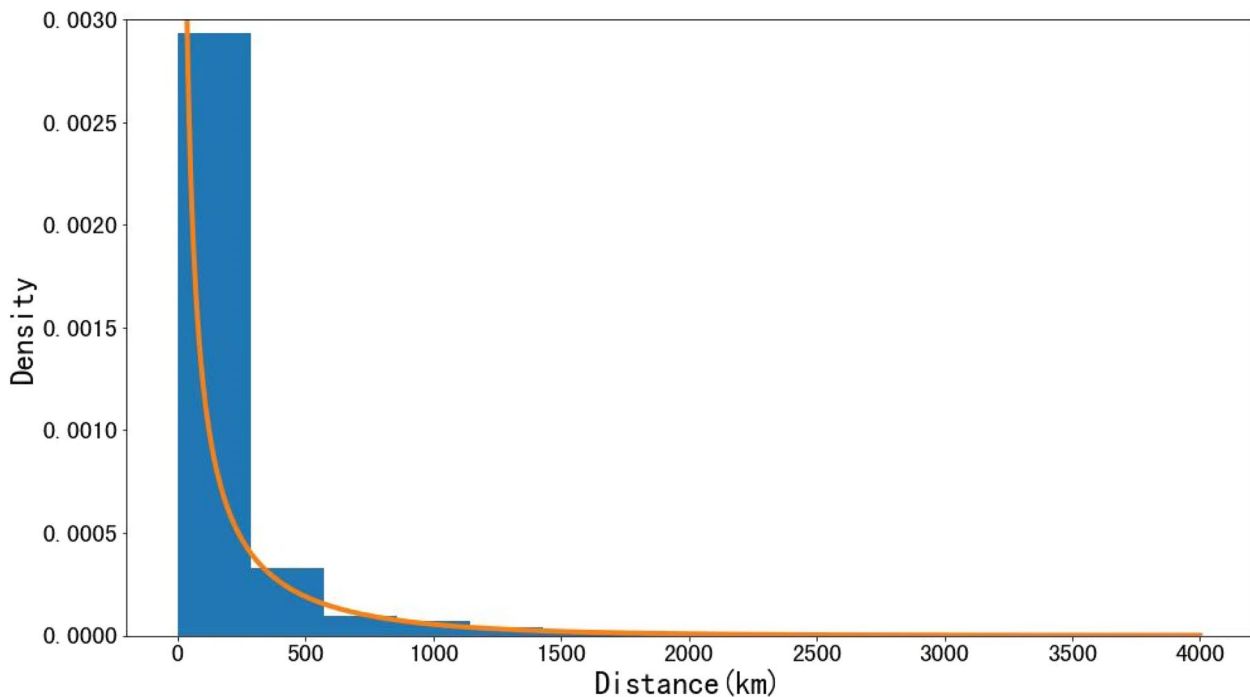


Fig. 1 Baseline distribution of the distances between the patients' origin and the hospital

Partitioning the baseline distance distribution

According to the theory of MSE-RPs [18], for a continuous distribution $F(x)$ defined on $[c, d]$, the k representative points $\mathbf{y} = \{y_1, y_2, \dots, y_k\}$, the corresponding k intervals $\Omega = \{\Omega_1, \Omega_2, \dots, \Omega_k\}$ and their respective probabilities $\mathbf{P} = \{P_1, P_2, \dots, P_k\}$ can be derived by minimizing MSE below:

$$MSE(y_1, \dots, y_k) = E(\min_{i=1, \dots, k} (X - y_i)^2) = \int_c^d \left(\min_{i=1, \dots, k} (x - y_i)^2 \right) f(x) dx$$

$$= \int_c^{\frac{y_1+y_2}{2}} (x - y_1)^2 f(x) dx + \int_{\frac{y_1+y_2}{2}}^{\frac{y_2+y_3}{2}} (x - y_2)^2 f(x) dx + \dots + \int_{\frac{y_{k-1}+y_k}{2}}^d (x - y_k)^2 f(x) dx,$$

where $f(x)$ is the probability density function of $F(x)$. Naturally, the interval Ω_i associated with y_i is its interval of integration. The cumulative probability p_i for each interval is given by

$$p_i = P(y = y_i) = \begin{cases} \int_c^{\frac{y_1+y_2}{2}} f(x) dx, & i = 1 \\ \int_{\frac{y_{i-1}+y_i}{2}}^{\frac{y_i+y_{i+1}}{2}} f(x) dx, & i = 2, \dots, k-1 \\ \int_{\frac{y_{k-1}+y_k}{2}}^d f(x) dx, & i = k \end{cases}$$

Figure 2 illustrates the distance intervals obtained from the partition of a right-skewed Gamma distribution, specifically for $k = 6$.

