Jonathan Cylus, Irene Papanicolas and Peter C. Smith

Using data envelopment analysis to address the challenges of comparing health system efficiency

Article (Accepted version) (Refereed)

Original citation:
Cylus, Jonathan, Papanicolas, Irene and Smith, Peter S. (2015) *Using data envelopment analysis to address the challenges of comparing health system efficiency.* Global Policy. ISSN 1758-5899

DOI: 10.1111/1758-5899.12212

Using Data Envelopment Analysis to address (some of) the challenges of comparing health system efficiency

Abstract

Efficiency is one of the most potent measures of health system performance and is of particular interest to policymakers because it seeks to assess the valued outcomes of a health system in relation to the resources that are sacrificed to achieve those outcomes. However, the production process of the health care system is a complex sequence, and most indicators are only able to capture part of that process; these indicators offer limited scope for analysis. While researchers have previously constructed composite indicators which combine partial measures into a single number, the weights used for aggregating data can be contentious and may not be universal across systems. In this article, we propose a method that uses data envelopment analysis (DEA) to construct composite health system efficiency indicators from several partial efficiency measures. DEA is most often used to compare the productivity of different producing entities, including health systems, but is not typically used in this manner. Among other noted benefits, it allows construction of composite indicators where different weights are attached to partial indicators for each country, allowing countries to be viewed according to the weights that cast each in the best light. Our application of this method suggests that there is reasonable consistency among the countries that are found to be efficient.

Key words: efficiency, performance, data envelopment analysis, composite indicators

I. Introduction

Rising health care costs and increasing concerns about fiscal sustainability have brought the issue of health system efficiency to the forefront of policy discussions. Most high income countries are trying to identify ways in which they can secure the same health outcomes for less, while many middle and low income countries are mindful of ensuring their health systems expand while providing value for money. Comparative efficiency indicators offer policy makers an important resource in their search for efficiency improvements. These types of indicators can not only be used to identify areas in the health system which may not be performing as well as they should be but also provide an indication of the countries to look towards to identify potential processes that may improve the value for money of the system (Papanicolas and Smith 2013).

While the concept of health system efficiency is deceptively simple - maximizing valued health system outputs relative to inputs - it becomes more difficult to make operational when applied to a concrete situation, particularly at the system level. Among the challenges in measuring health system efficiency are defining, and measuring, the valued system outputs and inputs that differ across institutions. In practice, different definitions in use cover a range of valued outputs such as 'overall' performance, quality of care, health gain, or volume of treatment. Thus, efficiency indicators essentially serve as a summary measure of the extent to which the inputs to the health system, in the form of expenditures and other resources, are used to secure these goals of the health system. Yet, the limitation of available metrics to measure the valued outputs and inputs at both national and international levels further restrain efforts to adequately represent the true efficiency of the system.

The challenge of identifying a set of valued outputs has implications for the conceptualization of both technical and allocative efficiency, both of which are important to policy makers. Technical efficiency examines the extent to which the unit is failing to reach the maximum level of health system output that *can* be produced for different levels of inputs (otherwise known as the production frontier).

Allocative efficiency relates to whether production is distributed across outputs to maximize the value to society. In order to determine the technically efficient points of production, it is necessary to identify the outputs of the production process and the maximum attainable outputs given existing inputs. Similarly, in order to determine what bundle of health services to provide, and thus identify the 'allocatively efficient' point of production, it is necessary to understand the preferences of the population being served. This will require consideration of whose preferences to consider, and how to trade-off one output for another.

Although it is one of the most fundamental health system performance metrics for researchers and policymakers (WHO 2000), the concept of health system efficiency is in practice heavily contested and its accurate measurement across countries difficult to realize (Reinhardt, Hussey et al. 2002). Few attempts have been made to create single comparative measures of health system efficiency, such as the World Health Report 2000 (WHO, 2000), but all face common challenges of dealing with lack of a clear conceptual framework, limited data, and selecting reliable and appropriate empirical techniques.

While frontier analysis – assessing country production relative to the best performers-- is generally agreed to be to the most reliable empirical approach to efficiency assessment, there is no consensus as to which theoretical or statistical criteria to use to select between existing empirical techniques (Street and Hakkinen 2009). The main approaches include both nonparametric data envelopment analysis (DEA) and parametric stochastic frontier analysis (SFA). The main differences between the DEA and SFA methods are in their approach to establishing the location and shape of the production frontier that all other systems are compared to, and estimating each system's position relative to that frontier. In practice, most empirical research work on the estimation of system efficiency uses one or both of these methods which have different strengths and weaknesses and thus make one or the other better suited to particular cases.

