Productive Efficiency and Heterogeneity of Health Care Systems: Results of a Measurement for OECD Countries

Peter Kotzian[*]

Technische Universität Darmstadt, Institut für Politikwissenschaft, Residenzschloß, D-64283 Darmstadt, Germany

Abstract: Health system efficiency is a major target of health policy but its conceptualization and measurement are still a problem in health economics. Because health status is influenced by many factors outside the health system, I argue that measurements of health system efficiency should focus on the process of turning financial input into additional health output rather than the levels of health status reached. When analyzing levels of health status using regression methods, the appropriate efficiency indicator is hence not a country-specific intercept based on the achieved health status, but a country-specific slope for input factors in the production function of health outcomes. The slopes represent health system efficiency, while the intercepts represent health relevant heterogeneity among countries. Using data on OECD members these slopes are estimated. Countries differ far more in their residual heterogeneity than in the rate by which their health system turns money into life years.

INTRODUCTION

The central aim of a health care system, henceforth abbreviated HCS, is the production of additional health. An HCS can be evaluated in several respects, such as equity, responsiveness, and quality but also regarding its productive efficiency. This latter aspect will be the focus of this paper. The HCS should achieve a sufficient level of health, but it should do so in an efficient way. The question of what achievement level shall be reached is also a political question. It covers aspects such as how much shall be invested in health, what illnesses and social groups shall be covered. Productive efficiency is an undisputed aim of health policy, but conceptualization and measurement of health system efficiency proved to be difficult.

Efficiency evaluations of the HCS typically take into account health expenditure and health levels achieved; see for this notion of technical efficiency [1, 2]. The empirical evaluation of a sample of health systems is usually based on a comparison of an HCS' actual achievement in reference to what the best performing HCS in the sample achieves, controlling for differences in input levels.

Efficiency measurement is subject to problems. The first problem consists of isolating the contribution of the HCS from other factors relevant for the observable health level.

The second problem arises from the fact that efficiency measurement is based on health levels. Following an efficiency concept proposed in particular by Greene [3], I argue that productive efficiency concerns primarily the production process, not the output levels reached. The predominant life style in a country may result in a high or low level of health status. A country may decide to spend much or little on health, and this too will be reflected in the health level. But neither lifestyle nor the political decision to spend much on health indicates whether the production of health is itself efficient or not. Consequently, I will refer to the productive efficiency of a country's HCS as the rate by which the HCS turns financial input into additional health outcomes, independently of the health levels prevailing in that country.

Productive efficiency is measured using production functions, which are basically regression models regressing health output on various input factors. In the existing literature on HCS efficiency measurement, the productive efficiency is based on the health level and statistically estimated by a country-specific intercept. I propose to measure efficiency by a country-specific slope for health expenditure, which will measure whether the country's HCS is using money more productively than HCS in other countries.

The paper proceeds as follows: a brief description will introduce to the general problems of efficiency measurement for health systems, followed by an introduction to the basic strategy of efficiency measurement using the productivity frontier approach and its application to health care. After elaborating the problems inherent to this approach and proposing an alternative, the empirical section will estimate the alternative efficiency indicator for a sample of OECD countries.

GENERAL PROBLEMS OF HEALTH SYSTEM EFFICIENCY MEASUREMENT

Any measurement of HCS efficiency, irrespective of the underlying efficiency notion, requires the measurement of

*Address correspondence to this author at the Technische Universität Darmstadt, Institut für Politikwissenschaft, Residenzschloß, D-64283 Darmstadt, Germany; Tel: +49 (0) 6151/16 6501; Fax: +49 (0) 6151/16 4602; E-mail: kotzian@pg.tu-darmstadt.de

inputs and outputs. Both are subject to measurement problems.

a)   The health output level, e.g., the life expectancy, in a country is strongly influenced by a range of factors: smoking habits, education, and income, but also factors like regulations on vehicle safety. A wide range of empirical studies found consistently that the HCS is only one factor for health status and one of limited impact at that [4-11]. Thus, any evaluation of an HCS based on the health levels reached is confronted with the problem of identifying the specific contribution of the HCS to health status.

b)   Econometric efficiency evaluations only refer to 'biological' health outcomes like life expectancy. But as Mooney argues, HCS in industrialized nations also produce 'beyond-health outcomes' and people are willing to pay for them [12, 13]. A much discussed beyond-health output is responsiveness, e.g., giving patients an active role in medical decision making. Further, Mooney emphasizes that utility arises also from the process by which health is produced. People are willing to pay for being treated in a respectful way or by more convenient therapies. Producing beyond-health outcomes requires resources while not increasing biological health outcomes. An HCS may appear as inefficient in a 'money-invested-life years-gained' sense, because citizens demand more beyond-health outputs. This is not a question of efficiency, but of preferences.

c)   With regard to the input of the HCS, one of the simplest input indicators is health care expenditure, henceforth abbreviated HCE. HCE can be seen as a proxy for many other input factors, such as manpower, pharmaceuticals or medical equipment. This indicator is usually standardized by using the per capita figure and purchasing power parities. But as Anderson *et al.* argue, the relative price of health services may differ substantially among countries, for instance if salaries for medical staff are relatively higher than the overall wage level in one country compared to another country [14]; see also work by Berndt *et al.* and Newhouse for the problems of measuring prices in health care over time and across countries in a valid and comparable way [15-17].

These general problems are relevant, because they affect and possibly bias those efficiency measurements, which are based on the productivity frontier approaches.