In this study, we used the Lloyd-Max method [19] to generate six representative points, along with their

corresponding intervals and cumulative probabilities, from the baseline distance distribution $Ga(0.1954, 0.0014)$. These are listed in Table 1.

Calculating the patient regional index

Adhering to the principle that greater distances should be assigned higher weights, we define the weight w_i of the i th distance interval to be inversely proportional to the probability p_i , that is, $w_i \propto \frac{1}{p_i}$. To ensure the sum of all weights equals 1, i.e., $\sum w_i = 1$, the weight of the i th distance interval is assigned as below:

$$w_i = \frac{\frac{1}{p_i}}{\sum_{i=1}^k \frac{1}{p_i}}, \quad i = 1, 2, \dots, k. \tag{1}$$

Following the probabilities of each interval provided in Table 1, the weights for these distance intervals were computed in accordance with Eq. (1) and are presented in the last row of Table 1.

Given the proportions of patients' origins distributed across these k distance intervals r_1, r_2, \dots, r_k , the patient regional index (PRI) is then defined as the weighted average of the patients' geographical distribution:

$$PRI = \sum_{i=1}^k w_i r_i. \tag{2}$$

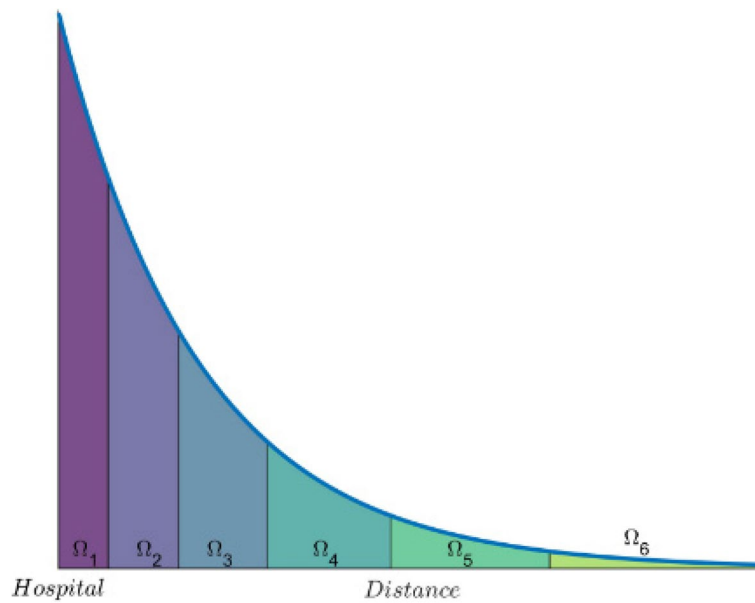


Fig. 2 Six distance intervals for a right-skewed Gamma distribution

Table 1 Representative points, associated intervals, cumulative probabilities, and corresponding weights derived from the baseline distance distribution $Ga(0.1954, 0.0014)$

<i>y</i>	17.8859	223.6800	539.6046	995.1722	1683.3197	2911.3003
Ω	[0, 120.7830)	[120.7830, 381.6423)	[381.6423, 767.3884)	[767.3884, 1339.2460)	[1339.2460, 2297.3100)	[2297.3100, 5000)
P	0.7566	0.1363	0.06302	0.0296	0.0117	0.0027
W	0.0026	0.01432	0.0310	0.0659	0.1663	0.7199

Since the distribution of patients’ origins varies across departments, the PRI scores derived from Eq. (2) will differ for different departments. Generally, a department will have a higher PRI score if it attracts a larger proportion of patients from more distant regions.

Finally, we provide a summary of the process used to construct the PRI from outpatient clinics data:

- Step 1:* Transform patient origin data into respective distances from the healthcare institution.
- Step 2:* Establish the baseline distribution for the population of distances.
- Step 3:* Obtain representative intervals and their associated probabilities from the baseline distribution.
- Step 4:* Determine the PRI for a specific specialty as a weighted average of the proportions of its patients within each representative interval.

Given that the PRI scores derived in the aforementioned manner were numerically small, we adopted the average PRI scores from 2017 as a benchmark, setting this score at 100. Subsequently, each PRI was adjusted as:

$$\text{Adjusted PRI} = \frac{PRI}{PRI_{2017}} \times 100.$$

For the sake of simplicity, in the remaining sections of this paper, any mention of the PRI will refer to the adjusted PRI.

A two-dimensional assessment model

The calculation of the PRI, as described above, primarily relies on the proportion of patients from different distances rather than their absolute numbers. This is because the number of patients can significantly vary from one department to another due to the unique characteristics of each specialty. For instance, in densely populated cities, pediatrics typically sees a larger patient volume, while some specialties like orthopedics primarily cater to a smaller population with physical deformities, resulting in fewer outpatient visits. By basing the PRI on proportions rather than total numbers, we can evaluate the regional influence of a specialty while mitigating the impact of the specialty’s inherent attributes.

