Received 18 July 2014
Revised 24 November 2014
Accepted 27 December 2014

Benchmarking hospitals through a web based platform

Maria Conceição A. Silva Portela
CEGE, Católica Porto Business School, Universidade Católica Portuguesa, Porto, Portugal
Ana Santos Camanho
Faculdade de Engenharia, Universidade do Porto, Porto, Portugal
Diogo Queiroz Almeida
Instituto de Ciências Biomédicas Abel, Universidade Católica Portuguesa, Porto, Portugal
Luiz Lopes
Católica Porto Business School, Universidade Católica Portuguesa, Porto, Portugal
Sofia Nogueira Silva
CEGE, Católica Porto Business School, Universidade Católica Portuguesa, Porto, Portugal, and
Ricardo Castro
Faculdade de Engenharia, Universidade do Porto, Porto, Portugal

Abstract

Purpose – In a context of international economic crisis the improvement in the efficiency and productivity of public services is seen as a way to maintain high-quality levels at lower costs. Increased productivity can be promoted through benchmarking exercises, where key performance indicators (KPIs), individually or aggregated, are used to compare health units. The purpose of this paper is to describe a benchmarking platform, called Hospital Benchmarking (HOBE), where hospital's services are used as the unit of analysis.
Design/methodology/approach – HOBE platform includes a set of managerial indicators through which hospital services' are compared. The platform also benchmarks services through aggregate service indicators, and provides an aggregate measure of hospital's performance based on a composite indicator of the service's performances. These aggregate indicators were obtained through data envelopment analysis (DEA).
Findings – Some results are presented for Portuguese hospitals for the trial years of 2008 and 2009, for which data is publicly available. Details for the service-level analysis are provided for a sample hospital, as well as details on the aggregate performance resulting from services performances.
Practical implications – HOBE's features and outcomes show that the platform can be used to guide management actions and to support the design of health policies by administrative authorities, provided that good quality and timely data are available, and that hospitals are involved in the design of the KPIs.
Originality/value – The platform is innovative in the sense that it bases its analysis on hospital's services, which are in general more comparable among hospitals than indicators of hospital overall performance. In addition, it makes use of DEA to aggregate performance indicators, allowing for user choice in the inputs and outputs to be aggregated, and it proposes a novel model to aggregate service's efficiencies into a single measure of hospital performance.

Keywords Benchmarking, Performance measurement (managerial), Data envelopment analysis (DEA), Health services, Web platforms, Aggregate performance measures, Portuguese health system
Paper type Technical paper

The authors acknowledge the financial support of the Portuguese Foundation for Science and Technology (FCT) through project PTDC/EGE-GES/112232/2009.

Benchmarking: An International Journal
Vol. 23 No. 3, 2016
pp. 722-739

1463-5771
DOI 10.1108/BIJ-07-2014-0067

1. Introduction

Benchmarking is the process through which organizations can learn from the best-in-class companies. As a result, benchmarking can be seen as a way to understand the current situation of a company in relation to others, as well as a process towards sharing practices, learning, and improvement (see e.g. Voss *et al.*, 1997). Within the health sector, benchmarking practices are increasing in most places of the globe as a way to control costs, increase efficiency, and increase quality. The importance of hospital benchmarking is undisputable and benefits many stakeholders: governments are interested on comparing hospital's costs and quality for financial reasons and for designing public health policies; hospital managers have interest on knowing the performance of their hospitals in comparison with similar entities; the general public is interested on comparisons between hospitals' quality, such that informed choices of the health provider are possible.

Therefore, multiple perspectives are at stake in hospitals' benchmarking. The literature mainly distinguishes between a clinical perspective and a managerial perspective that interact with each other (Chilingerian and Sherman, 2004). These can also be seen as the efficiency and quality perspectives assessed in Nayar *et al.* (2013), which, taken together, allow the measurement of the "value" of the health care provided. Clearly the managerial or efficiency perspectives are easier to assess than the quality perspective, but the latter is of particular use to clinical managers and to the general public.

The purpose of this paper is to present a novel benchmarking tool named HOBE (HOspital BEnchmarking – available at feg.porto.ucp.pt/hobe) and to show how an analytical benchmarking frameworks can be used by hospital and clinical service managers to gain insight of the current performance of their service(s) in relation to their peers and to improve their decision-making process and efficiency.

The potential users of HOBE are at the moment hospital managers and public authorities, as the indicators it presents are "biased" towards a managerial perspective. In spite of the quality dimension being considered within HOBE, difficulties in gathering detailed data implied that quality is currently not fully addressed. HOBE has been built based on data of Portuguese hospitals for the years 2008 and 2009, but the tool can be applied to any health system. The interest for the Portuguese case lies on the fact that benchmarking web-based tools have been proposed by the Troika team, which provided financial assistance during the recent financial crisis.

HOBE includes several functionalities that are not currently available in benchmarking platforms. In particular, it benchmarks medical and surgical services within the hospitals, rather than the hospital as a whole. As will be seen in the next section, more micro analysis of health units are generally recommended, but few studies or benchmarking tools use other level of analysis than the hospital. The benchmarking is performed individually for each indicator, but HOBE also incorporates an aggregate service performance measure, based on a set of selected indicators. Each hospital can also have access to an aggregate measure of performance, which aggregates its services' performances. The aggregation of indicators in HOBE is done through data envelopment analysis (DEA).

HOBE's use of DEA to aggregate indicators is a novelty in the health context, but not in other contexts. Bogetoft and Nielsen (2005) were amongst the first to present benchmarking solutions that use DEA and other frontier techniques integrated in online platforms. Another example is a web benchmarking tool that integrates DEA to compare warehouses described in Johnson and McGinnis (2011). In Portugal

there is a free web tool that allows the benchmarking of Portuguese secondary schools, named BESP – Benchmarking Escolas Secundárias Portuguesas) described in Portela *et al.* (2011). BESP is a visual tool showing the performance of selected schools on individual indicators, based on a percentile score. BESP has a customized tool that allows the selection of indicators to be used in the construction of a composite performance indicator for each school. Such aggregate indicator is constructed using DEA. The HOBE platform that we present in this paper shares some similarities with the BESP platform.

In conclusion, the main contribution of the HOBE platform described in this paper lies on the calculation in real time of aggregate performance measures, which are constructed at two levels: the service and the hospital. In addition, the research described in this paper contributes to the literature by analysing hospitals' performance through the performance of the clinical services the hospital provides.