This article proposes an alternative application of the non-parametric DEA methodology to create composite measures of efficiency indicators. The remainder of the paper is organized as follows. In section 2 we describe the limitations of existing methodologies for efficiency measurement and outline the motivation for our approach. In section 3 we propose our methodology by means of an example using available Organization for Economic Cooperation and Development (OECD) data. In section 4 we present the results of this case study, followed by a brief discussion in section 5. To conclude, we recommend how and why this method should be used in future efficiency analyses.

II. Benefits and Limitations of Composite Measures of Efficiency

The development of metrics that are able to compare health system efficiency has been on the agenda of researchers and policymakers for some time (WHO 2000; Hollingsworth and Wildman 2003) with most researchers relying on parametric (e.g. SFA) or non-parametric (e.g. DEA) analytic methods to facilitate assessment. Indicators designed to capture the efficiency of the entire production process—health costs relative to health outcomes—appear to be superficially most desirable despite the methodological challenges in their estimation (Jacobs 2006). In contrast, partial process measures, which capture only part of the transition from costs to health outcomes, are in practice easier to create and more useful to policymakers because they are better positioned to identify sources of inefficiency (Cylus and Smith 2013). As a result, most efficiency indicators are less ambitious and only measure the success by which discrete stages in the health care production process occur. Examples of common indicators include unit costs, average length of inpatient stay, or labor hours per episode of care.

However a fundamental concern when using partial process measures of efficiency is the limited scope they offer for analysis. No single partial indicator can accurately or effectively capture the relative efficiency of an entire system, since by definition partial indicators only assess the efficiency of a small portion of the health care production system. Apparent inefficiencies in one part of the system may be the result of constraints or opportunities elsewhere – for example, a low length of stay after stroke admission may be secured only at the expense of expensive investment in rehabilitation facilities elsewhere in the system, or indeed poor eventual health outcomes.

Researchers relying on narrow indicators to measure whether inputs are successfully converted into outputs for only a segment of health care production are therefore usually unable to draw broader conclusions about the system at large. One way of overcoming this – at least in part - may be to examine a composite of partial indicators of efficiency that explore a spectrum of stages and processes of the health system. The intention is that by assessing a set of indicators simultaneously, rather than piecemeal, a more secure picture of system efficiency can be secured.

Composite indicators have become increasingly used by researchers, including for World Health Report 2000 (WHO 2000), and the UN's Human Development Index (Sagar and Najam 1998). These indicators use relative weights to aggregate data series and can be more appealing to researchers and policymakers than assessing multiple measures separately because they are useful for conveying summary performance information (Smith 2002). However, in the construction of composite indicators there are numerous unresolved technical concerns related to weighting, using comparable units of measure, and ultimately aggregating (Esty, Levy et al. 2006). Generally, weights are based on relative preferences for whatever is being measured, often elicited through surveys (Gakidou 2000) or inferred based on some other measure such as the share of total expenditure for a type of service. However, asking survey respondents how important a certain objective or indicator is to them will often reveal large inter-personal variations and will not necessarily produce marginal trade-off valuations that can be used as weights (Munda and Nardo 2005). The weights used for aggregating data can therefore be contentious, arbitrary, and may not be universal across systems.

While there is a wide range of weighting techniques available (OECD/European Union/Joint Research Centre- European Commission 2005) and although economic methodology exists for determining relative valuations, there is no universally accepted methodology for deriving appropriate weights. Certainly none has been consistently incorporated into the development of health system performance composite indicators (Dolan, Gudex et al. 1996; Jacobs, Smith et al. 2004; Decancq and Lugo 2010). There is therefore no commonly used methodology for comparing countries' ability to perform well simultaneously on multiple indicators of health system efficiency that circumvents the weighting issue.

In response, we propose a method that uses DEA to construct composite indicators from several partial efficiency measures. DEA is one of the most widely used analytical techniques for assessing comparative efficiency (Farrell 1957; Charnes, Cooper et al. 1978). It uses linear programming methods to determine how well a producer, often referred to in the literature as a decision making unit (DMU) converts a set of inputs into a set of outputs. DEA works by using data from a selection of DMUs to create an efficiency frontier based on various measures of inputs and outputs, with DMUs assessed in relation to that frontier; DEA offers a conservative assessment of performance. The methodology chooses a different set of input and output weights for each DMU so as to maximize the DMU's apparent efficiency (Thanassoulis 2001). Unlike regression methods for estimating efficiency frontiers, DEA requires no restrictive assumptions and none of the stringent model testing that is required of statistical techniques. However, it can be vulnerable to data errors, because the DEA 'best practice' frontier is composed of a small number of highly performing organizations, and the performance of all other units is judged in relation to that frontier. For a more detailed explanation of this method see Jacobs et al. (2006).

