CrossMark

ORIGINAL RESEARCH ARTICLE

Health System Efficiency: A Fragmented Picture Based on OECD Data

Andreas Behr[1] · Katja Theune[1]

Abstract

Background Globally, health expenditure as a percentage of GDP has increased in recent years, so evaluating the health care systems used in different countries is an important tool for identifying best practices and improving inefficient health care systems.

Objective We investigate health system efficiency at the country level based on OECD health data. We focus on several aspects of health care systems to identify specific inefficiencies within them. This information hints at potential policy interventions that could improve specific parts of a country's health care system.

Methods A discussion is provided of ideal-typical evaluations of health systems, ignoring data restrictions, which provide the theoretical basis for an analysis performed under factual data restrictions. This investigation includes health care systems in 34 countries and is based on OECD health data. Health care system efficiency scores are obtained using data envelopment analysis (DEA). Relative productivity measures are calculated based on average DEA prices. Given the severe data limitations involved, instead of performing an all-encompassing analysis of each health care system, we focus on several aspects of each system, performing five partial analyses.

Results For each country, the efficiencies yielded by the five partial analyses varied considerably, resulting in an ambiguous picture of the efficiencies of the various health care systems considered. A synopsis providing comprehensive rankings of the analyzed countries is provided.

Conclusion Analysis of several aspects of the health care systems considered here highlights potential improvements in specific areas of these systems, thereby providing information for policymakers on where to focus when aiming to improve a country's health care system.

Key Points for Decision Makers

The analysis reveals strong within-country heterogeneity among the efficiencies of five different aspects of the health care system.

The results for several aspects of health care systems highlight potential improvements in specific areas of these systems.

Benchmark countries with highly efficient health care systems often present systems with high output and mediocre input or with mediocre output and low input, meaning that policymakers can select a role-model system according to their preferences.

1 Introduction

Providing health care is an important service of any country's government, and the efficiency of the health care system is a recurring and relevant key topic in health policy discussions. According to the Global Health Expenditure database of the World Health Organization (WHO), health expenditure as a percentage of GDP has increased strongly in recent years. Due to demographic changes, we can assume that this trend will continue in the coming decades. Hence, it is important to evaluate the efficiency of the

✉ Andreas Behr
  andreas.behr@vwl.uni-due.de

[1] University of Duisburg-Essen, 45117 Essen, Germany

△ Adis

health care systems of various countries. Efficiency analysis can point to potential input reductions or potential output increases. Therefore, benchmarking health care systems is important for identifying best practices and inefficient health care systems.

The contribution to the existing literature provided by the present paper is threefold. Firstly, we provide a theoretical discussion of ideal-typical measurements of health care efficiency, which serves as a yardstick when dealing with the severe data limitations that are often present when performing empirical analysis in this field. Secondly, because an all-encompassing analysis that includes inputs and outputs of very different aspects of the health care system can lead to compensating effects that hide specific strengths and weaknesses of that system, we perform five individual efficiency analyses, each focusing on different aspects of health care systems, thus avoiding any compensating effects. For example, Retzlaff-Roberts et al. [1] have shown that countries may be efficient at one aspect of health care (e.g., at improving life expectancy) but inefficient at other aspects (e.g., at reducing mortality). Thirdly, in addition to DEA efficiency scores, we provide a measure of relative productivity based on average prices, which prevents both the strengths of some health care systems from being exaggerated and unrealistic zero weights from being applied to specific inputs or outputs.

We investigate health care system efficiency at the country level based on OECD health data and apply data envelopment analysis (DEA). As the results and the scientific value of health care efficiency analysis depend crucially on the appropriateness of the inputs and outputs used, we discuss input and output choices in detail. In an empirical efficiency analysis, we consider several aspects of health care systems and conduct five partial analyses focusing on medical inputs and surgery provision; medical inputs and mortality prevention; lifestyle and life expectancy from birth; income, health expenditure, and life expectancy from birth; and relative expenditure, income inequality, and life expectancy from birth. We use data from 16–30 OECD countries in these five partial analyses.

For each country, we observed strong variations in efficiency across the five analyses; some countries are efficient at producing a particular health care output but very inefficient at producing other outputs. This emphasizes the value of disaggregated analysis, as one all-encompassing analysis including many inputs and outputs from different aspects of a health care system would mask specific inefficiencies in that system.

Furthermore, we find that different countries have achieved high efficiency in rather different ways; e.g., some health care systems yield low output but are highly efficient because only very low input quantities are invested in the first place. In terms of policy recommendations, the benchmark

health care system will of course depend on the preferences of the country that is choosing a health care system. These detailed results enable countries to prioritize and to focus on improving specific aspects of health production. As some inputs cannot be changed in the short term (e.g., health-related behavior), focusing on output improvements may be a more relevant approach for policymakers.

This paper is organized as follows: Sect. 2 contains some general methodological considerations. Section 3 provides a brief overview of the relevant literature on international comparisons of health care system efficiency. In Sect. 4, we discuss measurement issues, input and output selection, the database, and the methodology used. Section 5 contains the empirical results from the five separate analyses and a synopsis. Section 6 draws conclusions and contains policy implications.

2 General Methodological Considerations

Data envelopment analysis is a very popular method of obtaining efficiency scores for decision-making units (DMUs) in general and for countries' health care systems in our case. Its charm is its simplicity, as different health care systems are compared to the most efficient health care system, which most often is a synthetic health care system obtained as a linear combination of the observed health care systems of countries belonging to the reference set. Furthermore, the method is nonparametric, meaning that neither questionable assumptions about functional relations between inputs and outputs nor distributional assumptions need to be made.

2.1 Productivity, Relative Efficiency, and the Weighting Scheme

The productivity of a health care system is the ratio of the aggregated health output to the aggregated health input,

$$\frac{\text{Health output}}{\text{Health input}}. \tag{1}$$

The aim of efficient production is to maximize the output given a certain amount of input or to minimize the input given a certain amount of output. In our empirical efficiency analysis, we only consider the concept of relative efficiency. That is we simply compare the actual input amounts with lower hypothetical inputs of a (potentially synthetic) efficient health system given the actual output amounts. Hence, health care systems regarded as efficient serve as a benchmark when calculating the relative efficiency of the health care system being analyzed.

Note that a simple comparison of output relative to input is only possible if we sum the multiple different inputs for a health care system into a single input measure and the

multiple different outputs of the system into a single output measure. As the quantities of different goods cannot be summed, weights must be used when aggregating them.

We denote the outputs by y_r $(r = 1, \ldots, s)$, the inputs by x_i $(i = 1, \ldots, m)$, the output prices by u_r $(r = 1, \ldots, s)$, and the input prices by v_i $(i = 1, \ldots, m)$. j is an index that specifies the health care system considered; i.e., each value of $(j = 1, \ldots, n)$ refers to a different health system. Given the input prices v_i and the output prices u_r, we can compare the sum of weighted outputs with the sum of weighted inputs:

$$\frac{\text{Output}}{\text{Input}} = \frac{\sum_r y_r u_r}{\sum_i x_i v_i}. \qquad (2)$$

Note that scaling the output prices $c \cdot u_r$ and the input prices $k \cdot v_i$ alters the productivities but not the resulting relative efficiency measure, as the scaling factors cancel out. Hence, the resulting efficiency measure depends only on the relative input and output prices chosen.

2.2 Charnes–Cooper–Rhodes model

The ratio of weighted outputs to weighted inputs depends on the weighting scheme chosen. So how can we judge the efficiencies of DMUs if they depend on the weighting scheme applied? The idea behind the Charnes–Cooper–Rhodes model (CCR model) is to choose, for each DMU, the set of input and output prices that yields the maximum ratio of weighted output to weighted input, given a set of restrictions:

1. All input and output prices must be non-negative.
2. For all DMUs, the weighted output must not exceed the weighted input.

For the DMU under consideration (o), we choose v and u such that its productivity θ is maximized, under the constraint that, for all n DMUs, the weighted output does not exceed the weighted input and all prices are non-negative. Therefore, the problem is expressed as a linear programming problem which can be solved by means of the simplex algorithm. Obviously, the number of restrictions that have to be met here is rather large: it is the number of firms (n) plus the number of inputs (m) plus the number of outputs (s).