2. State of the art

Benchmarking is the process of "finding and implementing best practices" (Camp and Tweet, 1994). This search from best practices can be internal or external to the company. As a result, there are, amongst others, two important types of benchmarking. Internal benchmarking – when similar internal functions are compared; and competitive benchmarking – when a company (in whole or in parts) is compared to competitors (see e.g. Camp and Tweet, 1994). Mosel and Gift (1994) introduced the concept of collaborative benchmarking in health care that involves a combination of internal and external (or competitive) benchmarking, where the notion of "internal" is extended from one company to several companies belonging to a collaborative benchmarking network. The focus of the current paper is on competitive benchmarking.

Health benchmarking has been growing over the last decade for various interrelated reasons: first, the growing number of consultancy firms offering benchmarking services, second, the growing interest of governments in cutting costs, and therefore implementing mechanisms to control and assess them (together with effectiveness), third, the growing number of published research papers performing comparative analysis of health institutions (although not necessarily within broader benchmarking frameworks).

Being HOBE a web platform, our analysis of the state of the art in health benchmarking, focus on existing web platforms and on the literature that performs comparative analysis of health institutions through frontier techniques, as this is a novel feature within HOBE.

2.1 Hospital benchmarking platforms

Benchmarking through the web is not a new concept, and has been put into practice by several entities. In the health sector web platforms have been mainly developed by private entities that provide consultancy in the area of benchmarking, amongst other. There is a growing number of consultancy firms providing benchmarking services to hospitals and other health institutions. Some of these companies also provide a public service by publishing hospital rankings (e.g. US News Best Hospitals, or the Truven Health Analytics 100 Top Hospitals in the USA), by providing hospital guides (e.g. *Dr. Foster Hospital Guide*), or by awarding prizes to performance (e.g. Caspe Healthcare Knowledge System awards). In spite of providing some outcomes of the benchmarking exercises to the public, the actual access to the

benchmarking platforms of the above consultancy firms is only opened to the health institutions that purchased their services. As a result, in our revision of some web benchmarking tools, we focused on those that are available to the general public and can be consulted without restrictions.

In Table I, we show some functionalities of HOBE and some selected web benchmarking tools (most of which provided by public or governmental entities).

Generally publicly available benchmarking tools have information available at an aggregate level (the Acute Trust Dashboard benchmarks trusts in the UK, whereas the Care Quality Commission and ACSS (Administração Central do Sistema de Saúde) benchmark hospitals, in the UK and Portugal, respectively). Disease level of analysis is less common in publicly available tools (but are very common in commercialized tools). An interesting exception is the Portuguese SINAS (Sistema Nacional de Avaliação em Saúde) website developed by the Portuguese regulatory entity for health (ERS, 2016). SINAS is based on several indicators grouped into: clinical excellence, patient safety, comfort and fit of premises, focus on the patient, and patient satisfaction. The clinical excellence indicators report on the adequacy of procedures undertaken by the hospital based in diagnosis related group (DRG) patient information. For example, within the orthopedics service the indicators provided concern the surgical treatment of proximal femur fractures and total arthroplasty of the hip and knee. Within the latter, the indicators available include the administration of antibiotics in the hour before the surgery, prescription of prophylaxis recommended for thromboembolism, or interruption of the prophylactic antibiotic in the 24 hours after the surgery. To the author's knowledge the benchmarking at the service level (done within HOBE) has been less common both in the literature (see next section) and in available benchmarking tools.

Regarding interactivity of the benchmarking tools, in some cases this is non-existent, (like the Acute Trust Quality Dashboard, which for each selected hospital uploads a pdf file with a dashboard of the Trust on a number of indicators related to clinical effectiveness, patient experience, patient safety, and organizational context, or the Care Quality Commission that shows inspection reports on five quality dimensions), while in others, like the ACSS benchmarking, there is some level of interaction (indicators can be chosen on the platform, and graphs change depending on the selected indicators and on the selected hospital). HOBE presents the highest degree of interactivity with the user, allowing him/her to choose the indicators to be displayed, the reference set, the variables to be used in the construction of a composite indicator, as well as the hospital services that should be aggregated to compute a composite performance of the hospital.

	HOBE	Acute Trust Quality Dashboard	Care quality commission (UK)	SINAS	ACSS benchmarking
Level of analysis	Service	Trust	Hospital	DRG	Hospital
Interactivity	☑	☒	☒	☑	☑
Managerial info	☑	☒	☒	☒	☑
Quality info	☑	☑	☑	☑	☑
DEA	☑	☒	☒	☒	☒
Rankings/rating	☒	☒	Rating	Rating	Ranking
URL	feg.porto.ucp.pt/hobe	analytics.methods.co.uk/atqd/	www.cqc.org.uk/	www.ers.pt/	acss.min-saude.pt/benchmarking.aspx

Table I. Some benchmarking tools on the web and their functionalities

Most available tools present some measures of quality of the hospitals because their ultimate aim is to inform the general public and assist them in their provider's choice. However, the level of the information provided can vary considerably, from a high level of detail, as in SINAS, to just a few indicators of quality (as in ACSS benchmarking tool, where the quality variables considered are the readmissions within 30 days, percentage of cesarean births, percentage of hip fractures in the 48 h after surgery, and percentage of length of stay higher than 30 days). Managerial information is less common in publicly available platforms. Again ACSS benchmarking tool is mainly a managerial benchmarking tool because ACSS is the entity that manages the financing of Portuguese hospitals. HOBE, on the other hand, has been designed with the objective of serving both hospital managers and the general public, and therefore it provides detailed managerial performance indicators, which are of use for the former. HOBE also provides information on some quality indicators, but the level of detail is low (compared to that of ACSS).

Regarding the existence of rankings or ratings not all tools have this functionality. SINAS has a three-level rating system, with level III denoting superior performance and rating I base performance. The Care Quality Commission (an English independent regulator of health and social care) has also a four-level rating system (outstanding, good, requires improvement, and inadequate) for each of the dimensions inspected in hospital services (safety, effectiveness, care, leadership, and responsive to people's needs). On the other hand, ACSS benchmarking website, shows for each indicator, hospitals ranked from the best performer to the worst. HOBE does not present explicitly any ranking of hospitals. For each indicator the hospital knows its position in a rank (through the percentile), but does not know which hospitals are in the remainder positions. For the aggregate performance measures, the hospital knows its relative performance measure but does not know its position on a ranking. The option for not building rankings relates to the fact that differences between two places in a rank may be irrelevant, but this is not the general idea when a ranking is presented to the general public. The difference between being ranked first or third (e.g.) is perceived as significant, when in fact the values dictating the first and third position may be very close to each other. That is why within HOBE we opted by not showing ranks.

