Suggested Citation: Li, Y., Cao, L., Zhang, J., Jiang, Y., Han, Y., and Wei, J. (2021). "Energy Benchmarking in Healthcare Facilities: A Comparative Study." Journal of Construction Engineering and Management, 147(11), 04021159. For the published version, please refer to ASCE Database here: https://ascelibrary.org/doi/10.1061/(ASCE)CO.1943-7862.0002183

Energy Benchmarking in Healthcare Facilities: A Comparative Study

Yongkui Li, Ph.D.[1]; Lingyan Cao[2]; Jiansong Zhang, A.M.ASCE[3]; Yi Jiang, Ph.D[4]; Yilong Han, Ph.D.[5]; Jianjun Wei[6]

[1] Professor, Department of Construction Management of Real Estate, School of Economics and Management, Tongji Univ., Shanghai 200092, China. Email: lyk@tongji.edu.cn

[2] Ph.D. Candidate, Department of Construction Management of Real Estate, School of Economics and Management, Tongji Univ., Shanghai 200092, China; School of Construction Management Technology, Purdue University, 401 N Grant St., West Lafayette, IN 47907, USA. Email: cly900601@tongji.edu.cn

[3] Assistant Professor, School of Construction Management Technology, Purdue Univ., 401 N. Grant St., West Lafayette, IN 47907 (corresponding author). Email: zhan3062@purdue.edu

[4] Professor, School of Construction Management Technology, Purdue Univ., 401 N. Grant St., West Lafayette, IN 47907. Email: jiang2@purdue.edu

[5] Assistant Professor, Department of Construction Management of Real Estate, School of Economics and Management, Tongji Univ., Shanghai 200092, China. Email: yilong.han@tongji.edu.cn

[6] Ph.D. Candidate, Department of Construction Management of Real Estate, School of Economics and Management, Tongji Univ., Shanghai 200092, China. Email: 1510322@tongji.edu.cn

Abstract

Benchmarking the energy efficiency of buildings is important for improving the efficient use of energy and reducing carbon footprint, especially for healthcare facilities with high energy usage intensity. However, the historical difficulty of collecting energy data from a relatively large number of healthcare facilities has hobbled efforts to develop such an efficient benchmarking system. In this paper, we seek to stimulate such efforts by benchmarking the energy efficiency of healthcare facilities using three different methods, i.e., multiple linear regression (MLR), generalized additive model (GAM), and energy performance index (EPI). The analysis was applied using a unique dataset that contained information on energy consumption and various building features for 22 large-size public hospitals managed by the Shanghai hospital development center (SHDC). Findings suggest that different benchmarking methods yield substantively different energy performance ranking results. Furthermore, a comparative analysis of the three benchmarking methods was conducted in terms of goodness of fit, consistency, and robustness. The results show that MLR tends to be the most consistent and robust benchmarking model while GAM could bring the highest goodness of fit. The proposed methodology can assist hospital managers in identifying potential improvements for the energy efficiency of healthcare facilities.

Author Keywords: Healthcare Facilities; Energy Performance; Benchmarking; Generalized Additive Model

Introduction

Concerns about the resource shortages and environmental pollution issues with the use of fossil fuels are growing around the world, driven by rapid population and economic growth. According to the Intergovernmental Panel on Climate Change (IPCC), at least one-third of global primary energy consumption and greenhouse gas emissions were attributed to the existing buildings (IPCC 2014). Healthcare facilities, in particular, have received increasing attention due to their high energy usage intensity (Tejero-González et al. 2020). Compared to many other building types (e.g., hotel buildings, office buildings), healthcare facilities are complex buildings that are characterized by strict thermal comfort requirements and additional use of specialist medical equipment (Morgenstern et al. 2016). In the United States, healthcare facilities represent 4.8% of the total area in commercial buildings and account for 10.3% of their total energy consumption, yielding the second highest energy usage intensity after the food service industry (Bawaneh et al. 2019). In Germany, the average annual energy consumption of a hospital under normal climatic and operational conditions is 0.27MWh/m^2 (González González et al. 2018a). In China, the energy usage intensity of healthcare buildings in 2016 was 118 kilowatt-hours (kWh)/(m^2a), two times more than the ordinary public buildings (Jiang et al. 2018). The energy benchmarking of hospitals needs to be conducted as a specific area, as recommended by previous studies (Lomas and Ji 2009; Singer et al. 2009). Thus, improving the energy performance of existing buildings, and specific healthcare buildings can be seen as a significant lever from both environmental and economic perspectives.

To improve the energy performance of existing buildings, a critical step is to measure it in a transparent and objective way (Wei et al. 2018). This involves comparing the in-use energy performance of a single building against a performance baseline (Roth et al. 2020). Generally, three types of performance baseline can be calculated, i.e., previous performance, a reference performance level, or current performance of similar buildings (Roth and Rajagopal 2018). In the context of this study, calculating the last type of performance baseline is focused on, a process known as energy benchmarking. Consistent with the original definition of benchmarking given by Camp (1989), energy benchmarking is a process of searching for the best practice which can lead the organization to achieve superior energy performance. Energy benchmarking can provide the building owners the information about the energy performance of their buildings, making them more knowledgeable about energy efficiency improvement opportunities (Kontokosta et al. 2020). In addition to creating an energy performance baseline for a specific building, benchmarking is helpful for building owners and managers to set priorities for limited resources (e.g., such as capital and staff) (Cox et al. 2013).

Throughout the years, a significant number of energy benchmarking tools have been developed by governments and agencies to support energy efficiency programs. For example, the *EnergyStar* Portfolio Manager platform is a widely adopted interactive resource management tool in the United States and Canada, enabling building owners to benchmark the energy use of buildings (Arjunan et al. 2019). Likewise, all member nations in the European Union were suggested to implement building energy labelling schemes that can provide ratings to buyers in the real estate market (Annunziata et al. 2013). Inspired by these systems, other parts of the world have initiated their own benchmarking systems, such as Australia (Bannister 2012) and Singapore (Duarte et al. 2016). Considerable researches have reported that the widespread deployment of energy benchmarking systems has impacted the energy efficiency of the existing building significantly. For example, a study conducted by Roth et al. (2020) showed that 7% of savings had been achieved over four years since many cities throughout the U.S. mandated large-size buildings to benchmark their energy consumption. From an emission and cost reduction point of view, it is necessary to develop an efficient energy benchmarking system to enable scientific evaluation of building energy performance.

While benchmarking tools developed by different organizations and researchers in the scientific community provide indications of how one building compares to its peers, as Morgenstern et al. (2016) noted, energy benchmarking in healthcare settings has received less attention. Fundamentally, two possible reasons have been identified: the limited access to data from a relatively large number of healthcare facilities and the inherent differences between healthcare facilities. As Wei et al. (2018) pointed out, different benchmarking systems have their application situations with their own assumptions. Thus, the performance of the widely adopted benchmarking systems and their full potential in healthcare scenarios are still not clear. Moreover, the characteristics of energy consumption and influence mechanisms may differ from country to country. For example, research conducted by (Hong et al. 2014) showed that Chinese buildings consume much more energy for lighting than buildings in the United States because the latter's single-pane windows can introduce more natural light. Consequently, the following research question is asked: "Are there any effective benchmarking tools that can help determine or compare the energy performance of healthcare facilities?".