Nevertheless, it's crucial to recognize that the volume of outpatient visits serves as a significant measure of a specialty's proficiency, reflecting aspects such as patient demand, quality of care, efficiency, and capacity, among others. Hence, a two-dimensional assessment framework, leveraging outpatient big data, that incorporates both patient regional distribution and outpatient volume can provide a more comprehensive evaluation of a specialty's influence. Figure 3 briefly illustrates a schematic diagram of this joint assessment model for specialty influence based on outpatient big data.

Results

We calculated the PRI for each of the 16 specialties of interest over eight consecutive years. This calculation involved a weighted average of the proportions of patients from six distance intervals for each specialty, which we summarized annually. The proportions were weighted according to Eq. (2), using the weights specified in Table 1.

PRI trends amid healthcare reform and pandemic

Figure 4 illustrates the changes in the PRI across 16 specialties within the hospital from 2014 to 2021. For comparative analysis, we categorized the 16 specialties into two groups: non-surgical and surgical departments. The non-surgical departments included pediatrics, nuclear medicine, dermatology, endocrinology, traditional Chinese medicine, cardiovascular medicine, rheumatology, and respiratory medicine. The surgical departments comprised gynecology and obstetrics, urology, hepatobiliary and pancreatic surgery, breast surgery, ENT (ear-nose-throat), plastic surgery, orthopedics, and oncology.

Figure 4 reveals a gradual increase in the PRI for each specialty over time, reflecting the hospital's development and diversification of its patient origin. This trend was particularly noticeable in the surgical departments, implying a growing regional influence and suggesting higher patient loyalty compared to non-surgical departments. It indicates that once patients recognize a hospital's specialty, they are willing to travel longer distances for treatment.

However, we observed a decrease in the PRI for most specialties post-2017. This decline can be attributed to healthcare reforms initiated by the Chinese government in 2017, which encouraged patients with common diseases to seek initial treatment at primary healthcare institutions. This policy led to a significant reduction in out-of-town patients visiting this comprehensive hospital [22, 23].

Additionally, the COVID-19 outbreak at the end of 2019 restricted people's mobility, creating a noticeable inflection point in 2020 for the PRI of certain hospital specialties, especially surgical ones. Following the Chinese government's control of the epidemic, patient mobility was restored, and the PRI of all specialties showed a significant rebound. Thus, the fluctuation in the PRI effectively mirrors the impact of China's healthcare reform and the COVID-19 pandemic on specialty outpatient clinics.

Specialties overview using the joint assessment model

Beyond the scope of patient origin, the volume of outpatient visits in the outpatient information system, is another important metric of the proficiency of hospital specialties. This volume serves as a broad indicator of the scale of the specialty, the standard of medical technology, the efficiency of outpatient management,

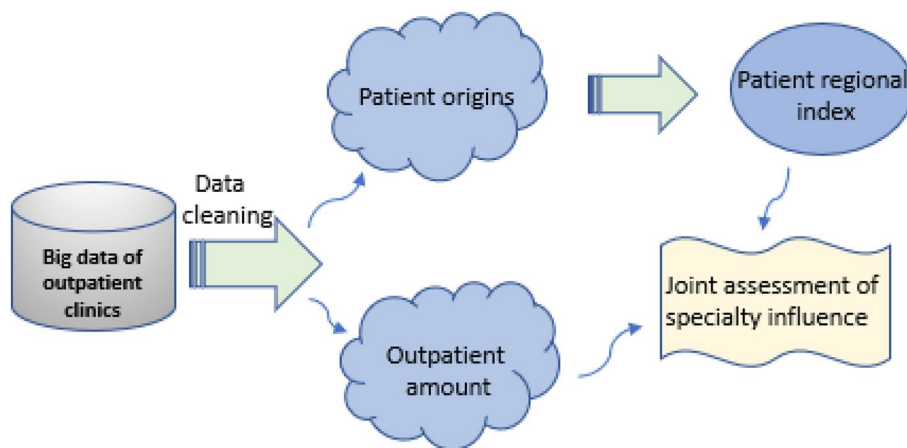


Fig. 3 A joint assessment model of specialty influence based on outpatient big data