The flexibility in the valuation weights attached to each input or output allowed by DEA has exposed the technique to fierce criticism. It may be perceived by some as a drawback from the point of view of ranking organizations, since ranking would require that all countries use the same set of weights (Stone 2002). However, we are interested not so much in ranking for the purpose of creating league tables, but in creating composite measures of efficiency that inform where each country stands relative to a set of 'best practice' peers. Indeed, using the concept of 'cross efficiency', one can also construct for each DMU a unique ranking of other DMUs using the weights selected by DEA for the DMU under scrutiny (Doyle and Green 1994). Therefore, for our purposes the weight flexibility is an important positive feature of DEA. Each organization can in principle be compared to the frontier according to an entirely different set of output weights, also allowing policy makers the option to weigh things according to societal or political preferences and to take into account allocative efficiency considerations. DEA also does not require that partial measures be normalized into a common scale prior to aggregation because results are independent of the units of measurement.

Using DEA to create composite indicators has previously been referred to as the "benefit of the doubt" approach (Melyn and Moesen 1991; Cherchye, Moesen et al. 2007). For example, Cherchye et al have used this technique in the context of creating the Technology Achievement Index (Cherchye, Moesen et al. 2008). Zhou et al outline the construction of composite indicators using DEA and apply their approach to modeling the development of sustainable energy in 18 countries (Zhou, Ang et al. 2007). However, there seem to be no existing DEA composite indicators that are comprised of partial comparative efficiency indicators for the health sector.

The intention of this methodology is to complement, not replace individual indicators, in the hope of offering a broader perspective on health system efficiency. Most importantly for our purposes, it allows differential weights to be attached to indicators for each country, allowing countries to be viewed according to the weights that cast each in the best light. By combining several efficiency indicators into a single measure, countries can see if there is evidence of system-wide efficiency effects, how they compare to other countries, which countries are the most efficient peers, and what areas are the priorities in order to improve their ranking.

III. Methodology

A variety of forms of the DEA model exist. As required when the variables used are in the form of ratios, this paper adopts the output-oriented representation presented by Banker, Charnes and Cooper (Hollingsworth and Smith 2003). The output orientation model indicates the extent to which a country could proportionately increase each of the partial efficiency indicators. No measure of inputs is required for our purposes (i.e. we assume that inputs are equal to 1). A mathematical illustration of the DEA model can be found in Box 1.

Box 1. Generic output-oriented DEA model

In algebraic terms, given *n* outputs and *m* inputs:

$$E_0 = \frac{\sum_{r=1}^{n} u_r \cdot y_{r_0}}{\sum_{i=1}^{m} v_i \cdot x_{i_0}}$$

where y_{r_0}= quantity of output *r* produced by unit 0; u_r=weight attached to output *r*; x_{i_0}=quantity of input *i*; v_i= weight attached to input *i*. For the models in this paper, the denominator is set equal to 1. E_0 is an indicator of efficiency where $1 \geq E_0 \geq 0$.

The model in linear programming terms is:

For unit 0 in a sample of *n* units,

Maximize: $h_0 = \sum_{r=1}^{s} u_r \cdot y_{r_0}$

Subject to: $\sum_{i=1}^{m} v_i \cdot x_{i_0} = 1$

$$\sum_{r=1}^{s} u_r \cdot y_{r_0} - \sum_{i=1}^{m} v_i \cdot x_{i_0} \leq 0 \quad j = 1, \dots, n$$

$$u_r \geq \varepsilon,\ r = 1, \dots, s$$

$$v_i \geq \varepsilon,\ i = 1, \dots, m$$

where ε is an infinitely small constant which restricts the weights to positive values.

To begin, we use OECD Health data to select simple efficiency measures to use as outputs in the DEA (OECD 2010). The OECD data do not contain many ready-made efficiency indicators, so some must be constructed manually using the available data. Since all efficiency measures by definition must represent some type of output to input ratio, each partial efficiency indicator is created by dividing some measure of output by some measure of input. Because DEA is based on the presupposition that greater output per input indicates greater efficiency, all output to input ratios must also be designed in such a fashion. Therefore, for any measure where a lower score indicates higher efficiency (e.g. average length of stay) the inverse is used. The use of such indicators implicitly assumes constant returns to scale because we ignore the size of countries.