2.2 Hospital evaluation studies

Frontier techniques can be seen as a suitable tool for the identification of benchmarks or best practices (see Hougaard and Tvede, 2002) within a benchmarking framework. This is because these techniques allow for the aggregation of individual indicators into a meaningful aggregate performance measure (representing technical efficiency, cost efficiency, effectiveness, or just aggregate performance) simultaneously allowing for the consideration of trade-offs between indicators (Bogetoft and Lars, 2011).

The number of studies applying frontier techniques to comparing health institutions has increased dramatically in recent years (Hollingsworth, 2008). The technique most frequently used in health evaluation studies is DEA, a technique developed by Charnes *et al.* (1978). Stochastic Frontier Analysis is also gaining popularity, partly due to the limitations of DEA in handling random errors (Hollingsworth, 2008).

The analysis on health care literature can range from something as small as a computerized tomography machine to something as big as a country's health system. Hospitals have, however, been the most common object of study, as in general they represent the biggest share of the public expenditure in health care (Hollingsworth, 2008). The use of hospitals as the unit of analysis in benchmarking exercises presents

some problems, as in general hospitals are complex and heterogeneous organizations, and should not be directly compared. To overcome this limitation, Hollingsworth (2008) and Chilingerian and Sherman (2011) suggest focusing in smaller, more homogeneous, units (like services, or practitioners).

When evaluating hospitals one can take (at least) two perspectives into account: the manager's perspective and the clinician's perspective (Chilingerian and Sherman, 2011). Under a managerial perspective a health unit is considered an entity transforming medical and other resources into intermediate outputs. Under a clinical perspective, efficiency requires that a health unit utilizes the minimal quantity of intermediate outputs or clinical resources (such as consultations, referrals, treatments, and drugs) to achieve a constant quality outcome, when caring for patients with similar diagnosis complexity and severity. In the health context it is usual to distinguish between inputs (or resources), outputs, and outcomes, where the latter are mainly related with the quality of the care provided. This distinction has its origin in the 1966 work of Donabedian (currently reprinted as Donabedian, 2005), where the author provides a framework to assess health care quality focusing on structure, process, and outcome measures. Structure is seen as the setting where care is given, and include infrastructures' quality, human resources' qualifications, and support processes and their efficiency (structure therefore can be seen as inputs). Process denotes what is actually done in giving and receiving care (Donabedian, 1988), including thus the physician performance and the patient intervention (process can be seen as the outputs). Finally, outcomes focus on the effects of all the layers of care in the actual health status of patients and populations.

Most studies in the literature focus on the managerial perspective, but the number of studies focusing on the clinician's perspective is increasing (Hollingsworth, 2008). Some examples of studies focusing on a managerial perspective of hospitals are: Grosskopf *et al.* (2001), Ozcan (2008), Sikka *et al.* (2009), Simões and Marques (2011), Kawaguchi *et al.* (2013). Examples of studies applying a clinical perspective can be found in Retzlaff-Roberts *et al.* (2004), Navarro-Espigares (2011), Nayar *et al.* (2013), or Stevanovic *et al.* (2005). The difficulty associated with the clinical perspective is the measurement of outcomes in health care, which are difficult to quantify and to obtain.

Not many studies have focused on comparing hospital's services. Some examples can be found in Laudicella *et al.* (2010), where obstetrician departments were compared, Puig-Junoy (1998) and Dervaux *et al.* (2009), who compared intensive care units of hospitals, or Kontodimopoulos and Niakas (2005), who compared Greek hemodialysis units. Groups of diseases have also been analysed in Kristensen *et al.* (2010) (diabetes) or Dismuke and Sena (2001) (heart failure and shock). In the Portuguese context, and to the author's knowledge, the first study applied at the service level was that of Castro (2011), who assessed the efficiency of the "internal medicine" service using data from 2008 and the DEA methodology. In a subsequent study, using the same data set, Almeida (2013) analysed the efficiency of several services of Portuguese hospitals for the year 2008 using DEA, and provided a measure of hospital performance that was an aggregation of the service efficiencies of the hospital. This is the approach that served as a basis for the development of the aggregate performance measure available in the HOBE platform, and it is therefore explained in more detail in the next sections. Other examples of studies applied to the Portuguese context can be found in Barros (2003), Moreira (2008), Menezes *et al.* (2006), Gonçalves (2008), Rego, *et al.* (2010), or Barros *et al.* (2013). These studies, however, focus on hospital's comparisons and on the impact of the management practices' reform in their performance (legislation in 2002 and also

later in 2005) implied a significant change in the management of hospitals, involving the separation between the financing and the service provider's role, which were accumulated by the State up to 2002).

3. Methodology

The methodology employed in the creation of HOBE involved several steps. The first step was the identification of all the requirements for the website and their priority. Functional requisites involved, among many others, the visualization of indicators, their formulas and explanation, the visualization of graphs with the indicators, their percentiles and their evolution over time. Non-functional requirements concerned the adequability of the website to the user needs, accuracy, attractiveness, comprehensiveness, efficiency and speed of computation, stability, etc.

The three most important steps of the HOBE creation process were the definition of the indicators, the data collection and organization into a database, and the definition of the models to aggregate the individual indicators.

In order to define the indicators, the team met with some hospital managers to gather their views on the most important indicators for their management practices. In addition, the team contacted with ACSS – the entity to which hospitals report data periodically – in order to have access to some data from all Portuguese hospitals. From the top management perspective clinical and managerial information are equally relevant, but service managers prefer clinical detailed information (as well as the general public). However, in order to get detailed clinical information one would need to use the DRG records for each episode. This information is available for all Portuguese hospitals, but hospitals and patients are anonymized and services are coded differently between hospitals. As a result, the indicators developed in HOBE focused primarily on managerial information and subsequently on quality information, but quality is only available at an aggregate level. Note that the consideration of a limited number of quality indicators (the use of aggregate-level indicators cannot account for many dimensions of service quality of a service) is a limitation in a benchmarking framework of services. As Gummesson (1998) notes, quality, profitability, and productivity are triplets that should be accounted for together, being impossible to dissociate in a service provider the productivity dimension (or efficiency) from quality (see also Grönroos and Ojasalo, 2004).