This maximization problem must be solved for each of the n DMUs, and $n + m + s$ constraints have to be met in each maximization problem. For firm o, the problem can be defined formally as

$$(\text{LP}_o) \qquad \max_{v,u} \theta = \frac{\sum_r y_{ro} u_r}{\sum_i x_{io} v_i}$$

$$\text{subject to} \quad \frac{\sum_r y_{rj} u_r}{\sum_i x_{ij} v_i} \leq 1 \quad (j = 1, \ldots, n) \qquad (3)$$

$$v_i \geq 0 \quad (i = 1, \ldots, m)$$

$$u_r \geq 0 \quad (r = 1, \ldots, s).$$

We only consider positive input prices $v > 0$ and positive input amounts $x > 0$, and we normalize the input of DMU o to 1. The maximization problem is then

$$(\text{LP}_o) \qquad \max_{v,u} \theta = \sum_r y_{ro} u_r$$

$$\text{subject to} \quad \sum_i x_{io} v_i = 1$$

$$\sum_r y_{rj} u_r \leq \sum_i x_{ij} v_i \quad (j = 1, \ldots, n) \qquad (4)$$

$$v_i \geq 0 \quad (i = 1, \ldots, m)$$

$$u_r \geq 0 \quad (r = 1, \ldots, s).$$

θ^* denotes the solution to the maximization problem. v^* and u^* are vectors of the optimal input and output prices. A DMU is efficient only if $\theta^* = 1$. In this case, its weighted output equals its weighted input and the restriction that the weighted output must not exceed the weighted input is just met. Aside from that restriction, we have the non-negativity constraints $v^* \geq 0$ and $u^* \geq 0$. If $\theta^* < 1$, then, for at least one DMU (and usually for several), we find that $\sum_r y_r u_r^* = \sum_i x_i v_i^*$. DMUs for which this equality holds belong to the reference set E_o' of the inefficient DMU o, where

$$E_o' = \left\{ j : \sum_r y_r u_r^* = \sum_i x_i v_i^* \right\}. \qquad (5)$$

The set of efficient DMUs in E_o' spans the "efficient frontier" for inefficient DMU o. Hence, all inefficient DMUs are measured relative to their specific reference sets.

3 Literature Review: International Comparisons of Health System Efficiency

There is a body of literature in which the efficiencies of the health care systems used in various countries are compared. Most of these studies make use of OECD health data or WHO health data. The World Health Report (WHR, e.g., [2]), published by the WHO, provides information on how well 191 countries are performing in relation to several health goals (improving health status and responsiveness, equality of financing) given their resources, and publishes health care system rankings. Efficiency is calculated as the ratio of observed performance to maximum performance, estimated by a fixed effects regression approach with observed inputs and outputs (see [3]; for more details, see, e.g., [4]). Evans et al. [3] evaluated the performance of 191 WHO member states using healthy life expectancy as outcome and health expenditure per capita and average years of schooling as input measures. They found efficiency to be positively related to health expenditure. Well-performing countries included Oman, Malta, Italy, France, and San Marino. The worst-performing countries were Zimbabwe, Zambia, Namibia, Botswana, and Malawi.

They concluded that countries with a good health level did not necessarily have an efficient health care system.

The WHO technique has been criticized for several reasons. Gravelle et al. [4] address the measure of efficiency, the fixed effects approach, the use of logarithms, and the omission of income from the WHO analyses. They observed substantial sensitivity of the results to the model's specifications, with the country's efficiency score and rank varying considerably depending on the specifications applied. In particular, they criticize the use of within-country variation, because most of the variation occurs between countries.

Hollingsworth and Wildman [5] compare parametric and nonparametric techniques [time-varying fixed effects, DEA, Malmquist indices, stochastic frontier analysis (SFA)] and emphasize that league tables hide valuable information on efficiency changes. In addition, they recommend that, because of their very different characteristics, OECD and non-OECD countries should be analyzed separately, and that further stratification by GDP or geographical region would also be useful.

Based on the critique of Gravelle et al. [4] concerning the between-country variation, Greene [6] used additional country-specific variables to account for some of the heterogeneity. Aside from health care expenditure and educational level, information on the income distribution (GDP and Gini coefficient), government effectiveness, dummy variables for tropical location and for OECD membership, population density, and an indicator relating to the allocation of health care expenditure between the private and public sector were included. The income level and the income inequality appeared to have sizeable impacts on the efficiency. Accounting for heterogeneity changed the rankings of the countries considerably.

As we use OECD data for our analysis, studies based on these data are reviewed here in more detail. Retzlaff-Roberts et al. [1] estimate the technical efficiency in the use of health care resources for 27 countries by applying input- and output-oriented DEA allowing for variable returns to scale (VRS). Based on the OECD health data for the year 2000, infant mortality and life expectancy at birth are defined as outputs. Inputs are categorized into two groups; one representing the social environment of the country (school expectancy, Gini coefficient, tobacco use) and the other consisting of health-related inputs [practicing physicians, inpatient beds, magnetic resonance imagers (MRIs), share of GDP allocated to health care]. They find 13 countries to be efficient according to both outputs: Australia, Canada, France, Greece, Ireland, Japan, Korea, Mexico, Norway, Spain, Sweden, Turkey, and the UK. This group includes countries with good health outcomes as well as those with poor health outcomes. Six countries are efficient at producing low infant mortality only, but not at producing high

life expectancy. Eight countries are inefficient with respect to both outputs: Austria, Belgium, Germany, Hungary, Netherlands, New Zealand, Switzerland, and the USA (i.e., again including countries with good and poor health outcomes). The authors conclude that a country can be technically efficient/inefficient at any level of health outcome. On average, inefficient countries could reduce infant mortality (with constant inputs) by 14.5% and increase life expectancy by only 2.1%. According to the input-oriented model, OECD countries can on average reduce inputs by 14.0 and 21.0%, respectively, without changing infant mortality and life expectancy.

Bhat [7] compares 24 OECD countries and uses the number of practicing physicians, nurses, inpatient beds, and pharmaceutical consumption as inputs. Output is defined based on three proportions of the population: those aged 0–19 years, those aged 20–64 years, and those aged 65 years or older, as health expenditure varies strongly with the age structure of the population. A DEA with constant returns to scale (CRS) identifies eight countries as efficient: Denmark, Japan, Netherlands, Norway, Portugal, Sweden, Turkey, and the UK. The lowest efficiencies are found for Belgium, Iceland, and Australia. Institutional arrangements seem to have an impact on efficiency, as public-contract and public-integrated countries are more efficient than countries with a public-reimbursement system. Here, public-reimbursement systems involve the retrospective indirect payment of providers for services. In the public-contract model, providers are paid a fee directly for their services, while the public-integrated model involves direct payment through global budgets and salaries. Control of spending is highest for the public-integrated model and lowest for public-reimbursement systems.

Afonso and Aubyn [8] analyze 24 OECD countries by applying two nonparametric approaches: free disposable hull (FDH) and DEA (input- and output-oriented, CRS and VRS). Just as in the study by Bhat [7], the numbers of doctors, nurses, and in-patient beds serve as inputs, and, in accordance with Retzlaff-Roberts et al. [1], infant survival rate and life expectancy are the health outcomes. Eight countries are found to be efficient when using the FDH and the DEA approaches: Canada, Japan, Korea, Portugal, Spain, Sweden, the UK, and the USA.

In a second paper, using a semiparametric two-stage approach, Afonso and Aubyn [9] find a strong relationship between health system inefficiencies and nondiscretionary inputs. An output-oriented DEA (VRS) based on principal components of base inputs (practicing nurses, physicians, beds, MRI) and outputs (infant survival rate, life expectancy, potential years of life not lost) identifies seven countries as efficient: Canada, Finland, Japan, Korea, Spain, Sweden, and the USA. On average, countries can increase their output by 40%. In a second stage, the fact that inefficiencies are not

necessarily caused by factors under the purview of the health care system is focused upon. The influence of nondiscretionary inputs on the efficiency is evaluated based on a Tobit regression. Here, the nondiscretionary inputs of GDP per capita, educational level, obesity, and tobacco consumption appear to have an impact on a country's efficiency. Correcting for environmental influences alters efficiency scores and country rankings. Countries that were poorly ranked before come closer to the efficient frontier (e.g., Denmark, Czech Republic, Hungary, Slovak Republic, and the UK), whereas the rankings of other countries decline (Canada, Sweden, USA, Japan).

A within-country comparison over time rather than a between-country comparison is conducted by Adang and Borm [10]. They apply an output-oriented DEA (CRS) and a Malmquist index over the period 1995–2002 and calculate Spearman correlation coefficients to identify the relationship between changes in productivity and changes in satisfaction in health care. Inputs are the share of GDP allocated to health care, the number of practicing physicians, and tobacco use. Outputs are life expectancy at birth and infant mortality. No association between the economic performance of the health care system and the change in satisfaction with the health care system is found.

A critical study involving between-country comparisons of health production efficiency is provided by Spinks and Hollingsworth [11]. Twenty-eight countries are analyzed based on OECD health data between 1995 and 2000. School expectancy years, unemployment rate, GDP per capita, and health expenditure per capita serve as inputs; life expectancy at birth serves as the health output indicator. An input-oriented DEA (VRS) analysis and Malmquist indices reveal that six countries were efficient in 1995 (Turkey, Mexico, Korea, Greece, Spain, and Japan) and eight countries were efficient in 2000 (Turkey, Mexico, Korea, Greece, Spain, Japan, Iceland, and Switzerland). Most of these efficient countries are lower ranked in GDP per capita and health expenditure per capita.

Table 1 provides an overview of the input and output measures used in previous research based on OECD data. An overview of the literature that has utilized the OECD data is given in Table 8 in the Appendix.

A recent study by Hsu [12] uses world development indicators and global development finance data for 46 countries between 2005 and 2007 to compare the efficiency of government health expenditure between Europe and Central Asia. Health expenditure per capita is used as input for a DEA. In addition, the effects of environmental variables such as the population density, GDP per capita, hospital beds, average years of primary schooling, and a regional effect (Central Asia versus Europe). Outputs are life expectancy at birth, infant mortality rate (reciprocal), and measles immunization. Their results indicate that health spending in Europe is less efficient than that in Central Asia.