The purpose of this study, therefore, is to present a framework that can be applied to gain a deeper understanding of the development of an energy benchmarking system for healthcare facilities. Such a framework also permits comprehension of the processes where energy performance is influenced. Generally, the healthcare facility is a broad term that can be used to describe several building typologies, such as hospitals, outpatient and inpatient centers, and community care facilities (Ahmed et al. 2015). Given that hospitals are accounting for more than half of the energy usage in healthcare systems (Bawaneh et al. 2019), this study focuses on energy consumptions in large-scale hospitals. To illustrate the research process, three methods, i.e., multiple linear regression (MLR), generalized additive model (GAM), and energy performance index (EPI), were selected for energy benchmarking. A unique dataset that contains information on energy consumption and various building features for 22 large-size public hospitals managed by the Shanghai hospital development center (SHDC) was utilized. Additionally, the ranking results of the three methods were compared in terms of goodness of fit, consistency, and robustness. It is worth noting that the framework outlined in this study is only one of many that might be suggested. This study aims to provide evidence as a basis for the development of meaningful energy performance targets for hospitals.

Literature Review

Current Energy Benchmarking Approaches

During the last 20 years, a wide variety of benchmarking models has been developed for assessing the energy performance of existing buildings. Generally, these models can be classified into four categories: end-use metrics, point-based rating system, simulation model-based method, and statistical analysis (Ghajarkhosravi et al. 2019). The end-use metric, also known as the energy performance index (EPI), is the most commonly used method to indicate the relative energy performance of a single building. However, the downside of EPI is that it does not account for other important factors that are known to have effects on energy performance, such as buildings' age, occupancy levels, weather, etc. (Chung 2011). The point-based rating system is used to evaluate the energy performance of a certain building concerning predefined guidelines and standards. A representative of the point-based rating system is Leadership in Energy and Environmental Design (LEED), which is the leading program for green buildings and communities worldwide. The main process of LEED is to certify green buildings by awarding points to buildings meeting some specific green building standards. In spite of its great success, LEED has been criticized for ignoring context and lack of scientific robustness (Zimmerman and Kibert 2007). The simulation model-based method can virtually assess the periodical load dynamics of buildings by mimicking the physical attributes, including properties of materials (e.g., concrete, brick) and building geometry (e.g., width and height of a surface), etc. (Kim et al. 2015; Yang et al. 2018). Simulation model-based methods are advantageous in considering various inputs contributing to energy use and can generate design and operational alternatives (Sokratis Papadopoulos 2019). However, existing studies have pointed out that simulation-based systems require extensive domain expertise and a time-intensive calibration process to yield accurate and reliable results (Ahmad and Culp 2006). The statistical model is based on the implementation of a function by inputting some observable data already available (Sowby and Burian 2018). This method is well employed when the physical features of the considered building are not known (Foucquier et al. 2013). Regression-based approaches as well as artificial neural networks, are commonly used statistical models (Hawkins et al. 2012).

In summary, each of these benchmarking methods has its pros and cons. Due to the great heterogeneity of healthcare facilities, an arbitrary choice of one single indicator or method may fail to accurately benchmark healthcare facilities. Thus, it is necessary to develop a comparative benchmarking methodology for energy managers so that more reliable results can be achieved.

Benchmarking Program in Healthcare Sector

Benchmarking can serve as a useful tool to measure the operating performance of healthcare organizations and thus facilitate the performance comparison within and outside of their organizations (Ettorchi -Tardy et al. 2012). Since its first appearance in the healthcare sector in 1990, when benchmarking was used to meet the needs of the Joint Commission on Accreditation of Healthcare Organizations (JACHO) in the U.S., several benchmarking frameworks in the healthcare context have been established. In 2001, the Organization for Economic Cooperation and Development (OECD) initiated the Health Care Quality Indicator (HCQI) Project to assess international health care quality that has ever been undertaken. In 2003, the Performance Assessment Tool for Quality Improvement in Hospitals (PATH) was designed as an internal tool for quality improvement in hospitals by World Health Organization (WHO) (Groene et al. 2008). Another famous benchmarking program is the Construction

Industry Institute (CII)'s National Healthcare Facility Benchmarking Program which focuses on the delivery of healthcare projects ranging from programming to activation/move-in phases (Choi et al. 2020). In addition to these programs developed by governments and agencies, there are also benchmarking studies conducted in academia. For example, Feibert et al. (2019) benchmarked the bed logistics process and the pharmaceutical distribution process of a hospital to improve process performance. Fry et al. (2016) conducted a benchmarking study on risk-adjusted adverse outcomes to identify the opportunity for care improvement. Morgenstern et al. (2016) constructed an energy benchmarking by taking into account the electricity usage of several department types, such as wards, theatres, laboratories, and some other departments. Kamaluddin et al. (2016) developed a typical base-case hospital building model via *EnergyPlus* software based on available surveyed data in Malaysia.

By reviewing the previous literature relevant to benchmarking programs in the healthcare context, two gaps have been identified. First, compared to the core business of healthcare facilities (e.g., clinical quality, patient care, service delivery), the energy performance of healthcare facilities has historically rarely been studied. Second, there is a lack of systematic benchmarking for energy performance at the hospital level instead of the building or department level, which is critical for setting energy-saving targets for hospitals.

Evaluation Criteria of Benchmarking Methods

To identify the most appropriate benchmarking method, a comparison between different methods is necessary. Thus, different evaluation criteria have been used in previous studies. For example, Keirstead (2013) employed three methods (i.e., ratio measures, regression residuals, and data envelopment analysis) to measure urban energy efficiency and concluded that each method has its own strengths and weakness in terms of the ease of interpretation, ability to identify outliers and provide consistent rankings. Gao and Malkawi (2014) validated the feasibility and robustness of their proposed clustering approach with the *EnergyStar* approach. Chen et al. (2018) compared their proposed Lorenz curve method with previous statistical methods. They summarized the advantages of that method from four aspects: reliability, applicable flexibility, generalized ability, and assumption of models. Papadopoulos and Kontokosta (2019) compared the performance of the GREEN grading and the EnergyStar-based scoring for NYC's large residential properties in terms of the goodness of fit and energy performance grade assignment. Ding and Liu (2020) compared the consistency, robustness, and explanatory ability of three methods (i.e., multiple linear regression, stochastic frontier analysis, and the descriptive statistics method) and found that the robustness of these methods depends on the specific benchmarking purpose. For example, if the building owners want to compare the energy performance of buildings in terms of ranking order, the descriptive statistics method would outperform the other two methods. When the energy performance is compared in terms of energy grades, stochastic frontier analysis tended to be the method with the highest robustness.

The ideal way for validating benchmarking methods is to compare the benchmarking results with the ground truth. However, it has been a widely acknowledged challenge across building benchmarking studies (Ding and Liu 2020; Francisco et al. 2020). Modern buildings encompass complex energy systems, making it challenging to acquire the practical energy efficiency of an individual building, let alone large-scale buildings (Ding and Liu 2020). To a certain extent, it is the lack of ground truth that makes benchmarking energy performance based on existing methods more critical and necessary, enabling building operators to identify the inefficient buildings with limited resources, such as data, time, and effort.

Methodology

The proposed methodology for the energy benchmarking process is capsulized in Fig. 1. As indicated, there are four major steps: data acquisition, methods selection, derivation of the ranking table, and comparison of methods. More details for the four steps can be found in the following section.

Insert Figure 1 about here

Data Acquisition

As a follow-up study from the performance benchmarking of healthcare facilities management reported by Li et al. (2020), the proposed models will be applied to a group of municipal public hospitals located in Shanghai, China. These healthcare facilities are managed by Shanghai Hospital Development Center (SHDC), a management institution that was set up by Shanghai Municipal People's Government in September 2005. SHDC needs to make informed decisions based on detailed knowledge of the energy performance of each hospital. Thus, there is a strong motivation for SHDC to find a way to measure the energy performance of hospital buildings. In this study, data that contained information on energy consumption and energy influencing factors of these hospitals were provided by SHDC.