Performance indicators in HOBE are divided into three main types: cost, quantity, and quality. The rationale for the separation into quantity and cost indicators is related to the fact that management deals mostly with volumes of outputs and volumes of resources, but their cost should also be taken into account, since the price of providing a certain service or acquiring certain resources may vary substantially across hospitals. In addition, in a context of economic difficulties, cost indicators assume particular importance. For ease of presentation we further divided quantity and cost indicators into those related to resources and those related to the outputs produced by the hospital. Overall a number of 80 indicators have been defined for 19 clinical services (those that were present in more than 15 national hospitals). We will not reproduce all indicators in this paper for sake of brevity. All indicators can be consulted in the website (in Portuguese) or provided by the authors upon request. Examples of indicators under each category are the following:

- cost indicators for the outputs produced: daily cost of inpatient admission, cost per outpatient, or hourly cost per surgery in operating room;

- cost indicators for resources: hourly cost for medical staff, hourly cost for nursing staff, or annual cost of clinical material per bed;
- volume indicators for resources: bed occupancy rate, weekly hours of work per doctor, or number of nurses per bed;
- volume indicators for the outputs produced: patients discharged from inpatient admissions per bed, average days of inpatient admission, or number of surgeries in the operating room per doctor; and
- quality indicators: mortality rate; re-admission rate, per cent of caesarian births, per cent mortality at birth, waiting times in emergency, waiting times for surgery.

Note that all indicators, except quality indicators are shown at the service level. Quality indicators are only available at the hospital level, but can be displayed at the service level if data were supplied to the platform (that data are not available at the moment).

After defining the indicators, the database that was going to feed HOBE was created. Accounting and production data supplied by ACSS were the main sources of data, provided in several disperse Excel files for two years (2008 and 2009). Data relates to almost all public hospitals in the country. We are not working with the full set of hospitals because some did not report data to ACSS in some years. For the first stage of HOBE construction it was decided not to import directly data from hospitals, as this would be a long process and implied the commitment of most hospitals in the country. As a result, the construction of the database involved the creation of routines to extract the variables from the corresponding available files. The database was constructed in SQLserver.

As mentioned above, after defining indicators and gather the data, the definition of the aggregating model of service performance was undertaken. As most of the resource variables could only be considered in cost terms (e.g. drugs, clinical material, complementary means of diagnosis, etc. could not be considered in aggregate meaningful volume measures) a perspective of cost efficiency was taken to construct an aggregate measure of service performance. The variables that were considered to construct this aggregate performance measure are shown in Table II.

The set of variables in Table II is the maximum available set of variables. This set can be customized to the service (e.g. some services may not perform surgeries), or according to the will of the user (details in the next section).

Using the overall or a subset of the variables in Table II, aggregate performance for each service is computed in HOBE through the DEA cost efficiency Model (1)

Inputs (costs)	Outputs
Costs with other supplies and services	Number of inpatient days
Costs with medical doctors	Number of days in ICU (intensive care units)
Costs with nurses	Number of first outpatient appointments
Costs with drugs	Number of patients assisted in emergency
Costs with clinical material	Number of ambulatory surgery
Costs with CDT (complementary means of diagnosis and therapy)	Number of conventional surgeries
Costs with surgery wards	Number of urgent surgeries

Table II. Complete set of inputs and outputs that can be chosen in HOBE

shown in Appendix. The resulting service efficiency score shows the maximum reduction to total cost that can be achieved keeping the current output production.

Note that ideally one should consider on the output side also quality variables. Indeed, the most cost effective services (those delivering their output at a minimum cost) are not necessarily the ones providing better quality of care. In fact the opposite can happen, as cost efficiency may be achieved at the expense of lower level of service quality (see e.g. Grönroos and Ojasalo, 2004). The consideration of quality variables on the output side would guarantee that the most cost efficient services were also the ones providing a service quality comparable or above that of its peers. In the case of HOBE we were limited by the availability of quality measures only at the hospital level, and could not incorporate this dimension into the analysis. We acknowledge, however, this as an important limitation of the study and one where future developments should focus.

After obtaining a cost efficiency measure for each service we further aggregated the service efficiencies into an aggregate indicator of performance for each hospital. This was computed based on the DEA Model (2) shown in the Appendix. This model uses a virtual input equal to one, and the outputs are the efficiencies of the services that the hospital wishes to aggregate. The use of DEA for performance assessments focusing only on achievements, rather than the conversion of inputs to outputs, in fact corresponds to the construction of a composite indicator. The unitary input underlying the evaluation of every decision making unit (DMU) can be seen a "helmsman" attempting to steer the DMU towards the maximization of outputs.

4. Results

4.1 HOBE's outcomes on individual indicators

The benchmarking functionality within HOBE aims at comparing, for each indicator, the hospital's service under assessment with a selected sample of hospitals. To perform benchmarking for a given hospital, the user must choose the service, the sample for comparison, and the year of assessment. The universe of comparison can be defined according to different criteria: regional health administration (North, Algarve, Alentejo, Centre, and Lisbon & Tejo Valley), financing group (Groups 1-4 – defined by ACSS for financing purposes), level of specialization (specialized vs non-specialized), university or non-university hospitals, and level of case mix (where we allow the comparative set to be constituted by hospitals that show a similar or more complex case mix than the hospital chosen).

The benchmarking results shown in the platform for each indicator appear in the form of graphs (see figures, imported from HOBE). Figure 1 shows the type of information provided for each individual indicator within HOBE. We show, on the left

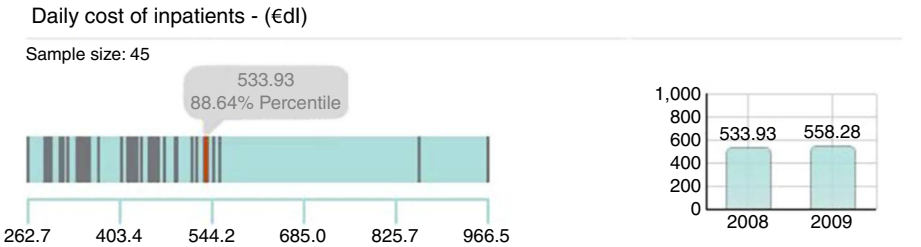

Figure 1.
Example of graphs shown for each indicator

hand side, a graph created by the team, which shows the distribution of values observed for all hospitals in the sample for a particular indicator (daily cost of inpatients in this case), along with the percentile where the hospital under analysis is situated. Figure 1 shows that for the service of general surgery of the hospital under analysis, the value of the indicator daily cost with inpatients is 533.93, corresponding to the percentile 88.64 per cent. The graph on the right hand side shows how the indicator has evolved over time for the service of this hospital.