Table 1 Inputs and outputs used in the OECD data

Inputs	Outputs
Health expenditure per capita	Life expectancy (at birth)
Percentage of health care expenditure	Infant mortality (survival) rate
Practicing physicians	Population aged 0-19 years; population aged 20-64 years; population aged 65 years and older
Nurses	Potential years of life not lost
In-patient beds	
MRI	
Pharmaceutical consumption	
Average years of schooling/ educational level	
Gini coefficient	
GDP per capita	
Unemployment rate	
Obesity	
Tobacco consumption	
Alcohol consumption	

MRI magnetic resonance imaging, *GDP* gross domestic product

Moreover, the number of hospital beds and education positively affect efficiency. Using the same data and a super slack-based measure (SBM), Hsu [13] further discriminates between efficient countries in order to rank them separately.

Medeiros and Schwierz [14] apply models with different combinations of inputs and outputs to assess the robustness of the results, and they conclude that efficiency scores vary strongly across efficiency models.

Summing up, we observe that a variety of input and output measures have been used in previous studies focusing on different sets of countries. However, the process by which the inputs affect the outputs is often insufficiently identified. As many studies do not include an ideal-typical characterization of the effect of health care provision on population health, the procedure used to select inputs and outputs often lacks the required theoretical underpinning. Based on this literature review, we considered it necessary to first try to characterize an ideal-typical measurement of the effects of a health system on population health, which would then serve as the foundation for operationalization under the restrictions of the OECD database.

The contribution of the present work to the existing literature is threefold. Firstly, it provides a theoretical discussion of ideal and operationalized inputs and outputs, which are important when dealing with severe data limitations. Secondly, it extends previous research by applying separate models and therefore emphasizing different objectives of the health care system. Thirdly, along with DEA efficiency

scores, it provides an efficiency measure based on average prices, which prevents the strengths of particular health care systems from being overemphasized and specific poorly performing inputs or outputs from being overlooked.

4 Methods

The results and the scientific value of health care efficiency analysis depend crucially on the appropriateness of the inputs and outputs considered in the analysis. We start our discussion by formulating ideal-typical concepts (as defined by Max Weber) of inputs and outputs. These ideal-typical concepts are then contrasted with the available inputs and outputs in the OECD database.

4.1 Ideal-Typical Input Measures

Ideally, the inputs should reflect all of the resources used by the health care system to improve the health status of the population. The inputs can be broadly classified into labor (in its different forms), materials, and the services provided by the capital stock in use. In classical economics, the total input is considered the sum of the direct labor input, the indirect labor input associated with the material consumed during the process of producing health care, and the indirect labor input associated with the portion of the capital stock that is consumed during the production process. In contrast to the classical use of labor coefficients to transform the quantities of all the different inputs into amounts of labor, other weighting schemes can be applied when aggregating the inputs.

The direct labor inputs can be broadly categorized into service by doctors, service by therapists (non-doctoral), services by nurses and care workers, services by pharmacists, and services by administrative staff. The material inputs can be further categorized according to commodity group into chemical goods, care materials, etc. The fraction of the capital stock consumed may be disaggregated into buildings, machinery (e.g., for drug production), and equipment (e.g., beds, MRIs, etc.; see Fig. 1).

Readers familiar with input–output tables could conceptualize the health system as one specific sector, which would provide the necessary information as quantities with different units of measurement. For instance, this information could be expressed in labor values, based on employment data for the relevant sectors (ignoring the issue of nonhomogeneous labor for the moment).

4.2 Ideal-Typical Output Measures

Starting with the idea of describing the health sector as an individual sector in the input–output analysis, the output
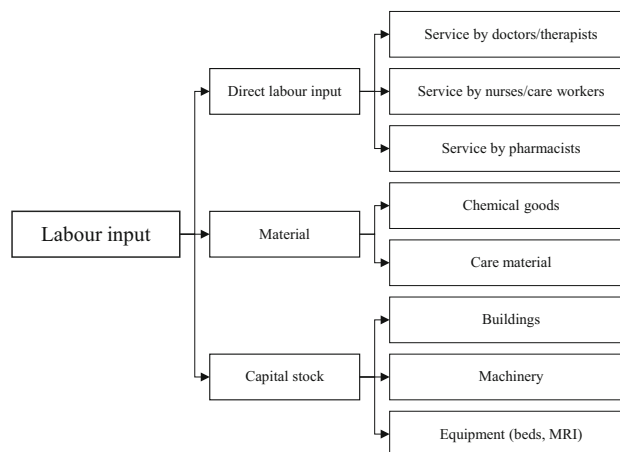

Fig. 1 Ideal-typical inputs. *MRI* magnetic resonance imaging

would be the aggregated value of health goods and services produced, thereby reflecting the individual valuation of the output from the buyers' (patients') perspective. But, as most of the products of the health system are not "sold" at market prices, this concept of subjective valuation of the output is not feasible.

Going back to an ideal-typical concept of health output, three quantities may best capture the output of the health sector: the amount of pain relief, the number of additional (quality-adjusted) life years, and the increase in well-being (e.g., achieved through the use of prostheses) caused by treatment within the health system. These initially purely theoretical concepts could perhaps be evaluated by summing the values of each measure in the various subsectors: hospitals, medical practices, pharmacies, etc. Alternatively, the aggregation could be performed by summing across patient categories (e.g., acute patients, patients with chronic conditions, etc.) or population subgroups (e.g., grouped by age and sex). Hence, ideal-typical outputs could be the number of extra pain-free hours (weighted by objective pain measures) due to treatment, the sum of the extra (quality-adjusted) life years (QALYs) gained by undergoing treatment within the health system, and the increase in well-being caused by treatment (see Fig. 2).

Obviously, the causal relation between treatments and measures often cannot be verified in practice. Furthermore, the intersubjective comparable measures of pain relief that need to be aggregated are purely theoretical, as is the calculation of additional (quality-adjusted) life years or the increase in well-being. Nevertheless, we consider this ideal-typical reasoning to be very important when judging the adequacy of available output indicators.

4.3 The OECD Database

The OECD collects health-related data on OECD countries from various data sources such as the Eurostat database,
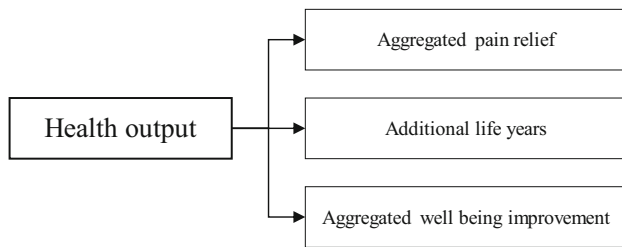
Fig. 2 Ideal-typical outputs

The World Bank, the Global Health Observatory data repository of the World Health Organization (WHO), and other OECD databases (OECD.Stat). These data are complemented by national databases (e.g., statistical offices, federal ministries).

The OECD health data provide information on health status (e.g., mortality, morbidity), nonmedical determinants of health (e.g., tobacco, alcohol, food consumption, obesity), health care resources (e.g., health employment and education, physical and technical resources), health care utilization (e.g., consultations, immunization, screening, hospital stay, surgical procedures, waiting times), health care quality indicators, the pharmaceutical market, expenditure on health, social protection, demographic factors (e.g., population, fertility, labor force), and economic factors (e.g., GDP, wages) (see [15] for example).

Differences in the data sources, definitions, and methodologies used for variable construction affect comparability across countries. For instance, when comparing the number of practicing physicians, some countries include doctors working in administration, management, academic, and research positions, while others only include doctors who provide care directly to patients. Similarly, when comparing the number of nurses, some countries also include midwives (considered to be specialist nurses) in their figures, while others do not. When comparing surgeries, it is important to realize that classification and registration practices differ among countries. For example, some countries group data on total hip replacements and partial replacements together, while others only provide total hip replacement data. Countries also apply different rules for registering technical equipment (e.g., MRI units or CT scanners): most OECD member states include all such equipment, regardless of whether it is used in a hospital, but some countries only include equipment used in hospitals (see [16]).

Moreover, a lack of standardization in health interview surveys (differences in wording and in response categories) complicates between-country comparisons. For example, the perceived health status can be retrieved from The European Union Survey on Income and Living Conditions (EU-SILC), but in this case there is no measurement standardization across OECD countries. Similarly,

countries may differ in the data they provide on health-related behavior such as tobacco and alcohol consumption (see [16]).

The main source of data on infant mortality in most European countries is the Eurostat database. Some of the between-country variation in infant mortality rates is probably due to differences in registering practices for premature infants. The methodology used to estimate life expectancy, data on which are again mainly retrieved from the Eurostat database, may also differ between countries. Life expectancy at birth for the whole population is calculated by the OECD Secretariat (see [16]).

These limitations in between-country comparability should be kept in mind when interpreting the empirical results. Moreover, we face the problem of missing data in the OECD database, which limits empirical analyses dramatically. For our analysis, we used the OECD health data for 2012. In the rather rare cases in which entries were missing for 2012 but available for 2011, we used the 2011 values to substitute for the missing 2012 values. Only six countries could be included in all five of the partial analyses, so the number of available scores for partial analyses varied between 1 and 5 for the 34 countries.

4.4 Available Input and Output Measures and their Adequacy

We scanned the complete available OECD health database to find the most adequate input and output indicators. Not unexpectedly, the available indicators do not remotely coincide with the ideal-typical measures discussed above. Note that we disregarded from the start all indicators with <20 valid data (i.e., with values from <20 countries).