After reviewing the data from 2005 through 2008 and constructing several indicators, we choose the following for a selection of countries because of their comparatively high frequency of availability:

a) Consultations per physician
b) Acute care occupancy rate
c) Inverse of average length of stay (ALOS)
d) Inverse of health spending share of gross domestic product (GDP)

For the most part, these indicators are self-explanatory.[1]

We acknowledge that the chosen indicators do not effectively cover the full-spectrum of the health care production process from health care costs to health outcomes; however, they are what could be created for a reasonably large set of countries using the data. We also recognize that they are highly imperfect indicators of partial efficiency, and that ideally an analysis would adjust for numerous exogenous influences on attainment. Conventional DEA would seek to model efficiency using all inputs (including environmental inputs) and all outputs. However, the purpose of this study is precisely to examine whether there is more information that can be gleaned from admittedly limited datasets when such comprehensive data are not available. Instead we treat the chosen indicators as a series of (partial)

[1] Consultations per physician is calculated manually by dividing the number of consultations per population by the number of physicians per population and is an indication of physician productivity. The acute care occupancy rate is the number of utilized bed-days divided by the number of available bed-days over the course of a year in acute care hospitals and is provided directly by the OECD Health Data. It is an indicator of the productive use of capital (hospitals). Average length of stay in acute care hospitals is the total number of occupied acute care hospital bed-days divided by the total number of acute care admissions and is also provided by the OECD; the inverse is used because a higher ALOS implies lower efficiency. It is a proxy for the costs per episode of hospital care. In principle, both average length of stay and acute care occupancy rate should be based on the same case-mix in each country, though casemix adjustment was not feasible in this application. Health spending share of GDP is total health care expenditure as a percentage of GDP; the data is presented in the OECD Health Data although again, the inverse is used for this analysis. This is an indication of the macroeconomic efficiency of the health system.

outputs in lieu of using each individual indicator's respective inputs and outputs. To maintain a consistent set of countries, we limit the analysis to the following countries with a full set of data: Austria, Belgium, Czech Republic, Estonia, Germany, Hungary, Luxembourg, Slovak Republic, Slovenia, Spain, and United Kingdom.

We create four DEA models using combinations of three partial efficiency indicators as outputs and assume uniform inputs (i.e. inputs equal to 1). We also run one DEA model using all four indicators, and experiment with a second four-indicator DEA model with a restriction on the weights; weight restrictions allow us to set rules regarding weights, if for instance, we wanted to ensure that all outputs were weighted greater than zero. Using different sets of partial efficiency indicators as outputs allows us to see if there is any consistency in the countries that score well. The intention is to determine whether any health systems are systematically performing better at multiple stages of health production, or whether the efficiency of different health systems is largely dependent on the efficiency indicators that are selected.

IV. Results

In each of the models containing three efficiency indicators as outputs, different combinations of four countries form the efficiency frontier; Hungary is the only country that is found to be efficient in all model specifications. Using all four partial efficiency measures as outputs, five countries—Estonia, Hungary, Slovak Republic, Slovenia, and the United Kingdom– form the efficiency frontier.

For more detailed analysis and to determine consistency, we compare one of the three indicator DEA models (referred to as 'three variable model') with the DEA containing all four indicators (referred to as 'four variable model') as an example. In the three variable model, consisting of consultations per physician, health share of GDP, and average length of stay as the selected indicators, the countries forming the efficiency frontier are Estonia, Hungary, Slovak Republic and Slovenia (Table 1). The inclusion of occupancy rate as an additional output for the four variable model causes the United Kingdom to be added to this set. The mean efficiency in the three variable model is 90.54%; in the four indicator model, mean efficiency is 97.69%. Note that the efficiency of inefficient DMUs can never decrease when additional outputs are added.

Table 1. Efficiency Scores		
	Three variable model	**Four variable model**
Austria	83.82%	98.75%
Belgium	80.28%	92.20%
Czech Republic	98.05%	98.05%
Estonia	100%*	100%*
Germany	75.69%	91.38%
Hungary	100%*	100%*
Luxembourg	88.77%	95.81%
Slovak Republic	100%*	100%*
Slovenia	100%*	100%*
Spain	87.69%	98.39%
United Kingdom	81.61%	100%*

*Forms efficiency frontier

Table 2 shows the 'peers' used to construct the frontier for each inefficient system, and the weights attached to each peer. In the three output specification, Hungary is a peer 4 times, Estonia 7 times, and Slovenia 5 times; in the four output specification, Hungary is a peer 6 times, the United Kingdom 5 times, and Estonia 2 times (Table 2).