Figure 2 illustrates a synthesis of cost indicators associated to the outputs produced by the general surgery service of the hospital.

The radar graph on the left hand side of Figure 2 shows the percentile where the particular service of the hospital stands on each indicator of the category chosen. As we selected the dimension relating to the cost of production the indicators shown are: hourly cost of surgery rooms (€hSR), cost per surgery (€/S), cost per outpatient surgery (€/OS), cost per emergency patient (€/EP), cost per patient in day procedures (€/PDH), cost per day procedure session (€/DS), cost per outpatient appointment (€/OA), cost of ICU per patient (€ICU/P), daily cost of inpatient in ICU (€dICU), inpatient costs per patient (€I/P), and daily cost of inpatients (€dI).

On the right hand side of Figure 2, it is shown the case mix of the hospital as well as the value of its percentile, which allows contextualizing the hospital in terms of the severity of the cases treated. Note that only the case mix at the hospital level is available. In particular, each hospital has four types of case mix, corresponding to the surgical and medical specialties, each subdivided into inpatients and outpatients. As the case mix may vary substantially among services of the same hospital and this information is not available for Portuguese hospitals, we did not adjust the indicators used in the HOBE platform for case mix. If the service under analysis is related to

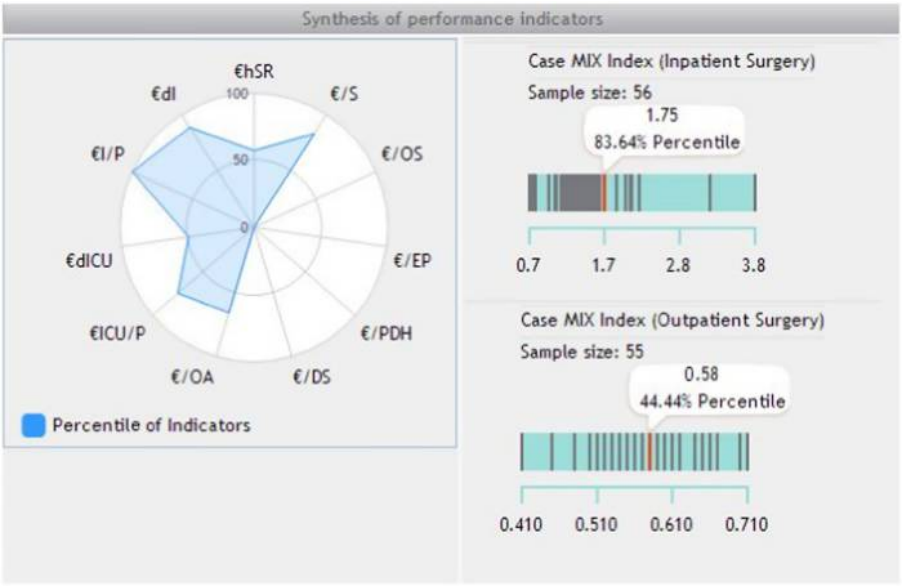

Figure 2.
Synthesis of performance on the set of indicators regarding cost of production

Note: The values of Case Mix Indexes relate to the Hospital as a whole and not to each service, being weighted by the number of patients when aggregated

surgery, the platform shows the case mix values related to all surgery services. Otherwise, the medical indicators of case mix are shown. For the case of this hospital we show the case mix of surgery inpatients and surgery outpatients, where the hospital lies on percentiles 83.64 and 44.44 per cent, respectively. This means that inpatients treated under surgical specialties at this hospital are amongst the most severe cases in the sample (only about 17 per cent hospitals show a higher case mix than this hospital), whereas the patients treated under outpatient surgery are not amongst the most severe cases in the sample of 56 hospitals (sample size is also shown next to the graphs).

4.2 HOBE outcomes on aggregate indicators for each service

In order to illustrate the aggregate service indicators we use the case of a general surgery service of a given hospital (Centro Hospitalar (CH) Setúbal). For this service, the full set of inputs and outputs in Table II was used to compute cost efficiency, but a customized set of variables could have been chosen through the interaction of the user. After the choice of inputs and outputs, HOBE returns an efficiency score for the service, estimated using Model (1) in the Appendix. The selection of appropriate inputs and outputs is crucial to obtain appropriate models for the purpose of the analysis, and as a consequence, robust results.

For the general surgery service of CH Setúbal the efficiency measure is 72.0 per cent (in a sample of 46 general surgery services). This score means that this hospital has scope to reduce overall costs (the sum of the costs considered in the assessment) to 72 per cent of observed levels. In addition to the efficiency score, other information is also shown in HOBE to inform hospitals regarding the potential sources of inefficiency. The radar chart in Figure 3 compares, for the general surgery service of CH Setúbal, the current values of the inputs and outputs with the targets provided by the DEA Model (1). Note that due to different units of measurement of the variables, the radars show normalized values. Thus, the hospital chosen appears in all indicators with a value of 1, and all other values shown in the radar are the targets, represented as a percentage of the current values observed in the service analysed.

Figure 3.
Radar comparing the input costs of the general surgery service with its targets

Figure 3 shows that the general surgery service for this hospital has very high costs for other supply services and medical doctors, when compared to target levels. On the other hand, the costs with complementary means of diagnostics and therapy (CDT) and surgery wards for this service are close to target levels. This means that, when compared to hospitals producing similar outputs for the general surgery service, CH Setúbal shows higher aggregate costs than its peers, and the higher costs are mainly a result of high supply costs and medical doctors costs. This information is also shown in a table, as reported in Figure 4. The knowledge of this information should trigger some actions. In particular it will be important to investigate why this service has shown such higher costs with doctors (does it have too many doctors when compared to other hospitals? Or does it pay higher salaries?), and with other supplies and services (was there any extra cost in this year that triggered this result?).

To pursue with this investigation, HOBE provides some other pieces of information, particularly the benchmarks of this hospital's service. The information reported is shown in Figure 5, where the inputs and outputs of the service analysed and those of its main peer (CH de Vila Nova de Gaia/Espinho, contributing in 52.01 per cent for the peer set) are shown.