Methods Selection

As mentioned previously, various methods have been employed in the field of energy benchmarking. However, it is not the aim of this study to create new benchmarking methods. Instead, the novelty of the proposed methodology lies in its original analysis of the evaluation performance of existing benchmarking methods from the perspective of their applications in healthcare settings. In addition, it is because each method has its own pros and cons that make it necessary to compare the different methods based on some evaluation criteria. An arbitrary choice of one single method may bring misleading policy conclusions (Cai et al. 2019). As mentioned previously, the main purpose of this study is to present a possible instrument and process for building owners and managers to establish an efficient energy benchmarking framework with limited resources. Therefore, three commonly used benchmarking methods, i.e., multiple linear regression (MLR), generalized additive model (GAM), and energy performance index (EPI) were selected for the benchmarking analysis.

MLR Method

Multiple linear regression (MLR) is selected for being the most popular method and the ability to obtain the most suitable subset of independent variables (Wei et al. 2018). As the extension of a simple linear regression model having only one independent variable, MLR can accommodate multiple independent variables, resulting in a better explanatory power (Sharma et al. 2020). According to Wei et al. (2018)'s research, the stepwise multiple linear regression method can be used to develop the MLR model. The basic principle of this method is to introduce significant factors into the model step by step and then remove those insignificant factors. As with Gao et al. (2019), the logarithmic transformation of both independent variables and dependent variables was prompted to address the heteroscedasticity of the model. Meanwhile, variance inflation factor (VIF) is used as the index for measuring multicollinearity, i.e., a phenomenon in which more than two independent variables are highly related in a multiple regression model (Lavery et al. 2019). There may be serious multicollinearity between the independent variables if the VIF value is greater than or equal to 10 (Chai et al. 2018). Otherwise, it can be concluded that there is no multicollinearity between variables. Finally, a more accurate regression model can be obtained. The developed MLR model is given by the following equation:

$$\ln Y = \beta_0 + \beta_1 \ln X_1 + \beta_2 \ln X_2 \cdots + \beta_p \ln X_p + \varepsilon \quad (1)$$

Where Y is the dependent variable, β_0 represents the intercept, β_i represents the coefficient, X_i are the significant influencing factors ($i=1, 2, \cdots, n$), and ε is the residual. An important performance metric is adjusted R^2, which can be used to indicate the goodness of fit as well as represent the amount of variance explained by the MLR model.

GAM Method

The generalized additive model (GAM) is a generalization of linear regression models, in which there are assumed additive functions and smooth components (Hastie and Tibshirani 1986). To be specific, the coefficients in GAM can be expanded as smooth functions of independent variables. Thus, they can account for non-linear relationships between dependent variables and independent variables, offering a middle ground between simple models (e.g., linear regression) and complex models (e.g., neural networks). In recent years, GAM has been applied in many fields, such as ecological fields (Jowett et al. 2008), and environmental issues (Zou et al. 2017). According to de Brogniez et al. (2015), a GAM model can be expressed as follows:

$$E(Y|X_1,X_2, \cdots,X_P)=\alpha+f_1(X_1)+f_2(X_2)+\cdots f_p(X_p) \tag{2}$$

Wherein $X_1, X_2, \cdots, X_P$ are the independent variables, Y represents the dependent variable, and $f_j(X_j)$ is a smooth function that can be estimated with nonparametric regression methods.

EPI Method

The energy performance index (EPI) is one of the most widely used metrics in energy benchmarking. Generally, EPI can be viewed as the ratio of the energy input to some energy-using factors, such as the number of services or goods (Abu Bakar et al. 2015), depending on the specific scenarios. For example, for hotel buildings, the number of guest room was considered in the calculation of EPI (Teng et al. 2017). González González et al. (2018) collected data from 23 public hospitals in Germany and analyzed the relationship between energy consumption and the built area, the number of workers, and the number of beds per hospital. They found that the most suitable reference to quantify the energy consumption of a hospital is the total energy consumption normalized by the number of beds. Healthcare buildings are complex systems that can be viewed as a combination of general buildings and hospital-specific buildings. Therefore, three types of EPI were adopted in this study, namely, EPI_{area}, EPI_{bed}, and EPI_{person}. EPI_{area} was selected because it is the most commonly used EPI in the conventional building sector for measuring energy performance (Bakar et al. 2015). EPI_{bed}, EPI_{person} were treated as hospital-specific indicators. As with González González et al. (2018a), EPI_{bed} is defined as the ratio between the energy consumption and the number of available beds. EPI_{person} refers to the energy consumption normalized by the total number of patients (*TNOP*). Following the method adopted in Li et al. (2020), *TNOP* is the sum of the number of outpatients, the number of emergency patients, and three times the number of outpatients. They can be calculated as follows:

$$EPI_{area}=\frac{Total\ energy\ consumption}{Gross\ floor\ area} \tag{3}$$

$$EPI_{bed}=\frac{Total\ energy\ consumption}{Number\ of\ available\ beds} \tag{4}$$

$$EPI_{person}=\frac{Total\ energy\ consumption}{Number\ of\ outpatients+number\ of\ emergency\ patients+3*number\ of\ inpatients} \tag{5}$$

Derivation of Ranking Table

Based on the established MLR or GAM model, predicted energy consumption for each hospital ($Y_{Predicted}$) can be calculated. Then, the energy efficiency ratio (EER) can be defined as the ratio between the actual energy consumption (Y_{Actual}) and the predicted energy consumption ($Y_{Predicted}$), just as shown in the following equation.

$$EER=\frac{Y_{Actual}}{Y_{Predicted}} \tag{6}$$

The smaller the value of *EER* is, the more energy-efficient is the building since it means the building uses less energy than the predicted value; otherwise, the less energy efficient is the targeted building. Based on the calculated EER, a ranking table can be derived. Ranking by EER is in descending order, meaning smaller EER values indicate the higher energy efficiency of hospitals.

Comparison of Methods

In this study, the performance of benchmarking methods is evaluated on three criteria: goodness of fit, consistency, and robustness.

Goodness of Fit

An essential step in developing a benchmarking system is to select a model that can generate a reasonable and reliable predicted value of total energy consumption, i.e., high goodness of fit. In this study, the adjusted R^2 was used to represent the goodness of fit.

Consistency

Three data-driven benchmarking methods, MLR, GAM, and EPI, were applied in this study. As mentioned previously, three types of widely used EPIs for healthcare facilities were selected, including EPI_{area}, EPI_{bed}, and EPI_{person}. Thus, there are a total of five ranking results obtained. Following the method adopted by Graafland and Eijffinger (2004) and Ding and Liu (2020), the Spearman correlation coefficient was used to test the consistency between these methods.

Robustness

In a practical situation, buildings will be divided into different grades according to their energy performance. Based on their ranking results, the 22 hospitals were divided into three groups (with different colors). The top 1/3 are denoted as Group 1 (rank 1 to 7), the middle 1/3 (rank 8 to 15) are denoted as Group 2, and the last 1/3 (rank 16 to 22) are denoted as Group 3. Following Ding and Liu (2020) and Cai et al. (2019), the Sankey diagram is used to mapping the interactions among the ranking results of the three methods.