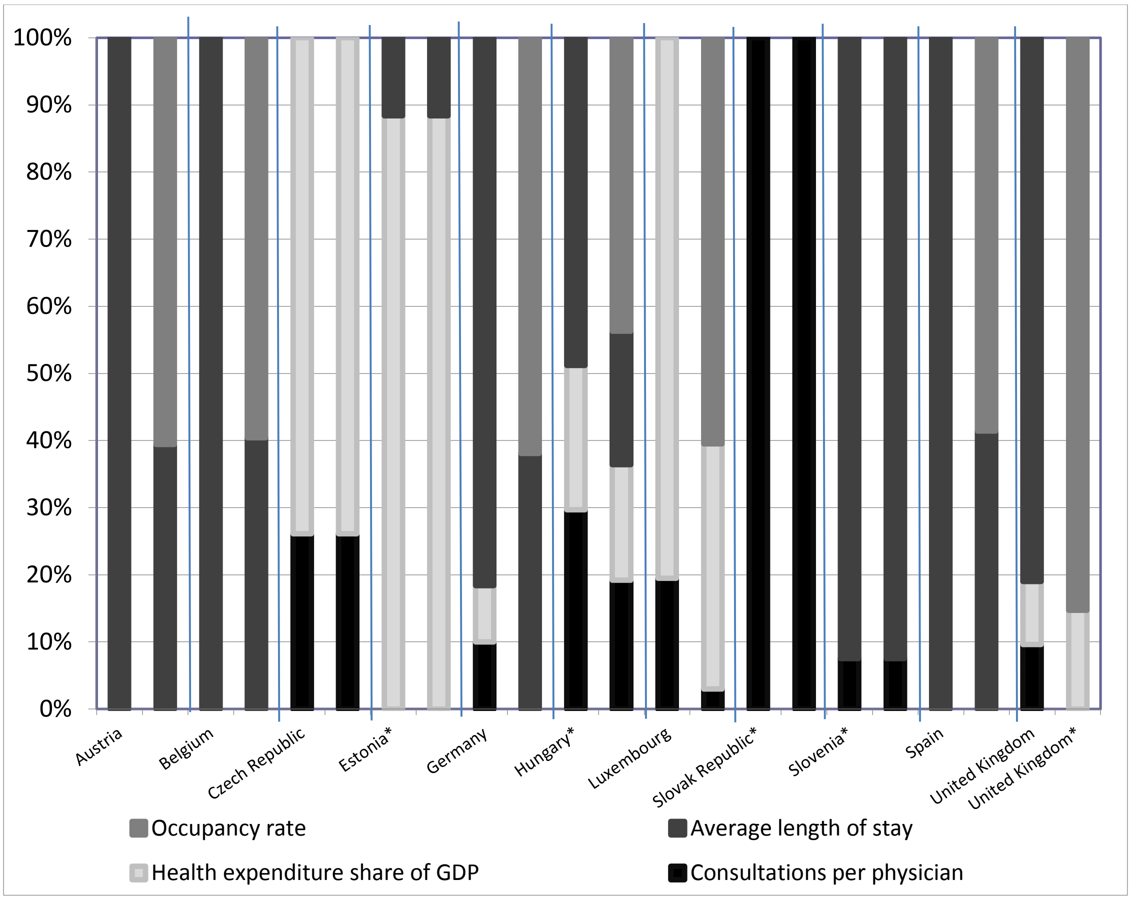
Table 2. Country benchmarks		
	Three variable model	**Four variable model**
Austria	Estonia (57.0%) Slovenia (43.0%)	Hungary (31.3%) UK (68.7%)
Belgium	Estonia (39.6%) Slovenia (60.4%)	Hungary (46.1%) UK (53.9%)
Czech Republic	Hungary (75.3%) Estonia (24.7%)	Hungary (75.3%) Estonia (24.7%)
Germany	Hungary (18.3%) Estonia (5.3%) Slovenia (76.4%)	Hungary (12.1%) UK (87.9%)
Hungary	Forms frontier for 4 countries	Forms frontier for 6 countries
Luxembourg	Hungary (27.7%) Estonia (72.3%)	Hungary (11.2%) UK (32.6%) Estonia (56.2%)
Slovak Republic	Forms frontier for no countries	Forms frontier for no countries
Spain	Estonia (22.4%) Slovenia (77.6%)	Hungary (60.1%) UK (49.9%)

United Kingdom	Hungary (32.5%) Estonia (34.5%) Slovenia (33.0%)	Forms frontier for 5 countries
Estonia	Forms frontier for 7 countries	Forms frontier for 2 countries
Slovenia	Forms frontier for 5 countries	Forms frontier for no countries

Note: For countries not on the efficiency frontier, the numbers in parenthesis correspond to the weight of the respective peer country that forms the frontier.