The benchmark hospital has higher levels of all outputs (see graph) and similar or lower costs (see graph) for all inputs, except CDT, than the hospital being assessed. The fact that the hospital being assessed spent more money with doctors than with

Inputs | Table

Inputs	Observed Values	Target Values	% of Change in Costs
Other Sup. Services	226069.33	84516.74	−62.61%
Medical Doctors	970733.87	254314.89	−73.80%
Nurses	732979.49	569763.70	−22.27%
Drugs	640356.04	516113.13	−19.40%
Clinical Material	122439.15	89634.43	−26.79%
CDT	890941.94	843274.69	−5.35%
Surgery Wards	857164.25	841361.76	−1.84%

Figure 4.
Table comparing the input costs of the general surgery service with its targets

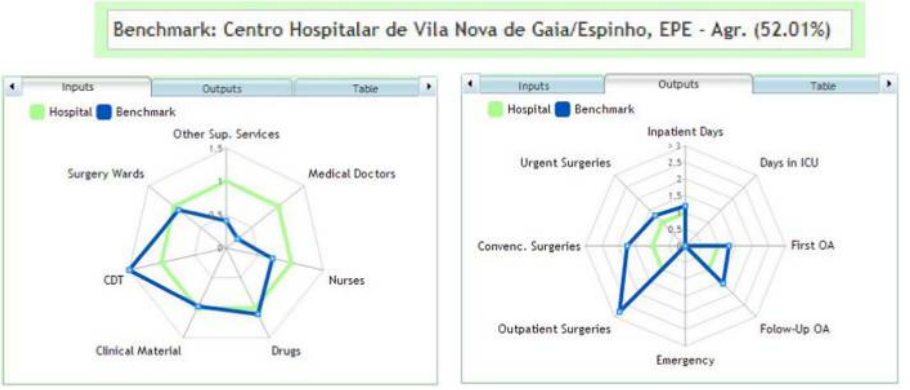

Figure 5.
Radars showing the inputs and outputs costs of the general surgery service of CH Setúbal with one of its benchmark

complementary means of diagnosis may be a sign that a wrong mix of treatment paths is being followed by this service. That is, could it be using too many medical hours instead of complementing these with other diagnosis means? These are questions that HOBE can trigger and push to action service managers of this Hospital.

The higher costs of general surgery service of CH Setúbal service, cannot be explained by higher outputs than benchmark services, and therefore other reasons may be behind the exaggerated costs. Clearly one of such reasons can be quality. For example this service may have amongst the best doctors in the country and therefore shows an excessive cost with doctors. This is the reason why, quality variables are important and should, whenever available, be included in the analysis. In our case, a constant quality assumption is being made on comparisons between services and/or hospitals.

The graphical information above can be complemented with the actual values of the evaluated hospital and its benchmarks. This information is also shown in HOBE (but not displayed here for sake of brevity).

4.3 Outcomes on aggregate indicators for the hospital

The above efficiency assessment can be done for each service of the hospital individually. However if a general overview of all service's efficiency is required, one needs to use HOBE's functionality of aggregate performance for the hospital. In that case, a summary table appears showing the cost efficiency of each service of a given hospital (see Figure 6), and also the number of patients treated within each service, to give an idea of the dimension and importance of the service (this information was used for the establishment of weight constraints included in the DEA model, as explained in the Appendix).

This information gives immediately an overall picture of the services where cost savings can be achieved. For HS Setúbal the cardiology service clearly deserves a closer investigation as well as the pulmonology service or the anesthesiology service. These services show a potential for costs savings around 70 per cent, meaning that the outputs they generate do not justify the high costs observed.

Services	Number of patients treated:	Efficiency:
Anaesthesiology	14031	33.1 %
Cardiology	22997	24.7 %
Gastroenterology	15652	64.8 %
General Surgery	35541	72 %
Gynaecology/Obstetrics	40095	100 %
Haematology	2745	100 %
Imunohemotherapy	7541	45.1 %
Infectious Diseases	12123	69.4 %
Internal Medicine	43411	100 %
Medical Oncology	28054	73.7 %
Nephrology	14682	93.3 %
Ophthalmology	32858	54.6 %
Orthopaedics	51801	100 %
Paediatrics	56989	100 %
Psychiatry	13522	45.5 %
Pulmonology	9369	24.8 %
Urology	11895	36.7 %

Figure 6.
Performance of the services of an efficient hospital (CH Setúbal)

An aggregate score for the hospital as a whole, can be computed by aggregating the service-level efficiencies. In that respect, HOBE allows the user to choose the services to be considered in this computation. After this selection, HOBE returns the aggregate efficiency value of the hospital. The aggregate score can be 100 per cent, meaning that the hospital is a benchmark for the remainder hospitals, as the cost efficiency of each of its services is the best observed (taking into account the weight of the service in the overall hospital). That is not the case of CH Setúbal that shows an aggregate efficiency score of 95.64 per cent (note that this high score is the result of low-service scores being observed for services whose weight on the aggregate efficiency score is low). HOBE returns for this hospital de graph in Figure 7 that compares its service's performance with that of its main peer (CH do Porto, in this case).

In this case, the hospital shows for several services lower performance than its benchmark (e.g. urulogy, ophtalmology and general surgery services). Note, however, that this hospital excels in the services of internal medicine, gynaecology/obstetrics, pediatrics, orthopaedics, haematology, and nephrology with a performance of 100 per cent. Note that in some cases the benchmark may appear to have some service efficiencies equal to zero – this happens when the service did not exist for the hospital, or could not be evaluated.

5. Conclusion and research implications

The aim of this paper was to present the benchmarking platform HOBE – that allows a comparative analysis of hospital services based on a set of individual indicators, or based on aggregate indicators constructed using DEA. We argue, as other authors did before, that the best way to benchmark hospitals is by comparing services, rather than entire institutions. In any case, if hospitals wish to have an idea of their overall performance, it is possible to estimate an aggregate performance score based of the performance of their services. Following this idea, we also present within HOBE models that allow the aggregation of service's performance and a comparison between hospitals based on their services' efficiencies. This analysis can be very valuable to hospitals, as their weaknesses and strengths may become clear from this benchmarking exercise.