4.4.1 Inputs

In the literature, health care expenditure is often used as an input measure. Unfortunately, however, some of this expenditure depends on the country's wage levels (for nurses, physicians, etc.) while the rest does not (e.g., expenditure on imported magnetic resonance imaging machines). Moreover, disentangling the effects of private and public spending on health outcomes is a challenging task (see [17]), and it is unclear how variations in the public financing of health care influence a country's health outcomes (see [6]). Not all of the public expenditure on health has a measurable effect on health outcomes (e.g., the effect of the expenditure on tooth cleaning). Furthermore, we face the problem of differentiating between health expenditure and consumption (e.g., designer and basic spectacles).

Nonfinancial health care resources often used in the literature include practicing physicians and nurses and inpatient beds per 1000 population. Moreover, to take into

account health care technology, technical equipment such as MRI units is considered medical input (see [1]). Medical services such as surgeries and transplantations are not usually considered in research, but are assumed to be highly relevant to health care efficiency.

Due to issues with availability and missing data, we cannot cover all of the aspects of the ideal-typical inputs mentioned above. We categorize the inputs used into the three broad categories of basic medical inputs, intermediate medical inputs (accomplishments, provided services), and financial inputs:

1. Basic medical inputs

   – hosp_beds: total hospital beds per 1000 population
   – physicians: practicing physicians per 1000 population
   – nurses: practicing nurses per 1000 population.

2. Intermediate medical inputs (used as outputs)

   – cataract: cataract surgery, total number of procedures per 100,000 population
   – bypass: coronary artery bypass graft, inpatient cases per 100,000 population
   – kidney: transplantation of kidney, total number of procedures per 100,000 population.

3. Financial inputs

   – healthgdp: health expenditure as a share of gross domestic product
   – healthspend: total health expenditure per capita in USA dollars based on PPP.

4.4.2 Outputs

Unfortunately, the ideal (or almost ideal)-typical output measures that have already been mentioned are not available in health data sets. Life expectancy and infant mortality are often used as health output indicators, but these outputs are, of course, incomplete measures of health status since they do not reflect the quality of life of the living. Life expectancy depends strongly on social, cultural, and environmental factors (especially at the time of birth). Infant mortality strongly depends on health-system and hygiene standards but focuses on a very specific aspect of health. In addition, life expectancy and mortality rates are probably affected by country-specific rates of accidents, homicides, and suicides. It is difficult to define health care outcomes in the same way for all countries using a single factor, because different countries have different policy priorities and aims (see [17]). Research shows that there are strong differences in the extent to which the health system influences performance measures, ranging from large

effects on measures such as waiting time to very small effects on mortality, which is strongly influenced by factors that are not within the purview of the health care system (see [18]). High infant mortality rates (e.g., in the USA) may be due to factors such as poverty rather than low health care system efficiency (see [1]). To account for mortality rates that are less strongly affected by these factors, we include 30-day mortality after ischemic stroke and acute myocardial infarction in our analyses.

We categorize the output indicators into two groups: specific outputs and unspecific outputs. While recognizing that this distinction is somewhat fuzzy, we regard the specific outputs as more closely related to the effects of the health system than the unspecific outputs. For instance, infant mortality depends crucially on the hygiene standards applied and the medical care provided during birth. This is not true of general life expectancy, which is a weighted average of the number of deaths in a given year across all ages. Therefore, the living conditions and health provision across a period of about 100 years determine the actual measure of life expectancy. This means that life expectancy can, at best, be only partially attributed to the present state of the health system. We use the following variables for our analyses:

1. Specific outputs

   – infmort: infant mortality, deaths per 1000 live births
   – ischemicstroke30: 30-day mortality after admission to hospital for ischemic stroke per 100 patients (based on admission data).
   – amim30: 30-day mortality after admission to hospital for acute myocardial infarction (AMI) per 100 patients (based on admission data).

2. Unspecific outputs

   – lifeexp: life expectancy at birth.

4.4.3 Economic Conditions and Lifestyle

It is not easy to correlate health status with the activities performed by health care systems. There are other non-health variables that are outside the range of influence of the health care system but can affect health, such as environmental conditions (e.g., temperature, pollution), socioeconomic factors (e.g., education, income, employment), and health-related behaviors and lifestyle (e.g., alcohol and tobacco consumption, hygiene) of the population. But these factors are not routinely measured and they are highly correlated with each other. For instance, many behavioral factors are correlated with educational attainment (see [3]) and measures of income. Therefore, we

face the problem of using a combination of macro (e.g., education, employment) and intermediate (e.g., smoking, alcohol) factors in the DEA model which could lead to biased results (see [11]). We categorize the analyzed variables into economic conditions and lifestyle inputs:

1. Economic conditions

   – gdp_ppp: GDP per capita in US dollars, based on PPP
   – gini: Gini coefficient of income distribution (disposable income after taxes and transfers).

2. Lifestyle input

   – alcohol: alcohol consumption in liters per capita, population aged 15+
   – tobacco: % of population aged 15+ who are daily smokers
   – obesity: % of population aged 15+.

4.5 Methodology

Unfortunately, a large number of values are missing from the OECD health data. An all-encompassing analysis would only be possible for an extremely small number of countries. Thus, we instead focus on several aspects of health systems and conduct five partial analyses, as displayed in Fig. 3. Note that the analyses differ in the number of countries they consider. Moreover, we assume that separate analyses of specific aspects of health care provision are useful for providing detailed information on how a country's health care system works.

The first analysis (analysis I) focuses on the efficiency of surgery provision. Outputs are cataract and bypass surgeries and kidney transplantations (per 100,000 population). We use the numbers of physicians, nurses, and beds (per 1000 population) as medical inputs.

In the second analysis (analysis II), we focus on the efficiency of mortality prevention, defined by infant mortality, mortality 30 days after stroke, and mortality 30 days after infarct. Medical inputs are again the numbers of physicians, nurses, and beds (per 1000 population).

The third analysis (analysis III) concentrates on the effects of lifestyle on life expectancy from birth. Alcohol consumption, smoking, and obesity are used as lifestyle inputs.

The focus of the fourth analysis (analysis IV) is on the effects of income and health expenditure on life expectancy from birth. Income is measured as the gross domestic product per capita (GDP/Pop.). Expenditure is measured as the health expenditure per capita (HExp/Pop.).

In the fifth analysis (analysis V), we concentrate on the effects of relative expenditure, measured as the ratio of health expenditure to GDP (HExp/GDP), and inequality, measured as the Gini coefficient, on life expectancy from birth.

We use an input-oriented data envelopment analysis (DEA), as described in Sect. 2, with constant returns to scale. The idea of DEA is that price vectors are selected for each unit that maximize efficiency given a set of constraints, i.e., DEA provides the "optimal" price vectors u^* and v^* for all n health systems. u^* and v^* vary strongly between units, and in many cases several input and output prices are zero. Even when DEA was first applied by Charnes et al. [19], some DMUs obtained zero weights, and all but one of the several inputs of some efficient units had zero weights.

Charnes et al. [20] question restrictions on the firm's individual optimal weights. Allen et al. [21] provide a discussion of value judgements of weights obtained by DEA and a priori weight restrictions by the researcher, and regard the existence of zero weights as a conceptual problem with DEA. Due to the nature of value judgements, they understandably conclude that "There is no all purpose method for translating value judgements into restrictions on DEA weights" (p. 30). To circumvent the introduction of value judgements, we decided to use means of DEA weights as an alternative to original DEA scores.

In detail, we use DEA weights averaged over all countries to aggregate inputs and outputs, and we construct an

Fig. 3 Five partial analyses. *GDP* gross domestic product, *Pop.* population, *HExp* health expenditure
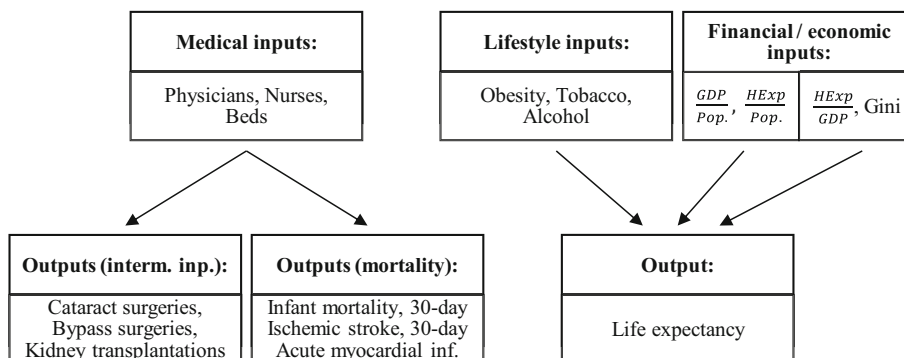
alternative relative productivity measure: the ratio of a country's productivity to the productivity of the most productive country. We obtain average "optimal" input and output price vectors $\bar{u}^*$ and $\bar{v}^*$ based on the n individual input and output price vectors. We calculate aggregates of the inputs and outputs based on $\bar{u}^*$ and $\bar{v}^*$ as follows:

$$\text{Productivity} = \frac{\text{output}}{\text{input}} = \frac{\sum_r y_r \bar{u}_r^*}{\sum_i x_i \bar{v}_i^*}.$$

The impact of varying the weights of the health system outputs is discussed in, for example, Lauer et al. [22]. They criticize the use of fixed output weights for all of the countries considered in the WHO health-system performance rankings. Fixed weights do not account for the development status or cultural traditions of a particular country, or for the different political objectives of different countries. This issue becomes more important as the range of income levels across the countries considered increases. Therefore, this issue is of less concern when investigating the (relatively similar) OECD countries than when considering the broader set of countries included in the WHO database. Hauck and Street [23] try to avoid the drawbacks of aggregating multiple objectives into one single index by applying a multivariate multilevel model.