Figure 1 summarizes the weights on the outputs used to construct the frontier (note that weights need not be unique, so this analysis should be treated with some caution). The addition of occupancy rate causes changes in many of the output weights for Austria, Belgium, Germany, Hungary, Luxembourg, Spain, and the United Kingdom (Figure 1). The change in weights is greatest for the average length of stay variable, which drops an average of 30.2% across all countries when moving from three indicators to four. The weights for Czech Republic, Estonia, Slovenia, and the Slovak Republic do not change when adding the occupancy rate variable, implying that achieving high occupancy rates has no relative importance to those systems when considered among this set of objectives. They rely for their high score on the other objectives. In particular, Slovak Republic weights consultations per physician at 100%, which suggests that it secures its good score solely because of its uniquely good performance on that indicator.

Figure 1. Output variable weights



Note: For each country, the bar to the left represents the output weights for the three variable model; the bar to the right represents the output weights for the four variable model. *Forms efficiency frontier

Countries with large slacks suggest that they have large potential improvements in those specific dimensions, in addition to the improvements suggested by the efficiency score. In both models, Austria is found to have the most potential additional improvement in the number of consultations per physician and the health share of GDP (Table 3). Czech Republic has the largest slack for average length of stay and occupancy rate.

Table 3. Slacks for countries not on the frontier

		Consultations per physician	**Health expenditure share of GDP inverse**	**Average length of stay inverse**	**Occupancy rate**
Three variable model:	Austria	511.99	0.032	0.000	N/A
	Belgium	100.42	0.016	0.000	N/A
	Czech Republic	0.00	0.000	0.031	N/A
	Germany	0.00	0.000	0.000	N/A

	Luxembourg	0.00	0.000	0.019	N/A
	Spain	47.85	0.003	0.000	N/A
	United Kingdom	0.00	0.000	0.000	N/A
Four variable model:	Austria	1180.00	0.025	0.000	0.00
	Belgium	861.48	0.019	0.000	0.00
	Czech Republic	0.00	0.000	0.031	2.83
	Germany	34.66	0.013	0.000	0.00
	Luxembourg	0.00	0.000	0.020	0.00
	Spain	831.47	0.015	0.000	0.00

Finally, we re-ran the four variable model with one weight restriction to test whether forcing the weight on consultations to be greater than or equal to the weight on health share of GDP would affect the results. This approach was selected because in the original four indicator model, consultations per physician were weighted less than health share of GDP for 4 countries. Again, we found similar countries to be efficient (Hungary, Slovak Republic, Slovenia, and UK); Estonia was also almost on the frontier. UK and Hungary each were peers 6 times, while the Slovenia was a peer only once.

V. Discussion

It has proved challenging to develop robust measures of comparative efficiency that are feasible to collect or estimate, that offer consistent insight into comparative health system performance, and that can be usable in guiding policy reforms (Hussey, de Vries et al. 2009). The example given above demonstrates that DEA can be used to create composite indicators consisting of partial efficiency measures and produces consistent results even when using different combinations of indicators. One major benefit of DEA is that the composite weights for each indicator by definition are endogenously determined to reveal the maximum overall efficiency for each country and thus are not subject to specific normative preferences, which is otherwise a concern when constructing composite indicators. Indicators also do not have to be transformed into common units prior to weighting. The method is transparent because the weights are clearly presented and can be imposed or restricted if necessary.

It is also possible to assess cross-efficiency, which attaches the weights determined by DEA for one country to the partial efficiency indicators of a comparator country; applying the same weights to all countries would allow for constructing of league tables. Another benefit is that the DEA approach not only provides an overall efficiency score for each country, but also reports the potential for each country to improve on inefficient dimensions.

The method is quite feasible and easy to implement. However it suffers from some drawbacks. To begin, the lack of data availability prevents large-scale implementation of this sort of approach. Few if any partial-efficiency indicators are readily available for most countries, which limits the number of countries, domains of care, and production processes that can be compared. Using the OECD data, only a handful of partial efficiency indicators could be selected or manually constructed; keeping the set of countries constant across multiple model specifications forced us to discard even more country observations as well, despite the fact that we pooled data from four years. This highlights a general problem in comparing the efficiency of health systems, and highlights the urgent need to develop a broader range of cross-country comparable partial indicators of efficiency.