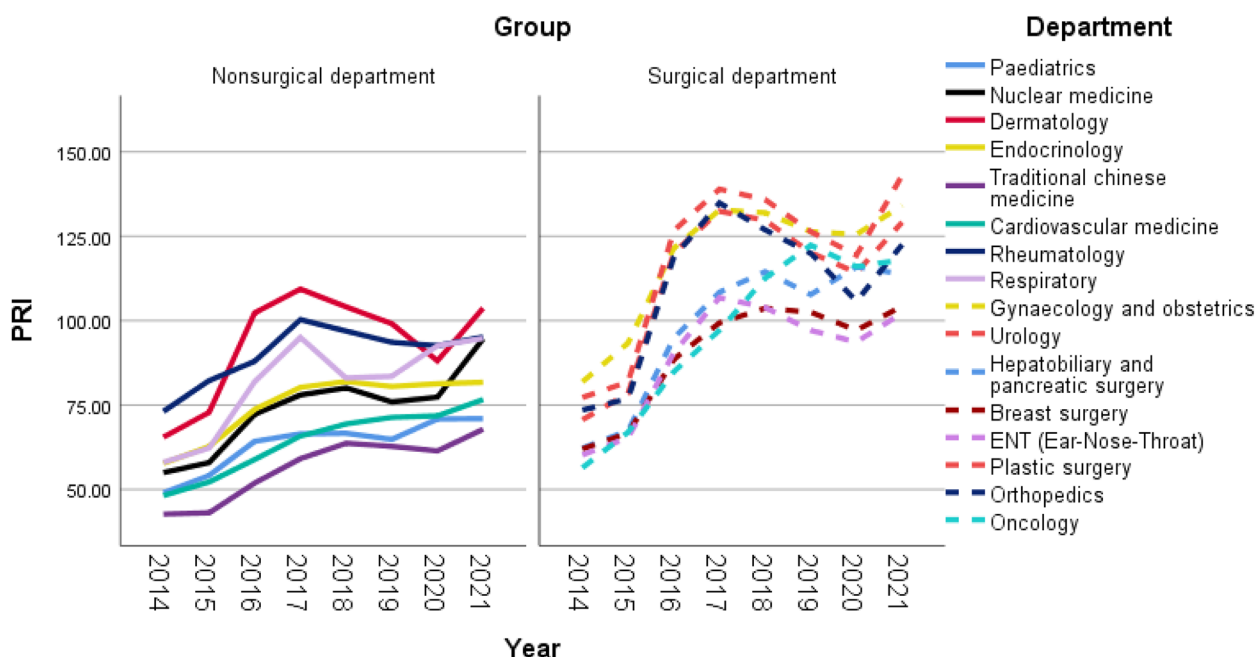


Fig. 4 Comparison of the PRI of specialties in a large comprehensive hospital, 2014–2021

and the extent of patient trust in the hospital [24]. For the purpose of comparison, the average number of outpatient visits per specialty in 2017 was utilized as a benchmark and assigned a score of 100. The number of visits per specialty was then adjusted as follows:

$$\text{Adjusted Outpatient Amount} = \frac{\text{Outpatient Amount}}{\text{Outpatient Amount}_{2017}} \times 100.$$

Figure 5 illustrates the shift in the number of visits to nonsurgical and surgical specialties in this hospital from 2014 to 2021. Among the nonsurgical specialties, pediatrics, Chinese medicine, dermatology, and endocrinology observed considerably higher outpatient volumes compared to the remaining four departments. Notably, there was a significant downturn in the volume of pediatric outpatients in 2020, a trend likely attributable to the effects of the pandemic. In the realm of surgical departments, gynecology and obstetrics were the most profoundly impacted by the pandemic, while the number of visits to other surgical departments showed a tendency to rise, rather than decline, in 2020. This trend underscores the resilience and capacity of this hospital’s specialties during challenging times.

By integrating the patients’ origin with outpatient volume, we could deliver a more holistic perspective on the strengths and unique features of the hospital’s various specialties. Figure 6 depicts a two-dimensional distribution in terms of the PRI and the adjusted outpatient

volume for 16 specialties within this large comprehensive hospital over an eight-year span. Identical symbols were employed to represent the same specialty values across different years.

Using the two-dimensional assessment model illustrated in Fig. 6, we broadly classified the hospital specialties into four primary categories. Category I encompasses specialties that demonstrate excellence through high outpatient volume and significant social influence. This category includes nonsurgical specialties such as Dermatology, and surgical specialties such as Gynecology and Obstetrics, which boasted an adjusted outpatient volume of approximately 150 or higher, and a PRI around 100. In fact, according to the ‘China Hospital and Specialty Reputation Ranking’, the Gynecology and Obstetrics department of this hospital holds the 2nd position in South China, affirming its superior specialty status.

Category II comprises specialties that, despite a smaller number of outpatient visits, maintain a high social reputation and attract patients from diverse regions. The Urology and Orthopedics departments fall under this category, with a PRI of 125 or above. These specialties are highly specialized, attracting a significant number of outpatients from distant locations.