To carry out the DEA, we have to transform some of our variables. The output (input) should be ≥ 0, and the output (input) should be "good" instead of "bad." For instance, Cheng and Zervopoulos [24] use a generalized directional distance function to handle desirable and undesirable outputs. Seiford and Zhu [25] propose a linear monotonically decreasing transformation for undesirable outputs and output-decreasing inputs. In our analysis, mortality rates and lifestyle inputs such as the amount of alcohol must be transformed. x denotes the original variable (e.g., mortality rate), $\tilde{x}$ is the scaled variable (with a mean of 0 and a standard deviation of 1), and x^* the transformed variable:

$$\tilde{x} = \frac{x - \text{mean}(x)}{\text{sd}(x)}$$
$$x^* = -\tilde{x} - \min(-\tilde{x}) + 1.$$

Note that x^* has a minimum value of 1 for the health system with the smallest output (i.e., the highest mortality rate) and a standard deviation of 1. The Gini coefficient is transformed into $1 - \text{Gini}$, as $1 - \text{Gini}$ is regarded as a sensible equality measure.

5 Empirical Results

In the following, we present results for the five analyses separately. The tables provide information on the inputs and outputs used, the relative inputs and outputs (measured

as the ratio of the country's weighted (by the mean input/output prices) sum of inputs/outputs and the maximum inputs/outputs of any country), the DEA efficiency scores, as well as the relative productivity of each country, based on the mean of the country-specific optimal prices across all countries. Relative productivity is by definition normalized to the interval [0, 1] and is therefore easily interpretable.[1] Note that the average implicit weights in the last row of each of the following tables refer to the scaled variables, and—for ease of interpretation—we also provide the original input and output values in the tables.

We observe that the optimal prices for inputs and outputs vary considerably across countries, with several prices being zero. This fact is discussed by Lauer et al. [22], who criticize the fact that a country can greatly increase its efficiency score by assigning zero weights to inputs and/or outputs it is not performing well in.

5.1 Analysis I: Surgeries

Table 2 shows the results for analysis I, comparing the efficiency of surgery provision.[2] According to the DEA efficiency scores for each of the 19 considered countries, we find 7 countries to be efficient (score of 1): Belgium, Canada, Estonia, Hungary, Korea, Spain, and Sweden. Very inefficiently performing countries are Lithuania (0.649), New Zealand (0.652), and Germany (0.658); for instance, Lithuania could potentially reduce its inputs by 35.1% but still achieve its current output.

The measure of relative productivity reveals Hungary to be the most productive country (1.000), followed by Estonia (0.979), and Korea (0.932). We observe low productivities for New Zealand (0.227) and Denmark (0.370). Comparing the DEA efficiency score with the relative productivity reveals, for the majority of countries, a strong divergence between both measures, with the latter being considerably lower than the former. For example, for Sweden and the UK, the relative productivity is about half the DEA efficiency score (with country-specific input prices strongly diverging from the mean prices). Therefore, variations in input and output prices appear to have a strong impact on the productivity and efficiency values of the countries considered (see also [22]). The correlation between DEA score and relative productivity based on average prices is 0.61. These results indicate that in DEA, countries emphasize their strengths and ignore aspects of health provision at which they are poor performers.

Table 2 Analysis I: medical inputs and surgeries

	Beds	Phys.	Nurses	Catar.	Bypass	Kidney	Rel.inp.	Rel.outp.	Eff.	Rel.prod.
Austria	5.460	4.900	7.830	11.250	0.435	0.049	0.488	0.989	0.865	0.819
Belgium	3.990	2.930	9.510	10.921	0.644	0.049	0.586	1.000	1.000	0.690
Canada	1.710	2.480	9.350	10.531	0.561	0.038	0.574	0.953	1.000	0.670
Czech Republic	4.560	3.670	8.060	10.399	0.501	0.041	0.499	0.932	0.834	0.754
Denmark	2.530	3.620	16.300	9.799	0.679	0.047	1.000	0.916	0.846	0.370
Estonia	3.550	3.280	6.170	10.734	0.327	0.045	0.383	0.928	1.000	0.979
Finland	2.920	3.010	14.120	9.538	0.472	0.036	0.866	0.857	0.762	0.400
Germany	5.380	3.960	12.610	9.643	0.684	0.032	0.777	0.903	0.658	0.470
Hungary	3.980	3.090	6.320	9.969	0.910	0.030	0.392	0.970	1.000	1.000
Israel	1.890	3.250	4.820	5.416	0.372	0.023	0.300	0.505	0.887	0.680
Korea	6.120	2.080	4.840	8.417	0.066	0.036	0.301	0.693	1.000	0.932
Lithuania	5.380	4.220	7.590	5.314	0.621	0.028	0.472	0.542	0.649	0.464
Luxembourg	3.960	2.780	11.920	9.618	0.465	0.000	0.732	0.859	0.815	0.474
New Zealand	2.610	2.700	9.970	3.270	0.437	0.024	0.613	0.344	0.652	0.227
Poland	4.320	2.230	5.560	3.480	0.546	0.026	0.344	0.381	0.935	0.447
Slovenia	3.620	2.540	8.160	8.914	0.406	0.011	0.503	0.793	0.860	0.637
Spain	2.300	3.820	5.240	6.116	0.177	0.055	0.327	0.529	1.000	0.654
Sweden	1.950	4.010	11.150	9.391	0.346	0.045	0.687	0.823	1.000	0.484
UK	2.310	2.750	8.210	6.198	0.287	0.040	0.506	0.554	0.942	0.443
Mean	3.607	3.227	8.828	8.364	0.470	0.034	0.545	0.762	0.879	0.610
Impl. pr.	0.045	0.270	8.589	0.069	0.154	0.060				

Phys. physicians, *Catar.* cataract, *Eff.* efficiency, *Prod.* productivity

5.2 Analysis II: Mortality

Table 3 presents the results for analysis II, focusing on the efficiency of mortality prevention. We find 6 out of the 22 considered countries to be efficient: Canada, Finland, Israel, Korea, Slovenia, and Sweden. Canada, Korea, and Sweden were also identified as efficient in the first analysis. Very inefficient are Austria (0.624), again Germany (0.645), and Mexico (0.403), which is an obvious outlier. A closer look at the used (relative) inputs and outputs reveals low input and a very low output in Mexico. We find that efficient countries achieve their efficiency in rather different ways; for example, by investing a medium quantity of input but producing comparably high output (e.g., Canada) or by investing a high quantity of input and producing very high output (e.g., Sweden). This fact is also mentioned by Afonso and Aubyn [8].

The measures of relative productivity reveal Israel to be the most productive country (1.000), followed by Spain (0.788), Korea (0.771), and Canada (0.762). As also seen in analysis I, we observe some divergence between the DEA efficiency score and the relative productivity for most countries, with the latter again being considerably lower than the former. Despite the marked differences between the values of these measures, the correlation between DEA score and relative productivity based on average prices is actually rather strong (0.77).

5.3 Analysis III: Lifestyles

Results for analysis III, focusing on the effects of different lifestyles on life expectancy, are summarized in Table 4. Here, we observe 16 countries, 5 of which are identified as efficient: Estonia, Luxembourg, Mexico, UK, and the United States. Very low efficiency scores are recorded for Japan (0.580), Sweden (0.626), and Norway (0.643). Interestingly, Sweden was found to be efficient in the first two analyses and Mexico to be substantially inefficient in analysis II. A closer look at inputs and outputs indicates that Mexico is characterized by the highest percentage of obesity in the population, but life expectancy is sufficient for it to be characterized as an efficient country.