Suggested Citation: Li, Y., Cao, L., Zhang, J., Jiang, Y., Han, Y., and Wei, J. (2021). "Energy Benchmarking in Healthcare Facilities: A Comparative Study." Journal of Construction Engineering and Management, 147(11), 04021159. For the published version, please refer to ASCE Database here: https://ascelibrary.org/doi/10.1061/(ASCE)CO.1943-7862.0002183

Case Settings and Results

Data Acquisition

Sample Description

Supported by the Shanghai hospital development center (SHDC), 22 large-scale municipal hospitals were selected as the sampled hospitals due to data availability. These 22 public hospitals included twelve general hospitals (numbered H1 to H12), five specialty hospitals (numbered H13 to H17), and five traditional Chinese medicine hospitals (numbered H18 to H22). Locations of these hospitals were plotted on Google Map (Fig. 2). Located in the hot summer and cold winter area, a large amount of energy consumed in these hospitals has been used for air-conditioning and space heating due to the long summer and short winter (Chen et al. 2009). Additionally, just as hospitals in other megacities such as Beijing, Shenzhen, these sampled hospitals of Shanghai are extremely busy providing medical services to locals as well as people from other regions of China (Wang et al. 2016), resulting in a continuous rise in energy cost as well as a high environmental impact. According to a local energy consumption monitoring report, the healthcare buildings of Shanghai have the highest energy usage intensity 177.8 kWh/(m^2a) among all building types in 2018 (SMMCHUD and SDRC 2019).

Insert Figure 2 about here

Fig. 3 presents the layout of one sampled hospital. Located in the same city, these hospitals have similar buildings typologies, such as assemblies and geometry. Each of the sampled hospitals can be regarded as a building group that consists of buildings with different functions, such as outpatient department, emergency department, inpatient wards, operating rooms, and supporting spaces (e.g., pharmacy and radiology). For these hospitals, the air conditioning system may be central air-conditioning systems, variable refrigerant volume (VRV) unit systems, large-sized test devices for air-cooled heat pump units, and split air-conditioning systems (Cao et al. 2020). Among the three common types of building structures, i.e., brick-wood, brick-concrete, and reinforced, most sampled hospitals are made of brick-concrete (Hong et al. 2016).

Insert Figure 3 about here

Energy Consumption Data

Monthly utility usage from January 2015 to December 2018 was obtained from the intelligent–Building Energy Support System (i-BESS), including two main fuel types: electricity and natural gas. The i-BESS is an external data repository developed by the Shanghai hospital development center (SHDC), tracking energy usage and utility cost for all its hospitals. The i-BESS consists of two sub-platforms, one platform is designed for SHDC, and the other one is for municipal hospitals managed by SHDC. The platform of SHDC consists of modules of land and building use information, equipment installation information (e.g., boiler, elevator, and medical gas systems), and specific energy and resources consumption information of different hospitals. For the platform of municipal hospitals, it is mainly used for operation monitoring, quality monitoring, and data analysis. Initially, different physical units were used to measure different types of energy, such as kilowatt-hours (kWh) for electricity and cubic meter (m^3) for natural gas. To compare these energy measurements with each other, we need to convert them to the same units. Therefore, electricity (kWh) and natural gas (m^3) usage were converted into standard coal consumption with the unit of kilograms of standard coal equivalent (kgce) and summed to provide a single approximate value for the energy use of each hospital. The conversion coefficient of electricity and natural gas are 0.327 kgce/kWh and 1.214 kgce/m^3, respectively (Wei et al. 2018). Then, the energy consumption can be calculated from the following equation:

$$\textit{Energy use (kgce)} = 0.327 \times \textit{Electricity use (kWh)} + 1.214 \times \textit{Natural use } (m^3) \quad (7)$$

Considering that building envelope deterioration tends to be slow, the monthly energy consumption data are aggregated to annual total energy consumption. In this way, the average annual energy use from 2015 to 2018 was calculated for each hospital (see Fig. 4). As noted in Fig. 4, most hospitals have a relatively steady trend of energy use across four years.

Insert Figure 4 about here

Energy Influencing Factors

According to the findings of previous research related to building energy performance measurement (Hong et al. 2014; Park et al. 2016), the desired dataset should consist of building envelope information (e.g., floor area, number of floors, building condition), occupants' demography (e.g., number of occupants), building function (building use ratio), and local climate statistics (e.g., outdoor temperature). In this study, twelve energy influencing factors were eventually used to describe features of the sampled hospitals by considering the procedure for selecting inclusive variables in previous literature and the data availability simultaneously (Table 1). Among them, the first three factors can be viewed as building envelope relevant, i.e., gross floor area, number of floors above ground, and number of available beds. The following four indicators are occupant relevant, including personnel expenses, number of outpatients, number of emergency patients, and number of inpatients. The personnel expense was used as a proxy indicator of staff input due to the limited access to the data of staff. When it comes to the building use category, each hospital was divided into four types of building use, namely, outpatient and emergency patient units (OEPUs), medical technical units (MTUs), inpatient units (IPUs), and facility management units (FMUs), according to the code for design of general hospitals in China (NHCPRC 2014). Therefore, the building use ratio (BUR) of the above-mentioned four building types were calculated as the gross floor area divided by the floor areas allocated to different building uses (Park et al. 2016), i.e., BUR_{OEPU}, BUR_{MTU},

BUR_{IPU}, and BUR_{FMU}. As can be seen from **Error! Reference source not found.**, most of the sampled hospitals are geographically concentrated. Thus, weather differences can be ignored across the group of hospitals.

Similar to the collection of energy consumption data, monthly data of the independent variables from the year 2015 to 2018 were collected and then aggregated and used as data on yearly basis. As is shown in Table 1, the majority of the data were downloaded from i-BESS, whereas the number of patients was drawn from Healthcare Information System (HIS). Remarkably, the data of annual personnel expenses were downloaded from the public website of the Shanghai government (Shanghai Municipal People's Government 2021). For brevity, the average of these data across the four years was calculated for the later benchmarking model development.

Insert Table 1 about here

Error! Reference source not found. gives the basic statistics for the above-mentioned variables, including their mean value, median value, maximum value, minimum value, and standard deviation.

Insert Table 2 about here

Methods Selection

MLR Method

A multiple linear regression model was developed using characteristics of the 22 sampled hospitals as independent variables. The final MLR model is shown as the following equation, in which three independent variables were employed as significant, including *GFA*, *FAG*, and *NIP*.

$$\ln TEC = 4.656 + 0.661 \ln GFA + 0.207 \ln FAG + 0.235 \ln NIP \quad (8)$$

It was expected that the increase in *GFA* and *NIP* would increase the energy

Suggested Citation: Li, Y., Cao, L., Zhang, J., Jiang, Y., Han, Y., and Wei, J. (2021). "Energy Benchmarking in Healthcare Facilities: A Comparative Study." Journal of Construction Engineering and Management, 147(11), 04021159. For the published version, please refer to ASCE Database here: https://ascelibrary.org/doi/10.1061/(ASCE)CO.1943-7862.0002183

consumption of the hospital because more medical services would be delivered. As *FAG* grows, the total energy consumption will also increase because more electricity will be used for vertical transportation such as elevators.

The developed model was found to be significant with a p-value less than 0.001 and have large goodness of fit with the adjusted R^2 of 0.957, and thus it could adequately explain the correlation between energy consumption and the various factors. Therefore, this regression model can be used for energy benchmarking of hospital buildings.

GAM Method

In this study, the GAM-based benchmarking model was implemented by the "mgcv" package (version 1.8.33) of the R program (version 4.0.2) (Rstudio Team 2020). Consistent with the result of the MLR model, the three significant independent variables in GAM were *GFA*, *NIP,* and *FAG*. **Error! Reference source not found.** summarizes all the estimation for GAM's smooth function, and the adjusted R^2 was 0.98. Therefore, the GAM model can be used for energy benchmarking.

Insert Table 3 about here

EPI Method

As a simple baseline benchmarking method, EPI represents the most straightforward measurement of energy efficiency. In this study, three types of EPIs were taken into account, i.e., energy consumption per gross floor area (EPI_{area}), energy consumption per bed (EPI_{bed}), and energy consumption per patient (EPI_{person}). Fig. 5 shows the distribution of EPI_{area}, EPI_{bed}, EPI_{person} among the 22 hospitals, respectively. The mean value of EPI_{area}, EPI_{bed}, EPI_{person} are 65.45 kgce/(m^{2}a), 8208.04 kgce/(bed•a), 3.74 kgce/(person•a), respectively.