Category III refers to specialties with a high outpatient volume, primarily serving local residents, and demonstrating robust operational capacity. The Pediatrics, Chinese Medicine, and Cardiovascular departments are

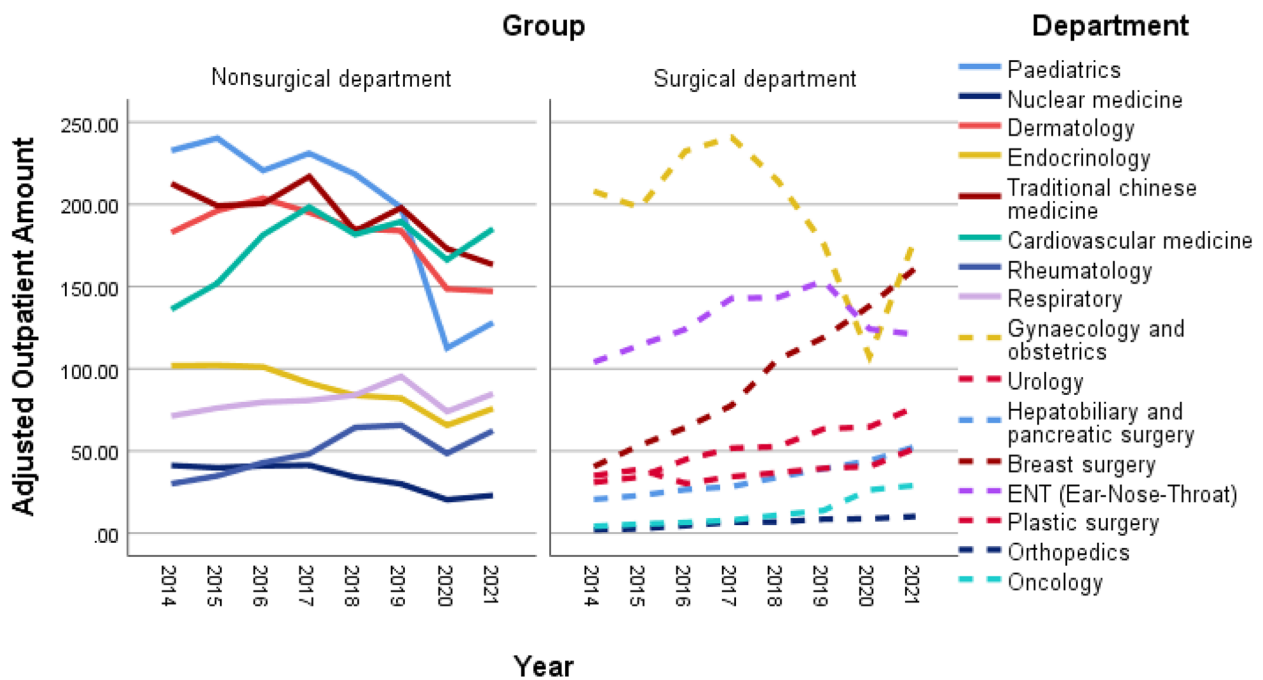


Fig. 5 Comparison of the outpatient volume of specialties in a large comprehensive hospital, 2014–2021