A tool like HOBE can be very important for hospital management as it allows service managers to understand the drivers of their service's performance, and it also

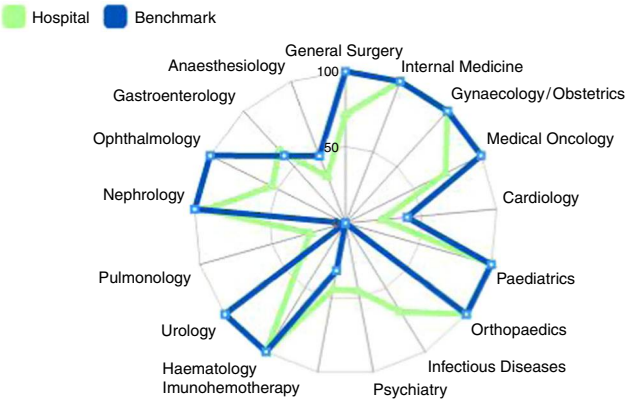

Figure 7. Aggregate performance of CH Setúbal

allows hospital managers to understand which services may be hindering/fostering the overall performance of the hospital. This tool has been presented to some hospital managers and the key performance indicators were developed in agreement with hospital managers and ACSS, which provided the data. Note that an entity like ACSS (central administration) has also interest on benchmarking exercises like that done through HOBE. In particular, ACSS benchmarks hospitals rather than services, and therefore the possibility of comparing service's performance is an added value to the current ACSS benchmarking platform. HOBE can and should be extended to include a more thorough analysis of quality of the health care provided, and that will be the subject of future research. This would improve the usefulness of the platform to the general public, and simultaneously it would allow the computation of performance measures accounting both or the cost and for the quality of the service provided.

Future work in HOBE will also involve uploading of more recent data into the system (such that it can indeed be used for managerial actions), and working in closer cooperation with cooperating hospitals to improve the indicators, and to nimble the process of importing data (which should ideally be performed by hospitals for their own interest and for the interest of their "clients").

References

Almeida, D.Q. (2013), "Hospital efficiency as an aggregate of services' efficiency: a DEA approach", master thesis, Faculdade de Engenharia da Universidade do Porto, Porto.

Barros, C.P., de Menezes, A.G. and Vieira, J.C. (2013), "Measurement of hospital efficiency, using a latent class stochastic frontier model", *Applied Economics*, Vol. 45 No. 1, pp. 47-54.

Barros, P.P. (2003), "Random output and hospital performance", *Health Care Management Science*, Vol. 6 No. 4, pp. 219-227.

Bogetoft, P. and Lars, O. (2011), *Benchmarking with DEA, SFA, and R*, Springer, Science & Business Media, New York, NY.

Bogetoft, P. and Nielsen, K. (2005), "Internet based benchmarking", *Group Decision and Negotiation*, Vol. 14 No. 3, pp. 195-215.

Camp, R.C. and Tweet, A.G. (1994), "Benchmarking applied to health care", *The Joint Commission Journal on Quality Improvement*, Vol. 20 No. 5, pp. 229-238.

Castro, R.A. (2011), "Benchmarking de Hospitais Portugueses – Modelação com Data Envelopment Analysis", Tese de Mestrado, Faculdade de Engenharia da Universidade do Porto, Porto.

Charnes, A., Cooper, W.W. and Rhodes, E. (1978), "Measuring the efficiency of decision making units", *European Journal of Operational Research*, Vol. 2 No. 6, pp. 429-444.

Chilingerian, J.A. and Sherman, H.D. (2004), "Health care applications: from hospitals to physicians, from productive efficiency to quality frontiers", in Cooper, W.W., Seiford, L.M. and Zhu, J. (Eds), *Handbook on Data Envelopment Analysis*, Kluwer Academic Publisher, Dordrecht, London and Boston, MA, pp. 445-493.

Chilingerian, J. and Sherman, H.D. (2011), "Health-care applications: from hospitals to physicians, from productive efficiency to quality frontiers", in Cooper, W.W., Seiford, L.M. and Zhu, J. (Eds), *Handbook on Data Envelopment Analysis*, Springer, Science & Business Media, New York, NY, Vol. 164, pp. 445-493.

Dervaux, B., Leleu, H., Minvielle, E., Valdmanis, V., Aegerter, P. and Guidet, B. (2009), "Performance of French intensive care units: a directional distance function approach at the patient level", *International Journal of Production Economics*, Vol. 120 No. 2, pp. 585-594.

Dismuke, C.E. and Sena, V. (2001), "Is there a trade-off between quality and productivity? The case of diagnostic technologies in Portugal", *Annals of Operations Research*, Vol. 107 No. 1, pp. 101-116.

Donabedian, A. (1988), "The quality of care: how can it be assessed?", *The Journal of the American Medical Association*, Vol. 260 No. 12, pp. 1743-1748.

Donabedian, A. (2005), "Evaluating the quality of medical care", *The Milbank Quarterly*, Vol. 83 No. 4, pp. 691-729.

ERS (2016), "SINAS – Sistema Nacional de Avaliação Em Saúde", available at: www.ers.pt/pages/118 (accessed 28 February 2016).

Gonçalves, L.F. (2008), "Análise da Eficiência dos Hospitais SA e SPA Segundo uma Abordagem de Fronteira de Eficiência", PhD, Instituto Superior de Ciências do Trabalho e da Empresa, Lisboa.

Grönroos, C. and Ojasalo, K. (2004), "Service productivity towards a conceptualization of the transformation of inputs into economic results in services", *Journal of Business Research*, Vol. 57 No. 4, pp. 414-423.

Grosskopf, S., Margaritis, D. and Valdmanis, V. (2001), "Comparing teaching and nonteaching hospitals: a frontier approach (teaching vs non-teaching hospitals)", *Health Care Management Science*, Vol. 4 No. 4, pp. 83-90.

Gummesson, E. (1998), "Productivity, quality and relationship marketing in service operations", *International Journal of Contemporary Hospitality Management*, Vol. 10 No. 1, pp. 4-15.

Hollingsworth, B. (2008), "The measurement of efficiency and productivity of health care delivery", *Health Economics*, Vol. 17 No. 10, pp. 1107-1128.

Hougaard, J.L. and Tvede, M. (2002), "Benchmark selection: an axiomatic approach", *European Journal of Operational Research Society*, Vol. 137 No. 1, pp. 218-228.

Johnson, A.L. and McGinnis, L.F. (2011), "Performance measurement in the warehousing industry", *IIE Transactions*, Vol. 43 No. 3, pp. 203-215.