Based on the relative productivity, we again find Sweden (0.543), Japan (0.565), and Norway (0.591) to perform rather inefficiently. The highest productivity is reported for Estonia (1.000), followed by the UK (0.911). Again, we observe some differences between DEA efficiency scores and relative productivity values; the correlation between DEA score and relative productivity is 0.78. For Mexico, we observe a large difference between scores based on

Table 3 Analysis II: medical inputs and mortality

	Beds	Phys.	Nurses	Infmort	Stroke30	Ami30	Rel. inp	Rel.outp.	Eff.	Rel.prod.
Australia	3.360	3.310	10.220	5.227	3.830	5.984	0.743	0.859	0.846	0.623
Austria	5.460	4.900	7.830	5.269	4.643	4.749	0.922	0.899	0.624	0.525
Belgium	3.990	2.930	9.510	5.015	3.858	5.333	0.741	0.829	0.814	0.603
Canada	1.710	2.480	9.350	4.593	3.550	5.468	0.551	0.778	1.000	0.762
Czech Republic	4.560	3.670	8.060	5.522	3.802	5.445	0.787	0.858	0.756	0.588
Denmark	2.530	3.620	16.300	5.184	3.915	5.557	0.883	0.851	0.771	0.520
Estonia	3.550	3.280	6.170	5.100	2.794	4.457	0.637	0.709	0.841	0.600
Finland	2.920	3.010	14.120	5.607	4.923	5.378	0.807	0.965	1.000	0.645
Germany	5.380	3.960	12.610	5.227	4.587	4.974	0.994	0.898	0.645	0.487
Iceland	2.290	3.570	15.160	6.156	4.223	5.423	0.830	0.936	0.991	0.607
Israel	1.890	3.250	4.820	5.100	4.447	5.535	0.482	0.895	1.000	1.000
Japan	7.940	2.290	10.540	5.691	5.624	4.233	1.000	1.000	0.958	0.539
Korea	6.120	2.080	4.840	5.396	5.512	5.019	0.696	0.995	1.000	0.771
Luxembourg	3.960	2.780	11.920	5.565	3.915	5.400	0.799	0.870	0.914	0.587
Mexico	1.530	2.120	2.550	1.000	1.000	1.000	0.318	0.185	0.403	0.313
New Zealand	2.610	2.700	9.970	4.423	4.083	5.445	0.645	0.816	0.950	0.682
Norway	2.340	4.230	16.530	5.565	4.895	5.310	0.917	0.958	0.879	0.563
Slovenia	3.620	2.540	8.160	5.945	2.766	5.535	0.651	0.791	1.000	0.655
Spain	2.300	3.820	5.240	5.311	3.606	5.153	0.560	0.818	0.955	0.788
Sweden	1.950	4.010	11.150	5.522	4.783	5.961	0.719	0.965	1.000	0.723
Switzerland	2.990	3.920	16.970	5.100	4.531	5.243	0.954	0.893	0.774	0.505
UK	2.310	2.750	8.210	4.888	3.662	5.198	0.576	0.799	0.952	0.747
Mean	3.423	3.237	10.010	5.109	4.043	5.082	0.737	0.844	0.867	0.629
Impl. pr.	0.076	0.073	0.032	0.101	0.152	0.050				

Phys. physicians, *Infmort* infant mortality, *Stroke30* 30-day mortality after admission to hospital for ischemic stroke per 100 patients, *Ami30* 30-day mortality after admission to hospital for acute myocardial infarction per 100 patient, *Eff.* efficiency, *Prod.* productivity

country-specific prices and on mean prices. While Mexico is regarded as efficient according to its DEA score, its efficiency in terms of relative productivity is below average.

5.4 Analysis IV: Per Capita Income and Health Care Spending

In the fourth analysis (Table 5), investigating the efficiency of income and health expenditure per capita, we observe only Mexico and Turkey to be efficient among the 30 countries considered. Luxembourg (0.222), Norway (0.278), Switzerland (0.335), and the USA (0.346) are characterized as very inefficient countries. All of these countries use high input quantities (primarily GDP for Luxembourg, health expenditure for USA), which is not sufficiently reflected in their life expectancies. Interestingly, the efficient countries Mexico and Turkey provide very low input and produce low output. This indicates that increased absolute health spending and income per capita have only moderate effects on life expectancy.

Based on the relative productivity, Mexico is the most productive country (1.000), followed by Turkey (0.943) and Chile (0.829). Countries with a low relative productivity are again Luxembourg (0.201) and Norway (0.276). The correlation between both efficiency measures is close to 1.

5.5 Analysis V: Relative Health Care Spending and Inequality

Analysis V focuses on the efficiency of relative health expenditure and inequality. Results for the 27 countries are presented in Table 6. Resembling the results of analysis IV, only two countries have an efficiency score of 1: Mexico and Turkey. The next most efficient countries are Israel (0.950) and the USA (0.943). In general, efficiency scores are close to 1; the lowest score of 0.765 is reported for the Slovak Republic. Mexico as well as Turkey report high Gini coefficients and low life expectancies. As income inequality has a negative but rather indirect effect on life expectancy, we find that the ultimately negative effect of

Table 4 Analysis III: lifestyle inputs and life expectancy

	Alcohol	Tobacco	Obese	Lifexp	Rel. inp	Rel.outp.	Eff.	Rel.prod.
Canada	8.100	16.100	50.900	81.450	0.852	0.980	0.710	0.635
Estonia	12.200	27.250	48.900	76.450	0.508	0.919	1.000	1.000
Finland	9.300	17.450	49.400	80.700	0.820	0.971	0.748	0.654
France	11.700	24.450	44.400	82.050	0.645	0.987	0.922	0.845
Japan	7.200	21.550	23.700	83.150	0.977	1.000	0.580	0.565
Luxembourg	11.300	16.800	59.200	81.450	0.726	0.980	1.000	0.745
Mexico	5.700	12.300	71.300	74.350	0.764	0.894	1.000	0.647
Netherlands	9.100	18.450	47.900	81.150	0.806	0.976	0.738	0.669
New Zealand	9.200	16.500	63.800	81.150	0.697	0.976	0.967	0.774
Norway	6.200	16.000	46.000	81.500	0.916	0.980	0.643	0.591
Slovenia	11.000	20.500	56.900	80.200	0.638	0.965	0.988	0.836
Sweden	7.300	12.750	47.100	81.750	1.000	0.983	0.626	0.543
Switzerland	9.900	20.450	41.000	82.750	0.814	0.995	0.732	0.676
Turkey	1.600	24.000	52.000	74.600	0.627	0.897	0.887	0.791
UK	9.700	20.500	61.900	80.950	0.591	0.974	1.000	0.911
USA	8.800	14.200	68.600	78.800	0.719	0.948	1.000	0.729
Mean	8.644	18.703	52.062	80.153	0.756	0.964	0.846	0.726
Impl. pr.	0.011	0.109	0.102	0.224				

Lifexp life expectancy at birth, *Eff.* efficiency, *Prod.*: productivity

high inequality (e.g., in Mexico) is not reflected in an accordingly low life expectancy.

Relative productivity is lowest for the Slovak Republic (0.742). Mexico has the highest relative productivity (1.000), followed by Israel (0.949) and the USA (0.942). We observe only small differences between DEA efficiency scores and relative productivity, and a very strong correlation between both efficiency measures (0.92).

5.6 Synopsis

Table 7 in the Appendix provides an overview of the DEA scores obtained in all five analyses of 34 countries. The mean is calculated on the basis of 1–5 DEA efficiency scores, and the ranking is based on the mean. We observe Iceland to be the most efficient country, followed by Turkey and Estonia. The lowest mean efficiency scores are reported for Ireland and Germany.

Efficiency scores vary strongly across the five analyses, with some countries performing efficiently in one or more analyses and very inefficiently in others (e.g., Mexico, Canada, and Sweden). Therefore, countries can indeed be efficient at producing specific health care outputs and inefficient at producing others. For example, Mexico is identified as efficient at producing life expectancy, due to its very low input, but inefficient at preventing mortality. This result emphasizes the value of disaggregated analysis, as a combined analysis would probably mask important specific inefficiencies.

We observe for some countries that financial inputs could be strongly reduced without degrading the current output. The potential increase in output with improvements in lifestyle inputs is much lower. For example, for Canada, we find that there is great potential to reduce financial inputs (GDP per capita, health expenditure per capita) without negatively impacting life expectancy, but a lower potential to reduce lifestyle inputs. Also, for Austria, we observe great potential to reduce medical inputs for mortality, but not to reduce inputs for surgeries.

6 Conclusions

In this study, we investigated the efficiencies of the health care systems of various OECD countries based on OECD health data. We focused on specific aspects of the health care systems and conducted five separate partial analyses investigating the effects of medical inputs on surgery provision, medical inputs on mortality prevention, lifestyle on life expectancy from birth, income and health expenditure per capita on life expectancy from birth, and health expenditure relative to GDP and income equality on life expectancy from birth.