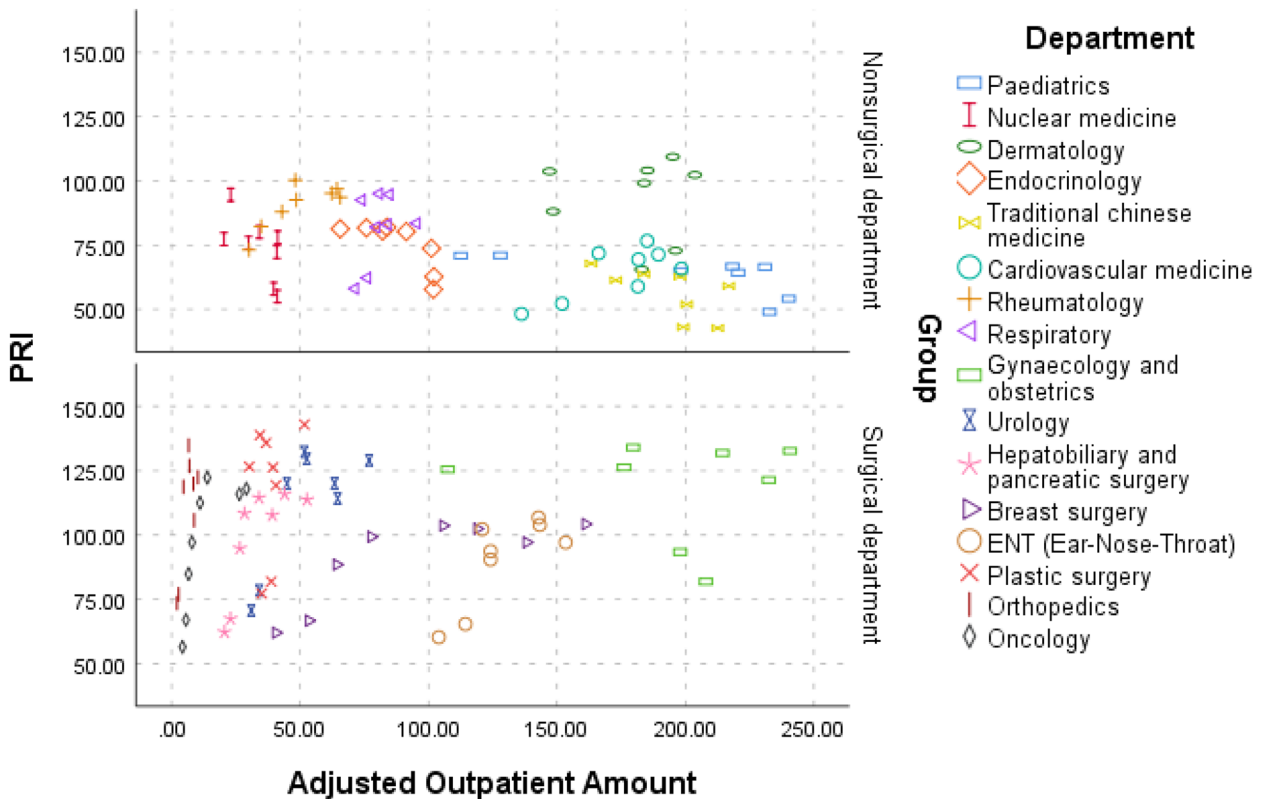


Fig. 6 Joint assessment of the specialties in a large comprehensive hospital, 2014–2021

included in this category, enjoying a strong reputation among local residents.

Lastly, Category IV includes specialties with smaller volumes, primarily serving local patients. The Nuclear Medicine and Rheumatology departments, due to the nature of their specialties, had a smaller volume of visits and an intermediate PRI.

We further selected Gynecology and Obstetrics (2017), Pediatrics (2019), Urology (2021), and Oncology (2014) as representatives of these four categories. We then generated a heat map detailing the distribution of their patients' origins (Fig. 7). This heat map confirms that the dual indicators of the PRI and outpatient volume can effectively capture the unique characteristics of each specialty department. For instance, the Department of Gynecology and Obstetrics enjoyed an outstanding reputation, attracting patients not only from Southern China but also in large numbers from Central and Eastern China. Some patients even traveled from as far as Northeast China. In comparison to Pediatrics, the Urology department, despite its smaller total patient visits, had a significantly broader geographical reach for its patient origins. Consequently, Urology's PRI was substantially higher than that of Pediatrics. In contrast, the Oncology department in 2014 was still in its nascent stages of development within the hospital. Its patients primarily resided in the hospital's immediate vicinity, resulting in both its PRI and outpatient volume being relatively low.

Conclusion

In this study, we developed a novel Patient Regional Index (PRI) to assess the influence of hospital specialties based on the statistical distribution of patient origins. By analyzing 10,097,795 outpatient records from a large comprehensive hospital in South China, we demonstrated that the PRI effectively captures the impact of significant healthcare events, such as the 2017 Chinese healthcare reforms and the 2020 COVID-19 pandemic. We also introduced a two-dimensional model that combines PRI with outpatient volume to provide a comprehensive characterization of various specialties. Based on the case study we have presented and discussed earlier, we wish to highlight the distinct advantages of the PRI as follows:

1. *Accessibility of data:* The dataset employed is straightforward and easily accessible. The methodology relies only two simple indicators (i.e., regional patient origins and outpatient volume), which can be readily collected from a hospital's outpatient system.
2. *Objective Weighting:* The weighting of distance intervals is determined by the statistical distribution of the data and relevant statistical theories, not by the subjective perceptions of stakeholders.
3. *Patient-Centric Robustness:* This index is robust and reliable as it reflects the choices of patients. It is a composite of the independent behavior of a substantial number of patients, rather than the subjective opinions of experts.