Assuming more plentiful data were made available, it is unclear which objectives should be measured and included in a DEA specification. The selection of partial efficiency indicators is not inconsequential, as weights for each indicator are entirely dependent on the other indicators that have been included in a model. This issue extends to the selection of variables that are collinear, which may present a problem as the calculated weights may not appropriately reflect trade-offs since it may not make sense to trade-off between such variables.

However, there is surprisingly little written on how to develop a satisfactory DEA model and there is no universally accepted modelling strategy. One can always debate whether a chosen model specification is most appropriate. DEA offers no tests to assist in choosing a preferred model and there has been little effort to examine the impact of misspecification on model results. Instead, the emphasis has been on exploring the sensitivity of results to data errors (Smith 1997). Many variants of the models used in this paper can be envisaged and tested which may include additional weight restrictions or add otherwise omitted indicators. For example, the additional constraint of the weight on consultations per physician being greater than the health share of GDP in the four indicator model was found to cause Estonia to become inefficient. Nevertheless, since combinations of the same group of 5 countries were found to be efficient in all models, there does appear to be some consistency amongst models using this particular set of indicators.

Additionally, since the indicators cannot account for country size, it may be inappropriate to make some comparisons. Similarly, there are likely other external influences on measured efficiency that are unaccounted for. Countless additional uncontrollable factors could in principle be introduced into a model, however each additional factor reduces the ability to discriminate between countries, and increases the number deemed efficient. With so few countries in our dataset, this presents a significant problem. Other general concerns relating to DEA are also important, such as its vulnerability to outliers, and the consequent need to examine robustness to the exclusion of such DMUs.

VI. Conclusion

The use of DEA offers an opportunity to create composite indicators made up of partial efficiency measures, and should be explored further to compare health care system efficiency. Given the current paucity of comparative data, examination of individual partial efficiency indicators does not permit broader conclusions regarding system-wide efficiency effects. We therefore feel that the standard piecemeal analysis can be usefully complemented with (but not replaced by) a DEA composite approach. Unlike traditional composite indicators, the DEA method reveals the elements that are responsible for each country's ranking, allowing better scrutiny of the reasons why a country may secure a favourable or unfavourable ranking. The intention is that this more explicit approach will help policymakers to direct reforms at the aspects of efficiency that need the most attention. From a technical perspective, the method also allows researchers to be detached from the selection of indicator weights, an otherwise contentious activity.

The analysis presented in this paper shows the extent to which countries are performing efficiently across four dimensions of the health care production process. The results must be interpreted with some caution, given the scope for endless debate about the precise specification of a "correct" model. However, the results do suggest consistency among the countries that are found to be efficient, since similar sets of countries form the efficiency frontiers in all model specifications. Wide variations in how countries weight different indicators are evident, with many countries attaching zero weights to some objectives in all models. It is likely that these countries in reality do value these objectives, and therefore it may make sense to restrict models to have only non-zero weights. In the full four-indicator DEA, Hungary was the only country where non-zero weights were found across all dimensions.

Despite some barriers to implementation, the approach offers improvements from a scientific and policy perspective because it identifies the areas that prevent a country from being considered efficient. The DEA approach facilitates detailed scrutiny of individual countries' performance, as in the examples for Austria and Czech Republic, both of which were found to have significant potential improvements in certain dimensions. This type of information can be helpful for decision-makers wishing to understand where the major scope for improvement lies in their country, and also what the relevant 'best practice' peers might be. Countries found to be inefficient may choose to examine why certain countries outperform them and explore whether the peers have adopted policies that are worthy of scrutiny. The method provides policymakers and other decision makers with a tool to squeeze further information out of existing data sources and to focus on areas of performance that demand further attention. Moreover, it allows comparative efficiency analyses to be estimated for Low and Middle-Income countries that have comparatively scaled down data collection efforts. To expand the use of this type of indicator in the future, the first step is to increase the number of partial efficiency indicators made available for more countries. There is clearly a role for organizations such as the European Commission, the OECD and WHO to specify, collate and report more such indicators, and to encourage countries to harmonize their efforts.

Composite indicators offer some important benefits, for example offering a rounded assessment of system efficiency and promoting accountability for the whole health system. Yet they are also often criticized for their lack of transparency in weighting and in identifying sources of inefficiencies. The use of DEA to create composites of partial efficiency indicators helps to address some of these issues. This paper highlights the value of DEA as an exploratory tool rather than offering a definitive judgement on health system performance; if used for that purpose it can offer useful diagnostic information on a country's general level of efficiency and on its performance in specific areas of activity.

Conflicts of Interest: We declare no conflicts of interest.

References

Charnes, A., W. W. Cooper, et al. (1978). "Measuring the efficiency of decision-making units." European Journal of Operational Research **2**: 429-444.