Insert Figure 5 about here

Derivation of Ranking Table

MLR Ranking

Based on the above MLR model, the predicted energy consumption of each hospital can be calculated. Table 4 presents the ranking results based on energy efficiency ratios calculated using the MLR method. As noted in Table 4, of the 22 hospitals, 11 have an energy efficiency ratio less than 1.

Insert Table 4 about here

GAM Ranking

Once the GAM model is identified, predicted energy consumption could be calculated. The calculation process of the EER score is similar to that of MLR, which is, dividing the predicted energy consumption by the actual energy consumption. Table 5 presents the ranking results of EER based on the GAM method. As can be observed, among the 22 hospitals, 14 of them have an EER of less than 1.

Insert Table 5 about here

EPI Ranking

Table 6 presents the quantitative results of EPI of the sampled hospitals. As explained in the previous section, a greater EPI ranking value indicates a higher energy efficiency.

Insert Table 6 about here

Comparison of Methods

Each of the above methods provided a collection of unique benchmarking results for the sampled hospitals. When a number of benchmarking methods are available, which one should facilities managers adopt is of great significance on the final benchmarking result. Thus, a comparative analysis of the three benchmarking methods was conducted in terms of the goodness of fit, consistency, and robustness.

Goodness of Fit

As previously mentioned, EPI is calculated as the total energy consumption divided by a dominant energy influencing factor. In other words, it can also be considered as a linear regression with only one independent variable. Thus, to get the adjusted R^2 of EPI methods (i.e., EPI_{area}, EPI_{bed}, EPI_{person}), three linear regression models were established by treating gross floor area (*GFA*), the number of available beds (*NAB*), and the total number of patients (*TNOP*) as the only input variable with *TEC* being the dependent variable. Fig. 6 gives the relationship between *TEC* and the three variables. The result indicated that almost 89% of the variation in the *TEC* could be explained by GFA.

Insert Figure 6 about here

Table 7 gives the adjusted R^2 of the five models, i.e., MLR, GAM, EPI_{area}, EPI_{bed}, and EPI_{person}. It can be observed that GAM can provide the highest goodness of fit with the adjusted R^2 is 0.98, followed by MLR and EPI_{area}.

Insert Table 7 about here

Consistency

According to Bonett and Wright (2000), a relatively accurate estimation of the Spearman Coefficient can be achieved when the sample size is larger than 20. Thus, the Spearman correlation coefficients with corresponding *p*-values based on the ranking of the five benchmarking methods are shown in **Error! Reference source not found.**. As pointed out by Akoglu (2018), there is a weak correlation between two variables when the Spearman coefficient ranges from 0.1 to 0.3. The correlation between the two variables becomes stronger with the increase of the Spearman coefficient. It can be seen that the rankings of MLR and the EPI_{area} are most consistent with the largest Spearman correlation coefficient of 0.743, whereas the GAM and the EPI_{area} have the weakest correlation with a 0.487 correlation coefficient. Therefore, MLR is the method that can bring the highest consistency, which is consistent with the conclusion drawn by Ding and Liu (2020). Interestingly, it was found that there is no significant correlation observed among the ranking of three commonly used EPIs (i.e., EPI_{area}, EPI_{bed}, EPI_{person}) in healthcare scenarios. The empirical results showed that energy consumption normalized by gross floor area (EPI_{area}) is the most suitable indicator for denoting the energy performance of healthcare facilities, which is consistent with the conclusion drawn by (González González et al. 2018b), a study focused on the energy consumption of hospitals in Spain. The results suggest that how the total energy consumption is normalized can impact the final ranking results significantly. Thus, more attention should be paid to the selection of the energy performance index despite its ease of use and interpretability.

Insert Table 8 about here

Robustness

As can be observed from previous analysis, both EPI_{bed} and EPI_{person} have relatively low goodness of fit and Spearman coefficients with the other three benchmarking methods. Thus, they will not be considered for further analysis. Fig. 6 is a Sankey diagram displaying the differences in energy performance ranking results of the three methods. The first column gives the ranking based on the MLR method. The second column gives the ranking based on the GAM method. The third column gives the ranking based on the EPI. As such, Fig. 7 provides a general overview of the variation of the rankings.

As noted in Fig. 7, most hospitals with the best or the worst energy efficiency remained consistent regardless of the method used. However, rankings of hospitals in the middle range varied greatly according to the method employed. For instance, H3 ranked 3 with the MLR method; however, it dropped to 12 with the EPI method and even lower to 22 with the GAM method. H8's GAM ranking was 15, rising to 12 with the MLR method and even higher, to 4 with the EPI method. In addition, the pairwise comparisons of the three benchmarking methods showed that 36.4% of the sampled hospitals were assessed with a different group between the MLR method and GAM method. The values were 22.7% and 54.5% by comparing MLR-EPI and GAM-EPI, respectively. Thus, MLR can also be perceived as the most robust benchmarking method.

Insert Figure 7 about here

Discussion

Different ranking results can be obtained based on the benchmarking method employed (see Fig. 8).

Insert Figure 8 about here

The essential purpose of energy benchmarking is to identify the best practice so that other hospitals can learn from them. In this study, the best practice should be the hospital with the highest ranking in terms of EER among all the sampled hospitals. As is noted in Fig. 8, the best performer in both MLR and GAM is H16, whereas the best performer in EPI is H21. Then, a further investigation on these two hospitals was conducted. H16 is a traditional Chinese medicine hospital with a gross floor area of 39850 m^2. TCM hospitals are medical institutions that treat the patients with service (e.g., acupuncture, scraping therapy) and traditional products (e.g., Chinese herbal piece, Chinese patent medicine) to maintain public health (Li et al. 2020). One factor that may have contributed to the low EPI of H16 is its advanced energy management level. According to Bao et al. (2015), a logistics intelligent management system was introduced to H21 before 2015, earlier than most other sampled hospitals. The system mainly consists of information on building construction, electricity, water, central air conditioning, water supply and drainage, boilers, elevators, medical equipment, and other systems. Among these subsystems, particular attention has been paid to an energy monitoring system that can monitor the detailed usage of energy sources in real-time. Consequently, the energy management level of a hospital can be improved, and the energy cost can be reduced further.

H21 is a mental health hospital specializing in the treatment of severe mental disorders. Different from those psychiatric hospitals that provide short-term or outpatient therapy for low-risk patients, H21 is specialized in the temporary or permanent care of residents requiring routine assistance, treatment, or a specialized and controlled environment due to psychological disorder. Thus, the majority of residents in H21 are inpatients with a long occupancy period. Compared to other typical hospitals such as general or specialty hospitals, the volume of energy-consuming activities (e.g., outpatient therapy or operating) is far smaller in a psychiatric hospital like H21, resulting in a relatively low total energy consumption. Thus, low EPI of H21 can be considered as the result of its specific function to some extent.

As can be observed in Figure 8, the worst performer of the MLR method was H15. The worst performer of the GAM method was H2, whereas the worst performer of EPI_{area} was H4. Interestingly, it was found that hospitals with the last three EER remain consistent across the three methods. Thus, a deeper investigation was conducted on the three hospitals, i.e., H2, H4, and H15. It was found that H2 is a general hospital built in 1907, a time when there were no strict regulations on energy conservation in China. H4 is also a general hospital with an even longer history than H2, which was built in 1844. H4 has many advanced large-scale medical equipment such as cell knives, ultrasonic knives, MRI, 128-row CT, digital X-ray camera, PET-CT, which are energy-intensive. H15 is a traditional Chinese medicine hospital equipped with various energy-intensive medical equipment, including 3.0T MRI machine, gene chip scanner, luminescence immunoassay analyzer, automatic modular biochemistry, bone and joint imager, etc.