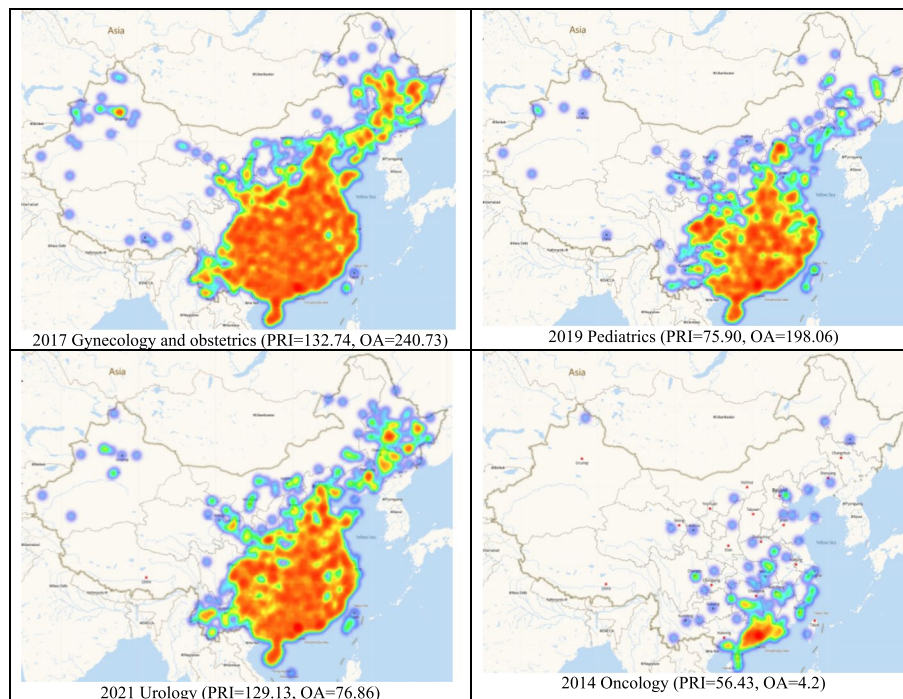


Fig. 7 Heat maps illustrating the distribution of patient origins across different specialty types. 'OA' denotes the adjusted outpatient volume

4. *Unbiased by Specialty Capacity*: The PRI is calculated using the proportion of patients from each regional interval rather than the total number of patients. This approach mitigates the distortion of other performance assessment methods for specialties with smaller volumes, thereby maximizing the fairness of the evaluation.
5. *Ease of Understanding and Implementation*: The assessment of specialty influence based on patient origins is easy to understand and implement. Therefore, it can be easily extended to evaluate the influence of an entire hospital, to allow for rankings and comparisons between different hospital specialties, ensuring consistent results.

Although our study is preliminary, it offers innovative ideas for effectively using big data from outpatient clinics to assess and rank hospital specialties. Compared with existing mainstream evaluation methods, which are often complex, our method is straightforward, easy to implement, and replicable, providing consistent conclusions. Therefore, our proposed assessment method can provide a meaningful reference that complements the existing evaluation systems.

Limitations

This study was confined to calculating and comparing the influence of a single hospital across various specialties and years due to data limitations. While outpatient big data is readily available within individual hospitals, obstacles persist in sharing this data across different institutions. The implementation of advanced data privacy technology is necessary to overcome these barriers, and that is the focus of our upcoming work.

Moreover, since the PRI is entirely reliant on outpatient data, there may be potential distortions for certain specialties. For instance, with the growth of the hospital, our study noted that the oncology department began to draw patients from an increasingly broader geographical range (PRI > 100). However, it was observed that a significant number of these patients were not attracted to the hospital due to the reputation of the oncology specialty per se, but rather due to the need for subsequent treatment in the radiation and chemotherapy clinics of the oncology department following surgeries from other specialties. As such, it is crucial to meticulously scrutinize patient origins when evaluating specific specialties using our method.

Abbreviations

PRI	Patient Regional Index
RPs	Representative Points
MSE	Mean Squared Error
API	Application Programming Interface

Acknowledgements

We thank Qianqian Lu, Beibei Liu and Jiahao Chen for their assistance with data processing. Additionally, we would like to express our sincere gratitude to the reviewers for their meticulous and thoughtful suggestions. Their contributions have greatly aided in enhancing the quality of our manuscript.

Authors' contributions

X.P. conducted the whole work and wrote the manuscript; M.H. did the main computation work; X.L. contributed to the methodology part and wrote the manuscript; T.Z. helped process massive amounts of data; G.L. interpreted the results and provided guidance; X.W. conceived the study and provided the data. All authors reviewed the manuscript and approved the submitted version.

Funding

This work was supported in part by the Guangdong Provincial Key Laboratory IRADS (No. 2022B1212010006), the Yat-sen Management Research Fund (No. GL2213), and the UIC research grant (No. R202108). None of these funding sources played any role in the design of the study, collection, analysis, and interpretation of data, or in writing the manuscript.

Availability of data and materials

The data presented in the study are included in the article, further inquiries about the dataset can be directed to the corresponding author (XW). Analysis code can be obtained upon request made with the first author (XP).