We applied an input-oriented data envelopment analysis (DEA) assuming constant returns to scale. DEA provides optimal input and output weights for each country, maximizing each country's efficiency, given a set of constraints. The obtained optimal weights of the linear problem in

Table 5 Analysis IV: absolute expenditure/income and life expectancy

	Hspend	Gdp	Lifexp	Rel. inp	Rel.outp.	Eff.	Rel.prod.
Austria	4.528	44.873	81.000	0.496	0.974	0.408	0.404
Belgium	4.225	41.397	80.450	0.457	0.968	0.439	0.435
Canada	4.304	42.128	81.450	0.465	0.980	0.437	0.432
Chile	1.476	21.412	78.700	0.234	0.946	0.831	0.829
Czech Republic	2.021	28.643	78.150	0.314	0.940	0.617	0.615
Denmark	4.512	43.560	80.100	0.481	0.963	0.416	0.411
Estonia	1.341	24.735	76.450	0.270	0.919	0.725	0.700
Finland	3.403	40.209	80.700	0.442	0.971	0.454	0.451
France	4.045	37.456	82.050	0.414	0.987	0.495	0.489
Germany	4.693	43.523	80.950	0.482	0.974	0.420	0.415
Hungary	1.697	22.494	75.150	0.247	0.904	0.755	0.752
Ireland	3.663	45.242	80.950	0.497	0.974	0.404	0.402
Israel	2.127	31.629	81.750	0.346	0.983	0.584	0.583
Italy	3.137	35.525	82.300	0.391	0.990	0.524	0.520
Japan	3.592	35.611	83.150	0.393	1.000	0.528	0.522
Korea	2.142	32.022	81.250	0.350	0.977	0.574	0.573
Luxembourg	4.371	91.850	81.450	1.000	0.980	0.222	0.201
Mexico	1.026	16.808	74.350	0.184	0.894	1.000	1.000
Netherlands	5.081	46.054	81.150	0.510	0.976	0.398	0.393
Norway	5.823	66.363	81.500	0.730	0.980	0.278	0.276
Poland	1.448	23.152	76.850	0.253	0.924	0.750	0.750
Portugal	2.523	27.000	80.450	0.298	0.968	0.674	0.668
Slovak Republic	1.977	25.718	76.200	0.282	0.916	0.670	0.667
Slovenia	2.483	28.450	80.200	0.313	0.965	0.637	0.633
Spain	2.957	32.770	82.500	0.361	0.992	0.569	0.565
Sweden	4.743	43.869	81.750	0.485	0.983	0.421	0.416
Switzerland	6.140	55.916	82.750	0.619	0.995	0.335	0.330
Turkey	0.890	17.935	74.600	0.195	0.897	1.000	0.943
UK	3.175	37.386	80.950	0.411	0.974	0.489	0.486
USA	8.454	51.496	78.800	0.578	0.948	0.346	0.337
Mean	3.400	37.841	79.935	0.417	0.961	0.547	0.540
Impl. pr.	0.007	0.025	0.029				

Hspend total health expenditure per capita in US dollars based on purchasing power parity, *Gdp* gross domestic product per capita in US dollars based on purchasing power parity, *Lifexp* life expectancy at birth, *Eff.* efficiency, *Prod.* productivity

multiplier form varied strongly among countries, with several input and output prices being 0, pointing to slacks.

Since DEA allows countries to emphasize the importance of specific inputs and outputs, allowing the efficiency to be exaggerated, we also used average prices for aggregation and constructed an alternative measure of relative productivity. Economically, zero weights are dubious, as the DEA scores then rest on dubious assumptions; for instance, that beds and physicians in Denmark and in the UK are costless (analysis I). The measure of relative productivity is therefore more informative if all countries face rather similar input prices. The correlation between DEA score and relative efficiency differed across the five analyses but was generally rather strong.

We observed efficiency scores to vary substantially across the five analyses, with some countries being efficient in one or more analyses and very inefficient in others. Therefore, countries were found to be efficient at producing specific health care outputs but very inefficient at producing others. Furthermore, we found that efficient countries could achieve their high efficiencies in several ways. The rankings of the 34 considered countries, based on the mean of 1–5 efficiency scores, revealed that Iceland was the most efficient country, followed by Turkey and Estonia. The lowest ranking were obtained for Ireland and Germany.

Inefficient countries have to decide whether to focus on input reduction or output improvement to become more efficient. The benchmark country chosen as a reference

Table 6 Analysis V: healthcare expenditure/GDP, Gini, and life expectancy

	Hgdp	Gini	Lifexp	Rel. inp	Rel.outp.	Eff.	Rel.prod.
Austria	0.101	0.276	81.000	0.964	0.979	0.817	0.817
Belgium	0.102	0.268	80.450	0.975	0.972	0.803	0.803
Czech Republic	0.071	0.256	78.150	0.990	0.944	0.810	0.767
Denmark	0.104	0.249	80.100	1.000	0.968	0.779	0.779
Estonia	0.057	0.338	76.450	0.881	0.924	0.917	0.844
Finland	0.085	0.260	80.700	0.985	0.975	0.796	0.796
France	0.108	0.306	82.050	0.924	0.992	0.863	0.863
Germany	0.108	0.289	80.950	0.947	0.978	0.832	0.831
Hungary	0.075	0.289	75.150	0.947	0.908	0.787	0.772
Ireland	0.081	0.304	80.950	0.927	0.978	0.849	0.849
Israel	0.070	0.371	81.750	0.837	0.988	0.950	0.949
Italy	0.088	0.327	82.300	0.896	0.995	0.893	0.893
Korea	0.067	0.307	81.250	0.923	0.982	0.899	0.856
Luxembourg	0.066	0.302	81.450	0.929	0.984	0.900	0.852
Mexico	0.061	0.457	74.350	0.723	0.898	1.000	1.000
Netherlands	0.110	0.281	81.150	0.957	0.981	0.824	0.824
Norway	0.088	0.253	81.500	0.995	0.985	0.797	0.797
Poland	0.063	0.298	76.850	0.934	0.929	0.860	0.800
Portugal	0.093	0.338	80.450	0.882	0.972	0.888	0.887
Slovak Republic	0.077	0.250	76.200	0.998	0.921	0.765	0.742
Slovenia	0.087	0.250	80.200	0.999	0.969	0.781	0.781
Spain	0.090	0.335	82.500	0.885	0.997	0.906	0.906
Sweden	0.108	0.274	81.750	0.967	0.988	0.822	0.822
Switzerland	0.110	0.285	82.750	0.952	1.000	0.845	0.845
Turkey	0.050	0.402	74.600	0.796	0.902	1.000	0.911
UK	0.085	0.351	80.950	0.864	0.978	0.911	0.911
USA	0.164	0.390	78.800	0.813	0.952	0.943	0.942
Mean	0.088	0.308	79.806	0.922	0.964	0.861	0.846
Impl. pr.	0.011	1.548	1.305				

Hgdp health expenditure as share of gross domestic product, *Gini* Gini coefficient, *Lifexp* life expectancy at birth, *Eff.* efficiency, *Prod.* productivity

when making this decision will depend on the preferences of the country that is making this choice. Our individual analyses allowed us to identify, for each country, which of the outputs/inputs had the greatest potential to be increased/reduced to obtain efficiency gains.

In analysis I (medical inputs and surgeries), we found that Germany was inefficient at producing surgeries. A closer look at inputs and outputs for Germany revealed that that its utilization of about 78% of the maximum input yields 90% of the maximum output (as produced by the efficient country of Belgium). As Germany is a rather rich country, it would probably be more logical to increase output to the highest observed level rather than reducing inputs. A comparison with Belgium, the relevant benchmark country, reveals that Germany produces far fewer kidney transplantations and cataract surgeries. Therefore, more efficient organization of transplantations would increase Germany's health care system efficiency in this specific aspect. According to this

reasoning, Spain—which also belongs to the reference set—would probably be less suitable for use as a benchmark due to its comparatively low output.

On the other hand, Denmark is the country with the highest sum of weighted inputs in most analyses, so it could focus on reducing inputs rather than on further increasing its already high output. Looking at the inputs reveals that there is great potential to reduce the number of practicing nurses, which is very high compared to all other countries. But, of course, the route to improving Danish health care efficiency depends on the preferences of the Danish population or its representatives.

Again using Germany as example, analysis II (medical inputs and mortality) revealed that there is high potential to increase outputs in order to become efficient. Using Japan—the country with the highest sum of outputs—as a benchmark, Germany has considerable potential to reduce its 30-day mortality after admission to hospital for ischemic stroke. But, as Germany is also the country with

the second highest use of inputs, there is also great potential for input reduction (particularly beds and nurses, which are well above average). Austria, which has very similar efficiency measures and similar relative inputs and outputs to Germany, has the greatest potential to reduce beds and physicians to increase efficiency.

In analysis III (lifestyle and mortality), we observed that Japan was the most inefficient country despite presenting the highest output. Japan uses a very high amount of weighted inputs, which, in the case of "bad" inputs, means that there is low alcohol and tobacco consumption as well as low obesity in Japan. As increasing these inputs is not realistic policy advice, Japan would have to increase its average life expectancy—which is already the highest in the world—to become more efficient. But the fact that many countries with rather unhealthy lifestyles have above-average live expectancies (e.g., France and the UK) indicates that lifestyle is a relatively minor determinant (among many others) of life expectancy.

The results of analysis IV (absolute expenditure/income and life expectancy) hinted strongly that absolute income and health spending per capita have no proportional effect on life expectancy. Some high-income and high-spending countries have extremely low efficiency scores (Luxembourg, Norway, Switzerland, and the USA). For instance, life expectancy in Switzerland is 99.5% of the highest observed value (for Japan), but the efficiency score for Switzerland is only 0.335. As reducing GDP is not a meaningful option for policymakers to increase the efficiency of the health care system, there is probably only some potential to decrease health expenditure. Similar arguments hold for analysis V (relative spending/inequality and life expectancy), as we again observed only very moderate effects of relative health care spending and equality on life expectancy.

Summing up, we found that analyses I and II hinted much more strongly at possible policy recommendations than analyses III–V did. We also agree with Medeiros and

Schwierz [14], who conclude that for countries with already high life expectancies, focusing on inputs seems to be a more relevant path for policymakers.

We conclude that performing separate analyses of specific aspects of health care provision is a useful approach for shedding light on how a country's health care system works in detail. It provides important information that can be used to pinpoint where policy interventions should be focused to improve specific parts of a country's health care system.

Compliance with Ethical Standards

Appendix

See Tables 7 and 8.