It has been widely acknowledged that sample size inefficiency will bring the concern of generalizability of studies' results, i.e., the potential to draw inferences from the sample to the broader study population (Vasileiou et al. 2018). However, as noted in Jenkins and Quintana-Ascencio (2020), greater sample size may be unavailable due to the limited cost and effort. Thus, it is important to determine the minimum sample size of statistical analysis. A review of the few articles shows that the minimum sample size varies in different scenarios. For example, Curtis et al. (2015) suggested that statistical analysis should have a minimum of 5 independent samples per group. Ford (2005) mentioned that a minimum sample size of 10-20 per predictor was acceptable for ecological studies. Others have offered advice based on the data shape at different sample sizes, assuming that a stable data shape can bring accurate inference (i.e., generalizability). According to (Jenkins and Quintana-Ascencio 2020), the data shape can be clearly identified when the sample size is larger than 8 and an accurate inference is available when the sample size is larger than 25. Thus, the statistical analysis in this study can be considered valid using a sample size of 22 hospitals.

Conclusion

A better understanding of the relative energy efficiency of buildings is crucial for building owners and managers to target energy savings opportunities, especially for healthcare facilities characterized by intense use of high-end equipment and accessories. To accomplish this, a three-way methodology (i.e., MLR, GAM, EPI) to calculate and benchmark the hospital energy efficiency ratios were developed against empirical data from 22 large-scale municipal hospitals in Shanghai, China. The findings suggest that different benchmarking methods yield substantively different energy performance ranking results. Interestingly, it was found that the last three energy-inefficient hospitals remain consistent across the three methods. Furthermore, these ranking results were compared in terms of the goodness of fit, consistency, and robustness. The results show that the MLR model tends to be the one with the highest consistency and robustness, whereas the GAM model can bring the highest goodness of fit in terms of adjusted

R^2. Therefore, it is recommended that energy managers should apply multiple benchmarking methods instead of an arbitrary choice of one single method, avoiding misleading policy conclusions drawn by using only one benchmarking method.

This study offers the following key contributions. First, we contribute to the extant literature on energy benchmarking by developing a methodology considering the energy characteristics of healthcare facilities. Given the importance of healthcare facilities in general, and particularly in the pandemic period, this study sheds some lights on the aspects of affordability and cost related to energy efficiency and consumption of healthcare facilities using real data. Second, this work contributes to the academic debate about which benchmarking method should be selected when there are various energy benchmarking tools available. This study presents a possible instrument and process to establish an efficient energy benchmarking framework by using several evaluation criteria, such as goodness of fit, consistency, and robustness. Furthermore, the proposed methodology in this paper could be applied to evaluation of any group of benchmarking methods when there is no or limited actual energy performance data, which is urgently needed both in research and practice for managing energy performance of healthcare facilities. It is worth noting that both the benchmarking methods and the evaluation criteria outlined in this study are only three of many that could potentially be used. This approach could be applied to other industry sectors as well, by replacing the healthcare-specific factors in this study with correspondingly more suitable ones for the industry of interest.

Our results have managerial implications. At the single hospital level, hospital managers can know how their hospitals perform in terms of energy performance against a group of similar hospitals. Once the performance gap is identified, they can conduct a deeper investigation of the best performer and then take suggested refurbishment strategies. At the level of management agency of hospitals such as SHDC, the benchmarking results can facilitate the exchange of experience and knowledge in the field of energy management between different hospitals, which is helpful for forming a collaborative benchmarking network. In this way, they can make informed decisions (e.g., allocation of limited resources, prioritization of energy retrofitting programs) to improve the overall performance of hospital groups.

This study has the following limitations. First, benchmarking results presented in this paper are only valid for the sampled hospitals and periods analyzed. Because of the limited number of cases, it is not appropriate to draw any far-reaching conclusions about the generality of the specific benchmarking methods beyond this scope. Second, due to the limitations of the database and missing values, only 12 independent variables were ultimately chosen for energy benchmarking. As the i-BESS evolves, new factors could be added to ascertain a more comprehensive ranking. Despite the lack of some energy influencing factors (e.g., the type of HVAC systems, hours of operation), the benchmarking models applied in this study can explain the majority of the variance in the practical energy consumption according to the high value of adjusted R^2. Thus, it is still reasonable to identify the best practice generated by the three benchmarking methods.

Data Availability Statement

Some or all data, models, or codes that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgements

This study is supported by China Scholarship Council (No. 202006260314), the Support Program for Young and Middle-Tech Leading Talents of Tongji University, the National Natural Science Foundation of China (No. 72001160), "Chenguang Program" under Grant (No. 18CG21) supported by Shanghai Education Development Foundation and Shanghai Municipal Education Commission.

References

Abu Bakar, N. N., Hassan, M. Y., Abdullah, H., Rahman, H. A., Abdullah, M. P., Hussin, F., and Bandi, M. (2015). "Energy efficiency index as an indicator for measuring building energy performance: A review." *Renewable and Sustainable Energy Reviews*, 44.1–11. https://10.1016/j.rser.2014.12.018.

Ahmad, M., and Culp, C. H. (2006). "Uncalibrated building energy simulation modeling results." *HVAC and R Research*, https://10.1080/10789669.2006.10391455.

Ahmed, T. M. F., Rajagopalan, P., and Fuller, R. (2015). "A classification of healthcare facilities: Toward the development of energy performance benchmarks for day surgery centers in Australia." *Health Environments Research and Design Journal*, https://10.1177/1937586715575910.

Akoglu, H. (2018). "User's guide to correlation coefficients." *Turkish Journal of Emergency Medicine*.

Annunziata, E., Frey, M., and Rizzi, F. (2013). "Towards nearly zero-energy buildings: The state-of-art of national regulations in Europe." *Energy*, https://10.1016/j.energy.2012.11.049.

Arjunan, P., Poolla, K., and Miller, C. (2019). "EnergyStar++: Towards more accurate and explanatory building energy benchmarking." *Applied Energy*, Elsevier, 276(June).115413. https://10.1016/j.apenergy.2020.115413.

Bakar, N. N. A., Hassan, M. Y., Abdullah, H., Rahman, H. A., Abdullah, M. P., Hussin, F., and Bandi, M. (2015). "Energy efficiency index as an indicator for measuring building energy performance: A review." *Renewable and Sustainable Energy Reviews*, Elsevier, 44.1–11.

Bannister, P. (2012). "NABERS: Lessons from 12 years of performance-based ratings in Australia." *Energy Systems Laboratory*.

Bao, Y., Yin, H., Hao, J., Zhang, J., and Zhou, C. (2015). "Study on the Construction of Logistics Intelligent Management System for Shanghai JS Hospital." *Refrigeration and Air Conditioning*, (1).44–48.

Bawaneh, K., Nezami, F. G., Rasheduzzaman, M., and Deken, B. (2019). "Energy consumption analysis and characterization of healthcare facilities in the United States." *Energies*, 12(19). https://10.3390/en12193775.

Bonett, D. G., and Wright, T. A. (2000). "Sample size requirements for estimating Pearson, Kendall and Spearman correlations." *Psychometrika*, https://10.1007/BF02294183.

de Brogniez, D., Ballabio, C., Stevens, A., Jones, R. J. A., Montanarella, L., and van Wesemael, B. (2015). "A map of the topsoil organic carbon content of Europe generated by a generalized additive model." *European Journal of Soil Science*, 66(1).121–134. https://10.1111/ejss.12193.