Declarations

Ethics approval and consent to participate

This is a retrospective study, the exemption from informed consent was approved by the ethics committee board of Sun Yat-sen Memorial Hospital, Sun Yat-sen University. All methods were carried out in accordance with relevant guidelines and regulations. All experimental protocols were approved by the ethics committee board of Sun Yat-sen Memorial Hospital, Sun Yat-sen University.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 25 May 2023 Accepted: 14 August 2024

Published online: 31 August 2024

References

1. U.S. News hospitals rankings and ratings. <https://health.usnews.com/best-hospitals>.
2. Austin JM, Jha AK, Romano PS, et al. National hospital ratings systems share few common scores and may generate confusion instead of clarity. *Health Aff (Millwood)*. 2015;34(3):423–30.
3. Bae JA, Curtis LH, Hernandez AF. National hospital quality rankings: improving the value of information in hospital rating systems. *JAMA*. 2020;324(9):839–40. <https://doi.org/10.1001/jama.2020.11165>.
4. Xiong XC, Gao JC, Zhang BY, Lu J, Luo L. Introduction of Chinese hospital ranking method from the aspect of theoretical framework, realistic choice and social effect. *J Hosp Manage Health Policy*. 2017;1:4.
5. Dong SP, Guo SY, He L, et al. Study of present situation and countermeasures of China's hospital rankings. *Chin Hosp Manage*. 2015;35(3):38–40.
6. Wei T, Ma LP, Chen Y, et al. Status Quo of clinical specialist ability evaluation in China. *Chinese Hospital Management*. 2019;39(3):14–17.
7. Lubalin JS, Harris-Kojetin LD. What do consumers want and need to know in making health care choices? *Med Care Res Rev*. 1999;56(Suppl 1):67–102.
8. Bilimoria KY, Birkmeyer JD, Burstin H, Dimick JB, Maddox KE, Dahlke AR, DeLancey JO, Pronovost PJ. Rating the raters: an evaluation of publicly reported hospital quality rating systems. *NEJM Catal*. 2019;5(4).

9. Ellis RJ, Yuce TK, Hewitt DB, Merkow RP, Kinnier CV, Johnson JK, Bilimoria KY. National evaluation of patient preferences in selecting hospitals and health care providers. *Med Care*. 2020;58(10):867–73. <https://doi.org/10.1097/MLR.0000000000001374>.
10. Cram P, Wachter RM, Landon BE. Readmission reduction as a hospital quality measure: time to move on to more pressing concerns? *JAMA*. 2022;328(16):1589–90. <https://doi.org/10.1001/jama.2022.18305>.
11. Exworthy M, Peckham S. Access. Choice and travel: implications for health policy. *Soc Policy Adm*. 2006;40:267–87.
12. Avdic D, Moscelli G, Pilny A, Sriubaite I. Subjective and objective quality and choice of hospital: evidence from maternal care services in Germany. *J Health Econ*. 2019;68(102229):1–19.
13. Mukamel DB, Amin A, Weimer DL, et al. When patients customize nursing home ratings, choices and rankings differ from the government's version. *Health Aff (Millwood)*. 2016;35(4):714–9.
14. Zhan JJ, Fu HQ. Hospital reputation, spatial distance and patient choice—evidence from hospital discharge data. *China Econ*. 2022;22:343–64.
15. Miao L, Chen W, Ding J, Chen R. Study on the medical service radius of large general hospital. *Prog Mod Biomed*. 2013;13(27):5381–6.
16. Cox DR. Note on grouping. *J Am Stat Theory*. 1957;52:543–7.
17. Gray RM, Neuhoff DL. Quantization. *IEEE Trans Inform Theory*. 1998;44(6):2325–83. <https://doi.org/10.1109/18.720541>.
18. Fang KT, He SD. The problem of selecting a specified number of representative points from a normal population. *Acta Math Appl Sin*. 1984;7(3):293–306.
19. Flury B. Estimation of principal points. *J R Stat Soc C*. 1993;42:139–51.
20. Lloyd SP. Least squares quantization in PCM. *IEEE Trans Inform Theory*. 1982;28(2):129–37.
21. Max J. Quantizing for minimum distortion. *IEEE Trans Inform Theory*. 1960;6(1):7–12.
22. Yip W, Fu H, Chen AT, et al. 10 years of health-care reform in China: progress and gaps in universal health coverage. *Lancet*. 2019;394(10204):1192–204. [https://doi.org/10.1016/S0140-6736\(19\)32136-1](https://doi.org/10.1016/S0140-6736(19)32136-1).
23. Zhang LY, Cheng G, Song S, et al. Efficiency performance of China's health care delivery system. *Int J Health Plann Manage*. 2017;32(3):254–63. <https://doi.org/10.1002/hpm.2425>.
24. Zhao X, Li W, Wang Y, Jiang L. Evaluation of the number of visits to Chinese medical institutions using a logistic differential equation model. *Complex*. 2021;2021. <https://doi.org/10.1155/2021/7943651>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.