Cherchye, L., W. Moesen, et al. (2007). "An Introduction to 'Benefit of the Doubt' Composite Indicators." Social Indicators Research **82**(1): 111-145.

Cherchye, L., W. Moesen, et al. (2008). "Creating composite indicators with DEA and robustness analysis: the case of the Technology Achievement Index." Journal of the Operational Research Society **59**(2): 239-251.

Cylus, J. and P. Smith (2013). Comparative efficiency. Health system performance comparison: an agenda for policy, information and research. P. Smith, I. Papanicolas and J. Figueras. Buckingham, Open University Press.

Decancq, K. and M. A. Lugo (2010). Weights in Multidimensional Indices of Well-Being: An Overview, Available at SSRN: http://ssrn.com/abstract=1571124.

Dolan, P., C. Gudex, et al. (1996). "Valuing health states: a comparison of methods." Journal of health economics **15**(2): 209-231.

Doyle, J. and R. Green (1994). "Efficiency and Cross-Efficiency in DEA: Derivations, Meanings and Uses." The Journal of the Operational Research Society **45**(5): 567-578.

Esty, D. C., M. A. Levy, et al. (2006). Pilot Environmental Performance Index. Y. C. f. E. L. Policy. New Haven.

Farrell, M. J. (1957). "The Measurement of Productive Efficiency." Journal of the Royal Statistical Society Series a-General **120**(3): 253-290.

Gakidou, E., Murray, C., Frenk, J., (2000). Measuring Preferences on Health System Performance Assessment. GPE Discussion Paper 20. Geneva, World Health Organization.
Hollingsworth, B. and P. Smith (2003). "Use of ratios in data envelopment analysis." Applied Economics Letters **10**(11): 733-735.
Hollingsworth, B. and J. Wildman (2003). "The efficiency of health production: re-estimating the WHO panel data using parametric and non-parametric approaches to provide additional information." Health Economics **12**(6): 493-504.
Hussey, P., H. de Vries, et al. (2009). A Systematic Review of Health Care Efficiency Measures. Health Services Research.
Jacobs, R., P. Smith, et al. (2004). Measuring performance: An examination of composite performance indicators. CHE Technical Paper Series 29. York, The University of York, Centre for Health Economics.
Jacobs, R., Street, A., and Smith, PC.,, Ed. (2006). Measuring Efficiency in Health Care. , Cambridge: Cambridge University Press. .
Melyn, W. and W. Moesen (1991). Towards a synthetic indicator of macroeconomic performance: unequal weighting when limited information is available. Public Economic research Paper 17. KU Leuven, CES.
Munda, G. and M. Nardo (2005). Constructing Consistent Composite Indicators: the Issue of Weights. JRC32434, European Commission Joint Research Centre.
OECD (2010). OECD Health Data.
OECD/European Union/Joint Research Centre- European Commission (2005). Handbook on Constructing Composite Indicators: Methodology and User Guide, OECD Publishing.
Papanicolas, I. and P. e. Smith, Eds. (2013). Health system performance comparison: an agenda for policy, information and research., Maidenhead: Open University Press. .
Reinhardt, U. E., P. S. Hussey, et al. (2002). "Cross-National Comparisons Of Health Systems Using OECD Data, 1999." Health Affairs **21**(3): 169-181.
Sagar, A. D. and A. Najam (1998). "The human development index: a critical review." Ecological Economics **25**(3): 249-264.
Smith, P. (1997). "Model misspecification in Data Envelopment Analysis." Annals of Operations Research **73**(0): 233-252.
Smith, P. (2002). Developing Composite Indicators For Assessing Health System Efficiency. Measuring Up: Improving Health System Performance in OECD Countries. OECD. Paris, OECD.
Stone, M. (2002). "How not to measure the efficiency of public services (and how one might)." Journal of the Royal Statistical Society **Series A, 165**(3): 405-422.
Street, A. and U. Hakkinen, Eds. (2009). "Health system productivity and efficiency" in Smith, PC., Mossialos, E., Papanicolas, I. and Leatherman, S. (2009) Performance measurement for health system improvement: Experiences, challenges and prospects. , Cambridge: Cambridge University Press. .
Thanassoulis, E. (2001). Introduction to the theory and application of data envelopment analysis: a foundation text with integrated software. Dordrecht, Kluwer Academic Publishers.
WHO (2000). World Health Report 2000. Geneva, WHO.
Zhou, P., B. W. Ang, et al. (2007). "A mathematical programming approach to constructing composite indicators." Ecological Economics **62**(2): 291-297.