Cai, B., Guo, H., Ma, Z., Wang, Z., Dhakal, S., and Cao, L. (2019). "Benchmarking carbon emissions efficiency in Chinese cities: A comparative study based on high-resolution gridded data." *Applied Energy*, Elsevier, 242(March).994–1009. https://10.1016/j.apenergy.2019.03.146.

Camp, R. C. (1989). "Benchmarking: The search for industry best practices that lead to superior performance." *Quality Progress*, https://10.5860/choice.27-2173.

Cao, L., Li, Y., Zhang, J., Jiang, Y., Han, Y., and Wei, J. (2020). "Electrical load prediction of healthcare buildings through single and ensemble learning." *Energy Reports*, Elsevier Ltd, 6.2751–2767. https://10.1016/j.egyr.2020.10.005.

Chai, J., Liang, T., Lai, K. K., Zhang, Z. G., and Wang, S. (2018). "The future natural gas consumption in China: Based on the LMDI-STIRPAT-PLSR framework and scenario analysis." *Energy Policy*, https://10.1016/j.enpol.2018.04.049.

Chen, S., Yoshino, H., Levine, M. D., and Li, Z. (2009). "Contrastive analyses on annual energy consumption characteristics and the influence mechanism between new and old residential buildings in Shanghai, China, by the statistical methods." *Energy and Buildings*, 41(12).1347–1359. https://10.1016/j.enbuild.2009.07.033.

Chen, Y., Tan, H., and Berardi, U. (2018). "A data-driven approach for building energy benchmarking using the Lorenz curve." *Energy and Buildings*, Elsevier Ltd, 169.319–331. https://10.1016/j.enbuild.2018.03.066.

Choi, J., Leite, F., and de Oliveira, D. P. (2020). "BIM-based benchmarking for healthcare construction projects." *Automation in Construction*, Elsevier B.V., 119. https://10.1016/j.autcon.2020.103347.

Chung, W. (2011). "Review of building energy-use performance benchmarking methodologies." *Applied Energy*.

Cox, M., Brown, M. A., and Sun, X. (2013). "Energy benchmarking of commercial buildings: A low-cost pathway toward urban sustainability." *Environmental Research Letters*, 8(3). https://10.1088/1748-9326/8/3/035018.

Curtis, M. J., Bond, R. A., Spina, D., Ahluwalia, A., Alexander, S. P. A., Giembycz, M. A., Gilchrist, A., Hoyer, D., Insel, P. A., Izzo, A. A., Lawrence, A. J., Macewan, D. J., Moon, L. D. F., Wonnacott, S., Weston, A. H., and McGrath, J. C. (2015). "Experimental design and analysis and their reporting: New guidance for publication in BJP." *British Journal of Pharmacology*.

Ding, Y., and Liu, X. (2020). "A comparative analysis of data-driven methods in building energy benchmarking." *Energy and Buildings*, Elsevier B.V., 209. https://10.1016/j.enbuild.2019.109711.

Duarte, C., Raftery, P., and Schiavon, S. (2016). "SinBerBEST technology energy assessment report."

Ettorchi -Tardy, A., Levif, M., and Michel, P. (2012). "Benchmarking: A method for continuous quality improvement in health." *Healthcare Policy*, https://10.12927/hcpol.2012.22872.

Feibert, D. C., Andersen, B., and Jacobsen, P. (2019). "Benchmarking healthcare logistics processes–a comparative case study of Danish and US hospitals." *Total Quality Management and Business Excellence*, https://10.1080/14783363.2017.1299570.

Ford, E. D. (2005). "A Primer of Ecological Statistics." *The American Statistician*, https://10.1198/tas.2005.s32.

Foucquier, A., Robert, S., Suard, F., Stéphan, L., and Jay, A. (2013). "State of the art in building modelling and energy performances prediction: A review." *Renewable and Sustainable Energy Reviews*, Pergamon.

Francisco, A., Mohammadi, N., and Taylor, J. E. (2020). "Smart City Digital Twin-Enabled Energy Management: Toward Real-Time Urban Building Energy Benchmarking." *Journal of Management in Engineering*, 36(2). https://10.1061/(ASCE)ME.1943-5479.0000741.

Fry, D. E., Pine, M., Nedza, S. M., Locke, D. G., Reband, A. M., and Pine, G. (2016). "Benchmarking hospital outcomes for improvement of care in Medicare elective colon surgery." *American Journal of Surgery*, https://10.1016/j.amjsurg.2016.01.037.

Gao, J., Zhong, X., Cai, W., Ren, H., Huo, T., Wang, X., and Mi, Z. (2019). "Dilution effect of the building area on energy intensity in urban residential buildings." *Nature Communications*, Springer US, 10(1).1–9. https://10.1038/s41467-019-12852-9.

Gao, X., and Malkawi, A. (2014). "A new methodology for building energy performance benchmarking: An approach based on intelligent clustering algorithm." *Energy and Buildings*, Elsevier Ltd, 84.607–616. https://10.1016/j.enbuild.2014.08.030.

Ghajarkhosravi, M., Huang, Y., Fung, A. S., Kumar, R., and Straka, V. (2019). "Energy benchmarking analysis of multi-unit residential buildings (MURBs) in Toronto, Canada."

Journal of Building Engineering, https://10.1016/j.jobe.2019.100981.

González González, A., García-Sanz-Calcedo, J., and Rodríguez Salgado, D. (2018a). "Evaluation of Energy Consumption in German Hospitals: Benchmarking in the Public Sector." *Energies*, MDPI AG, 11(9).2279. https://10.3390/en11092279.

González González, A., García-Sanz-Calcedo, J., and Salgado, D. R. (2018b). "A quantitative analysis of final energy consumption in hospitals in Spain." *Sustainable Cities and Society*, Elsevier Ltd, 36.169–175. https://10.1016/j.scs.2017.10.029.

Graafland, J. J., and Eijffinger, S. C. W. (2004). "Corporate social responsibility of Dutch companies: Benchmarking, transparency and robustness." *Economist*, 152(3).403–426. https://10.1023/B:ECOT.0000036578.63438.30.

Groene, O., Klazinga, N., Kazandjian, V., Lombrail, P., and Bartels, P. (2008). "The World Health Organization Performance Assessment Tool for quality improvement in hospitals (PATH): An analysis of the pilot implementation in 37 hospitals." *International Journal for Quality in Health Care*, https://10.1093/intqhc/mzn010.

Hastie, T., and Tibshirani, R. (1986). "Generalized additive models." *Statistical Science*, https://10.1214/ss/1177013604.

Hawkins, D., Hong, S. M., Raslan, R., Mumovic, D., and Hanna, S. (2012). "Determinants of energy use in UK higher education buildings using statistical and artificial neural network methods." *International Journal of Sustainable Built Environment*, https://10.1016/j.ijsbe.2012.05.002.

Hong, L., Zhou, N., Feng, W., Khanna, N., Fridley, D., Zhao, Y., and Sandholt, K. (2016). "Building stock dynamics and its impacts on materials and energy demand in China." *Energy Policy*, Elsevier, 94(June 2015).47–55. https://10.1016/j.enpol.2016.03.024.

Hong, T., Yang, L., Xia, J., and Feng, W. (2014). "Building energy benchmarking between the United States and China: Methods and challenges." *Lecture Notes in Electrical*

Engineering, 263 LNEE(VOL. 3).473–486. https://10.1007/978-3-642-39578-9_50.

IPCC. (2014). *Climate Change 2014 Mitigation of Climate Change*. *Climate Change 2014 Mitigation of Climate Change*.

Jenkins, D. G., and Quintana-Ascencio, P. F. (2020). "A solution to minimum sample size for regressions." *PLoS ONE*, https://10.1371/journal.pone.0229345.