Table 7 Overview: DEA scores

Country	e1	e2	e3	e4	e5	Mean	Rank
Australia		0.846				0.846	13.0
Austria	0.865	0.624		0.408	0.817	0.678	28.0
Belgium	1.000	0.814		0.439	0.803	0.764	19.0
Canada	1.000	1.000	0.710	0.437		0.787	15.0
Chile				0.831		0.831	14.0
Czech Republic	0.834	0.756		0.617	0.810	0.754	22.0
Denmark	0.846	0.771		0.416	0.779	0.703	26.0
Estonia	1.000	0.841	1.000	0.725	0.917	0.897	3.0
Finland	0.762	1.000	0.748	0.454	0.796	0.752	23.0
France			0.922	0.495	0.863	0.760	21.0
Germany	0.658	0.645		0.420	0.832	0.639	33.0

Table 7 continued

Country	e1	e2	e3	e4	e5	Mean	Rank
Hungary	1.000			0.755	0.787	0.847	12.0
Iceland		0.991				0.991	1.0
Ireland				0.404	0.849	0.626	34.0
Israel	0.887	1.000		0.584	0.950	0.855	8.0
Italy				0.524	0.893	0.708	25.0
Japan		0.958	0.580	0.528		0.689	27.0
Korea	1.000	1.000		0.574	0.899	0.868	4.0
Lithuania	0.649					0.649	31.5
Luxembourg	0.815	0.914	1.000	0.222	0.900	0.770	18.0
Mexico		0.403	1.000	1.000	1.000	0.851	10.0
Netherlands			0.738	0.398	0.824	0.653	30.0
New Zealand	0.652	0.950	0.967			0.856	7.0
Norway		0.879	0.643	0.278	0.797	0.649	31.5
Poland	0.935			0.750	0.860	0.848	11.0
Portugal				0.674	0.888	0.781	16.0
Slovak Republic				0.670	0.765	0.718	24.0
Slovenia	0.860	1.000	0.988	0.637	0.781	0.853	9.0
Spain	1.000	0.955		0.569	0.906	0.857	6.0
Sweden	1.000	1.000	0.626	0.421	0.822	0.774	17.0
Switzerland		0.774	0.732	0.335	0.845	0.672	29.0
Turkey			0.887	1.000	1.000	0.962	2.0
UK	0.942	0.952	1.000	0.489	0.911	0.859	5.0
USA			1.000	0.346	0.943	0.763	20.0

Table 8 Literature overview: OECD data

Paper	Data	Inputs	Outputs	Methods	Results
Retzlaff-Roberts et al. [1]	OECD health data (2000), 27 countries	Environment: school expectancy, Gini coefficient, tobacco use; healthcare: practicing physicians, inpatient beds, MRI, share of GDP allocated to health care	Infant mortality, life expectancy at birth	DEA (input- and output-oriented, VRS), each output individually, both outputs	13 countries are efficient in relation to both outputs (Australia, Canada, France, Greece, Ireland, Japan, Korea, Mexico, Norway, Spain, Sweden, Turkey, UK), 8 countries are inefficient in relation to both outputs (Austria, Belgium, Germany, Hungary, Netherlands, New Zealand, Switzerland, USA), a country can be technically efficient/inefficient at any level of health outcome
Bhat [7]	OECD health data (2002, 2003), 24 countries	Practicing physicians, nurses, inpatient beds, pharmaceuticals	Population aged 0-19years; population aged 20-64 years; population aged 65 years and older	DEA (CRS)	8 efficient countries (Denmark, Japan, Netherlands, Norway, Portugal, Sweden, Turkey, UK), lowest efficiency: Belgium, Iceland, Australia

Table 8 continued

Paper	Data	Inputs	Outputs	Methods	Results
Afonso and Aubyn [8]	OECD health data (2002), 24 countries	Number of doctors, nurses, inpatient beds	Infant survival rate, life expectancy	Free disposable hull (FDH), DEA (input- and output-oriented, VRS, CRS)	Under FDH and DEA: 8 countries are efficient (Canada, Japan, Korea, Portugal, Spain, Sweden, UK, USA)
Afonso and Aubyn [9]	OECD health data (2005), 19 (21) countries	Number of practicing physicians, nurses, acute care beds (per 1000), MRI; (nondiscretionary: GDP per capita, educational level, obesity, tobacco consumption)	Infant survival rate, life expectancy, potential years of life not lost	Two stages: DEA (output-oriented, VRS), (truncated) regression analysis (bootstrapping), PCA	7 efficient countries (Canada, Finland, Japan, Korea, Spain, Sweden, USA), a 40% increase in output is possible on average
Adang [10]	OECD health data, 1995–2000/ 2002, Eurobarometer survey	Share of GDP allocated to health care, number of practicing physicians, tobacco use	Life expectancy at birth, infant mortality	DEA (output-oriented, CRS), Malmquist index	No association between economic performance of health care system and change in satisfaction with health care system
Spinks and Hollingsworth [11]	OECD health data (2004), 1995, 2000, 28 countries; WHO data	School expectancy years, unemployment rate, GDP per capita, health expenditure per capita	Life expectancy at birth	DEA (input-oriented, BCC), Malmquist index	1995: 6 efficient countries (Turkey, Mexico, Korea, Greece, Spain, Japan); 2000: 8 efficient countries (Turkey, Mexico, Korea, Greece, Spain, Japan, Iceland, Switzerland)

References

1. Retzlaff-Roberts D, Chang CF, Rubin RM. Technical efficiency in the use of health care resources: a comparison of OECD countries. Health Policy. 2004;69:55–72.
2. WHO. The world health report—health system financing. The path to universal coverage. Geneva: World Health Organization; 2010.
3. Evans DB, Tandon A, Murray CJL, Lauer JA. Comparative efficiency of national health systems: cross national econometric analysis. BMJ. 2001;323:307–10.
4. Gravelle H, Jacobs R, Jones AM, Street A. Comparing the efficiency of national health systems: a sensitivity analysis of the WHO approach. Appl Health Econ Health Policy. 2003;2(3):141–7.
5. Hollingsworth B, Wildman J. The efficiency of health production: re-estimating the who panel data using parametric and non-parametric approaches to provide additional information. Health Econ. 2003;12:493–504.
6. Greene W. Distinguishing between heterogeneity and inefficiency: stochastic frontier analysis of the World Health Organization's panel data on national health care systems. Health Econ. 2004;13:959–80.
7. Bhat VN. Institutional arrangements and efficiency of health care delivery systems. Eur J Health Econ. 2005;6(3):215–21.
8. Afonso A, Aubyn MS. Non-parametric approaches to education and health efficiency in OECD countries. J Appl Econ. 2005;8:227–46.
9. Afonso A, Aubyn MS. Relative efficiency of health provision: a DEA approach with non-discretionary inputs. Working Papers of the Department of Economics at ISEG, vol. 33. Lisbon: Lisboa School of Economics and Management; 2006.
10. Adang EMM, Borm GF. Is there an association between economic performance and public satisfaction in health care? Eur J Health Econ. 2007;8(3):279–85.
11. Spinks J, Hollingsworth B. Cross-country comparisons of technical efficiency of health production: a demonstration of pitfalls. Appl Econ. 2009;41:417–27.
12. Hsu YC. The efficiency of government spending on health: evidence from Europe and Central Asia. Soc Sci J. 2013;50:665–73.
13. Hsu YC. Efficiency in government health spending: a super slacks-based model. Qual Quant. 2014;48(1):111–26.
14. Medeiros J, Schwierz C. Efficiency estimates of health care systems. Eur Econ Econ Pap. 2015;549:1–54.
15. OECD. OECD health statistics 2015: definitions, sources and methods. Paris: OECD Publishing; 2015.
16. OECD. Health at a glance 2015: OECD indicators. Paris: OECD Publishing; 2015.
17. Häkkinen U, Joumard I. Cross-country analysis of efficiency in OECD health care sectors. OECD Economics Department Working Paper 554. Paris: OECD Publishing; 2007
18. Smith PC, Mossialos E, Papanicolas I. Performance measurement for health system improvement: experiences, challenges and prospects. Technical report. Geneva: World Health Organization; 2008.
19. Charnes A, Cooper WW, Rhodes E. Measuring the efficiency of decision making units. Eur J Oper Res. 1978;2:429–44.
20. Charnes A, Cooper WW, Wei QL, Huang ZM. Cone ratio data envelopment analysis and multi-objective programming. Int J Syst Sci. 1989;20(7):1099–118.
21. Allen R, Athanassopoulos A, Dyson R, Thanassoulis E. Weights restrictions and value judgements in data envelopment analysis: evolution, development and future directions. Ann Oper Res. 1997;73:13–34.
22. Lauer JA, Lovell CAK, Murray CJL, Evans DB. World health system performance revisited: the impact of varying the relative

importance of health system goals. BMC Health Serv Res. 2004;4(19).

23. Hauck K, Street A. Performance assessment in the context of multiple objectives: a multivariate multilevel analysis. J Health Econ. 2006;25:1029–48.

24. Cheng G, Zervopoulos PD. Estimating the technical efficiency of health care systems: a cross-country comparison using the directional distance function. Eur J Oper Res. 2014;238:899–910.

25. Seiford LM, Zhu J. Modeling undesirable factors in efficiency evaluation. Eur J Oper Res. 2002;42:16–20.