Kawaguchi, H., Tone, K. and Tsutsui, M. (2013), "Estimation of the efficiency of Japanese hospitals using a dynamic and network data envelopment analysis model", *Health Care Management Science*, Vol. 17 No. 2, pp. 101-112.

Kontodimopoulos, N. and Niakas, D. (2005), "Efficiency measurement of hemodialysis units in Greece with data envelopment analysis", *Health Policy*, Vol. 71 No. 2, pp. 195-204.

Kristensen, T., Laudicella, M., Ejersted, C. and Street, A. (2010), "Health economics cost variation in diabetes care delivered in English hospitals", *Diabetic Medicine*, Vol. 27, pp. 949-957.

Laudicella, M., Olsen, K.R. and Street, A. (2010), "Examining cost variation across hospital departments: a two-stage multi-level approach using patient-level data", *Social Science & Medicine*, Vol. 71, pp. 1872-1881.

Menezes, A.G.R., Marco, F. and Vieira, J.C. (2006), "Eficiência técnica dos hospitais portugueses 1997-2004: uma análise (regional) com base num modelo de fronteira estocástica", CEEAplA Working Paper No. 08/2006, Açores.

Morais, P. and Camanho, A.S. (2011), "Evaluation of performance of European cities with the aim to promote quality of life improvements", *Omega*, Vol. 4 No. 39, pp. 398-409.

Moreira, S. (2008), "Análise dos Hospitais-Empresa: Uma Aplicação da Data Envelopment Analysis", Boletim Económico, Spring, (Primavera).

Mosel, D. and Gift, B. (1994), "Collaborative benchmarking in health care", *The Joint Commission Journal on Quality Improvement*, Vol. 20 No. 5, pp. 239-249.

Navarro-Espigares, J.L. and Torres, E.H. (2011), "Efficiency and quality in health services: a crucial link", *The Service Industries Journal*, Vol. 31 No. 3, pp. 385-403.

Nayar, P., Ozcan, Y., Yu, F. and Nguyen, A.T. (2013), "Benchmarking urban acute care hospitals: efficiency and quality perspectives", *Health care Management Review*, Vol. 38 No. 2, pp. 137-145.

OECD (2008), *Handbook on Constructing Composite Indicators: Methodology and User Guide*, Euroepan Comission, OECD.

Ozcan, Y. (2008), *Health Care Benchmarking and Performance Evaluation: An Assessment Using Data Envelopment Analysis*, Springer, New York, NY.

Portela, M.C.A.S. (2014), "Value and quantity data in economic and technical efficiency measurement", *Economics Letters*, Vol. 124 No. 1, pp. 108-112.

Portela, M.C.A.S., Camanho, A.S. and Borges, D. (2011), "BESP – benchmarking of Portuguese secondary schools", *Benchmarking: An International Journal*, Vol. 18 No. 2, pp. 240-260.

Puig-Junoy, J. (1998), "Technical efficiency in the clinical management of critically ill patients", *Health Economics*, Vol. 7, pp. 263-277.

Rego, G., Nunes, R. and Costa, J. (2010), "The challenge of corporatisation: the experience of Portuguese public hospitals", *European Journal of Health Economics*, Vol. 11 No. 4, pp. 367-381.

Retzlaff-Roberts, D., Chang, C.F. and Rubin, R.M. (2004), "Technical efficiency in the use of health care resources: a comparison of OECD countries", *Social Science & Medicine*, Vol. 69 No. 1, pp. 55-72.

Sikka, V., Luke, R.D. and Ozcan, Y. (2009), "The efficiency of hospital based clusters: evaluating system performance using data envelopment analysis", *Health Care Management Review*, Vol. 34 No. 3, pp. 251-261.

Simões, P. and Marques, R.C. (2011), "Performance and congestion analysis of the Portuguese hospital services", *Central European Journal of Operations Research*, Vol. 19 No. 1, pp. 39-63.

Stevanovic, V., Feek, C. and Kay, R. (2005), "Using routine data for benchmarking and performance measurement of public hospitals in New Zealand", *Benchmarking: An international Journal*, Vol. 12 No. 6, pp. 498-507.

Voss, C.A., Ahlstrom, P. and Blackmon, K. (1997), "Benchmarking and operational performance: some empirical results", *Benchmarking for Quality, Management and Technology*, Vol. 4 No. 4, pp. 273-285.

Appendix

Consider a set of j ($j = 1, \ldots, n$) of hospitals using an aggregate cost C_j to produce a set of r ($r = 1, \ldots, s$) outputs (y_{rj}). The DEA cost model for evaluating a given service of hospital o is shown in (1) (see e.g. Portela (2014) for a discussion on the use of aggregate costs to measure cost efficiency):

$$Min\left\{\theta \left| \sum_{j=1}^{n} \lambda_j C_j \leqslant \theta C_o, \sum_{j=1}^{n} \lambda_j y_{rj} \leqslant y_{ro}, \quad r = 1, \ldots, n, \ \lambda_j \geqslant 0 \right.\right\} \tag{1}$$

θ represents the efficiency of the service being assessed and λ represents the intensity associated to decision making unit (DMU) j, that is the contribution of DMU j for the benchmark set of unit o. This model has the disadvantage that it does not allow for the assessment of the technical efficiency or the allocative efficiency. This is compensated by the fact that it only requires the total cost of each resource, requiring neither the individual price nor the quantity of the resource.

Resulting from Model (1) there is an optimal efficiency value for each service k, ($k = 1, .., K$) of hospital o. The DEA model used for evaluating hospitals is a composite indicator (see OECD, 2008 or Morais and Camanho, 2011), which aggregates all efficiencies of the hospital's services subject to certain constraints on the weights assigned to each service. These constraints impose that the weights assigned to each service in relation to a base service (say Service 1) should be at most 20 per cent lower and at least 20 per cent higher than the relationship between the number of patients handled by each service (represented by N_k). This model is represented in (2) where the decision variables are the weights (u_k) assigned to each service's efficiency of hospital o:

$$Max\left\{\sum_{k=1}^{K} u_k \theta_{ko}^{*} \,|\, \sum_{k=1}^{K} u_k \theta_{kj}^{*} \leqslant 1, \quad j = 1, \ ..., \ n, \ 0.8\frac{N_k}{N_1} \leqslant \frac{u_k}{u_1} \leqslant 1.2\frac{N_k}{N_1}\right\} \tag{2}$$

Corresponding author
Maria Conceição A. Silva Portela can be contacted at: csilva@porto.ucp.pt