Ji, R., and Qu, S. (2019). "Investigation and evaluation of energy consumption performance for hospital buildings in China." *Sustainability (Switzerland)*, 11(6). https://10.3390/su11061724.

Jiang, Y., Guo, S., and Hu, S. (2018). "China Building Energy Use 2018 China Building Energy Use 2018 Building Energy Research Center of Tsinghua University." (December).

Jowett, I. G., Parkyn, S. M., and Richardson, J. (2008). "Habitat characteristics of crayfish (Paranephrops planifrons) in New Zealand streams using generalised additive models (GAMs)." *Hydrobiologia*, https://10.1007/s10750-007-9108-z.

Kamaluddin, K. A., Imran, M. S., and Yang, S. S. (2016). "Development of Energy Benchmarking of Malaysian Government Hospitals and Analysis of Energy Savings Opportunities." *Journal of Building Performance*, 7(1).72–87.

Keirstead, J. (2013). "Benchmarking urban energy efficiency in the UK." *Energy Policy*, https://10.1016/j.enpol.2013.08.063.

Kim, J. B., Jeong, W., Clayton, M. J., Haberl, J. S., and Yan, W. (2015). "Developing a physical BIM library for building thermal energy simulation." *Automation in Construction*, Elsevier B.V., 50(C).16–28. https://10.1016/j.autcon.2014.10.011.

Kontokosta, C. E., Spiegel-Feld, D., and Papadopoulos, S. (2020). "The impact of mandatory energy audits on building energy use." *Nature Energy*, Springer US, 5(4).309–316. https://10.1038/s41560-020-0589-6.

Lavery, M. R., Acharya, P., Sivo, S. A., and Xu, L. (2019). "Number of predictors and multicollinearity: What are their effects on error and bias in regression?" *Communications in Statistics: Simulation and Computation*, https://10.1080/03610918.2017.1371750.

Leathwick, J. R., Elith, J., and Hastie, T. (2006). "Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions." *Ecological Modelling*, https://10.1016/j.ecolmodel.2006.05.022.

Li, Y., Cao, L., Han, Y., and Wei, J. (2020). "Development of a Conceptual Benchmarking Framework for Healthcare Facilities Management: Case Study of Shanghai Municipal Hospitals." *Journal of Construction Engineering and Management*, 146(1). https://10.1061/(ASCE)CO.1943-7862.0001731.

Liu, Y., Paciorek, C. J., and Koutrakis, P. (2009). "Estimating regional spatial and temporal variability of PM2.5 concentrations using satellite data, meteorology, and land use information." *Environmental Health Perspectives*, https://10.1289/ehp.0800123.

Lomas, K. J., and Ji, Y. (2009). "Resilience of naturally ventilated buildings to climate change: Advanced natural ventilation and hospital wards." *Energy and Buildings*, https://10.1016/j.enbuild.2009.01.001.

Morgenstern, P., Li, M., Raslan, R., Ruyssevelt, P., and Wright, A. (2016). "Benchmarking acute hospitals: Composite electricity targets based on departmental consumption intensities?" *Energy and Buildings*, Elsevier B.V., 118.277–290. https://10.1016/j.enbuild.2016.02.052.

NHCPRC. (2014). *Code for design of general hospital: GB 51039-2014. China Planning Press.*

Papadopoulos, S., and Kontokosta, C. E. (2019). "Grading buildings on energy performance using city benchmarking data." *Applied Energy*.

Park, H. S., Lee, M., Kang, H., Hong, T., and Jeong, J. (2016). "Development of a new energy benchmark for improving the operational rating system of office buildings using various data-mining techniques." *Applied Energy*, Elsevier Ltd, 173.225–237.

https://10.1016/j.apenergy.2016.04.035.

Roth, J., Brown IV, H. A., and Jain, R. K. (2020). "Harnessing smart meter data for a Multitiered Energy Management Performance Indicators (MEMPI) framework: A facility manager informed approach." *Applied Energy*, Elsevier Ltd, 276.115435. https://10.1016/j.apenergy.2020.115435.

Roth, J., and Rajagopal, R. (2018). "Benchmarking building energy efficiency using quantile regression." *Energy*, Elsevier Ltd, 152.866–876. https://10.1016/j.energy.2018.02.108.

Rstudio. (2020). "Integrated Development for R. RStudio." PBC, Boston, MA.

Shanghai Municipal People's Government. (2021). "Shanghai Municipal People's Government." <https://www.shanghai.gov.cn/nw2319/index.html> (May 6, 2021).

Sharma, V., Caldas, C. H., and Mulva, S. P. (2020). "Identification and Prioritization of Factors Affecting the Overall Project Cost of Healthcare Facilities." *Journal of Construction Engineering and Management*, 146(2).04019106. https://10.1061/(asce)co.1943-7862.0001692.

Singer, B. C., Mathew, P., Greenberg, S., Tschudi, W., Sartor, D., Strom, S., Vernon, W., Nash, M., and Burch, L. (2009). *Hospital Energy Benchmarking Guidance - Version 1 . 0. Hospital Energy Benchmarking Guidance*.

SMMCHUD, and SDRC. (2019). *Energy consumption monitoring and analysis report of Shanghai state agency office buildings and large public buildings at 2018*.(in Chinese).

Sokratis Papadopoulos. (2019). "Machine Learning Applications To Advance the Field of Urban Energy Dynamics." (May).1–2.

Sowby, R. B., and Burian, S. J. (2018). "Statistical Model and Benchmarking Procedure for Energy Use by US Public Water Systems." *Journal of Sustainable Water in the Built Environment*, 4(4). https://10.1061/JSWBAY.0000864.

Tejero-González, A., DeFreitas-Barros-Galvão, V. M., Zarzuelo-Sánchez, A. M., and SanJosé-

Alonso, J. F. (2020). "Energy use optimization in ventilation of operating rooms during inactivity periods." *Building Research and Information*, Taylor & Francis, 0(0).1–17. https://10.1080/09613218.2020.1817723.

Teng, Z. R., Wu, C. Y., and Xu, Z. Z. (2017). "New energy benchmarking model for budget hotels." *International Journal of Hospitality Management*, Elsevier Ltd, 67.62–71. https://10.1016/j.ijhm.2017.07.010.

Vasileiou, K., Barnett, J., Thorpe, S., and Young, T. (2018). "Characterising and justifying sample size sufficiency in interview-based studies: Systematic analysis of qualitative health research over a 15-year period." *BMC Medical Research Methodology*, https://10.1186/s12874-018-0594-7.

Wang, T., Li, X., Liao, P. C., and Fang, D. (2016). "Building energy efficiency for public hospitals and healthcare facilities in China: Barriers and drivers." *Energy*, Elsevier Ltd, 103.588–597. https://10.1016/j.energy.2016.03.039.

Wei, Z., Xu, W., Wang, D., Li, L., Niu, L., Wang, W., Wang, B., and Song, Y. (2018). "A study of city-level building energy efficiency benchmarking system for China." *Energy and Buildings*, https://10.1016/j.enbuild.2018.08.038.

Yang, Z., Roth, J., and Jain, R. K. (2018). "DUE-B: Data-driven urban energy benchmarking of buildings using recursive partitioning and stochastic frontier analysis." *Energy and Buildings*, Elsevier B.V., 163.58–69. https://10.1016/j.enbuild.2017.12.040.

Zimmerman, A., and Kibert, C. J. (2007). "Informing LEED's next generation with the Natural Step." *Building Research and Information*, 35(6).681–689. https://10.1080/09613210701342367.

Zou, B., Chen, J., Zhai, L., Fang, X., and Zheng, Z. (2017). "Satellite based mapping of ground PM2.5 concentration using generalized additive modeling." *Remote Sensing*, 9(1).1–16. https://10.3390/rs9010001.