EFFICIENCY MEASUREMENT: THE PRODUCTIVITY FRONTIER APPROACH

The two standard approaches of efficiency measurement are data envelopment analysis and productivity frontier approaches. Both are similar, and I will focus on the productivity frontier approach, as the dominant one [3, 18-21].

Productive efficiency measurement is based on input/output-ratios. Using input factors (labor, capital, technological knowledge), a productive system, be it a firm or an HCS, can produce a certain output level. Using data on inputs used and outputs obtained by a sample of productive systems, one can use regression techniques to estimate a production function. Applied to health, this production function allows one to make several statements:

The first is the estimated contribution of each input factor to the output level. For instance, if data on the share of HCE spent on medicines and on medical technology is used in the analysis, the production function would allow statements about the relative contribution of spending on pharmaceuticals compared to spending on medical technology.

Secondly, while the output level increases with the quantity of inputs used, there is an upper limit due to biological restraints, which is fixed in the short run. As this level is approached, investing in health production reveals diminishing returns.

Thirdly, contrary to 'normal' firms there are health outcomes even in the absence of explicit health production. The overall health outcome is not the result of a multiplicative combination of input factors. It does not drop to zero if one input factor is missing, not even if there is no HCS at all.

Fourthly, even after controlling for differences in input levels the cases differ with regard to the health level reached. The core assumption of the productivity frontier approach is that these residual differences in the outcome levels are attributable to the HCS and can be interpreted in terms of efficiency. It may not be clear, what makes the HCS inefficient, maybe the inefficient HCS has a bigger administrative overhead or agents in the HCS extract economic rents. But an inefficient HCS will always produce less output and remain at a lower health level than an efficient one, even if the input level is the same or if the differences in input composition are controlled for.

The most efficient cases with the highest output levels constitute the productivity frontier: the maximum of output achievable for each level of inputs. Efficiency is defined as the distance to this productivity frontier. Most of the empirical literature on productive efficiency is implementing this concept of efficiency as the distance to an overall productivity frontier.

To estimate this distance, empirical applications estimate a case-specific inefficiency term or intercept, using variants of the regression model proposed by Schmidt and Sickles [22]:

$$Y_{it} = a + bX_{it} + v_{it} - u_i \qquad\qquad (1)$$

Respectively

$$Y_{it} = a_i + bX_{it} + v_{it}, \qquad\qquad (2)$$

Y_{it} is the level of health status in country i at period t. The v_{it} term is a random error term, capturing health-relevant random events in country i during the period t. The nonrandom inefficiency component is captured by u_i respectively a_i as the country-specific intercept. Both terms are connected by the relationship $u_i = a^* - a_i$ where a^* is the intercept of the best performing case, i.e., the HCS located on the production frontier. While the production function for

each case has the same form - bX_{it} is identical for all cases - a case's actual health output level depends on its set of input factors X_{it} but also on its intercept, a_i. An inefficient case is located below the overall frontier because its a_i is lower than that of the best case. The content of the country intercept is the distance to a benchmark constituted by the most efficient cases, and thus reflects the difference in efficiency [1, 3, 23].

PROBLEMS OF THE PRODUCTIVITY FRONTIER APPROACH AND AN ALTERNATIVE APPROACH

There are two problems when applying this approach to health production.

a)   The first problem arises from the fact that the level of health is influenced by so many other factors [5, 7-9, 24]. As Greene emphasizes, countries are heterogeneous in many health relevant factors [3]. This heterogeneity will be reflected in a country's level of health status. If not explicitly specified in the regression equation, the effect of all these health relevant factors will be captured by the country-specific intercept, and attributed to the HCS. By using panel data, all features that are constant over time, are put into the country-intercept (a_i) respectively the inefficiency term (u_i) and are interpreted in terms of efficiency. For instance if citizens in a country have a healthy life style, this will be reflected in the country's intercept and will be interpreted as a more efficient HCS because the country is closer to the frontier. This argument can be extended to factors which change over time: assume that a society's demand for beyond-health outputs increases. This increases HCE but not the health level, and this will be interpreted as increasing inefficiency. None of this is related to efficiency. Allowing for changes in the distance to the frontier over time as Hollingsworth and Wildman did, one may easily find spurious changes in efficiency [23]. While it is possible to control for biasing health relevant factors, not all of them are known. For even fewer of them the data necessary to actually do this are available. In particular if for each case data are available for a few points in time only, random events may influence the distance to the frontier, making it hard to distinguish between the error term and systematic inefficiency.

b)   The second, conceptual problem concerns the interpretation of the proximity to a single productivity frontier which was obtained by analyzing all countries in a sample as efficiency. My central point is that what the productive efficiency of an HCS is all about is the production of additional health: for instance, years of life in addition to those a population would have anyway. The HCS as a productive system transforms money into life years. This affects the resulting health level. An HCS's efficiency is not primarily determined by the level it achieves, but by the way it does so. To increase the health level, investing not in the HCS but in an anti-smoking-campaign might be much more efficient in a "value-for-money" sense;

see for example Tengs [25]. But if a society decides to invest in health production by means of the HCS, the interesting question is how much health the society gets in return for this investment. This depends on how productive the HCS uses the money. In statistical terms this "efficient" usage is proxied by the b one estimates for HCE in a regression of a health status indicator on HCE using data for an HCS. Instead of estimating the distance of each case to a common production function, one has to estimate different production functions for different cases.

This idea was raised for instance by Tsionas and Huang, who also elaborate and apply this approach to industry productivity [26, 27]. An application to health is Greene who raises the distinction between heterogeneity and efficiency, which are mixed up if intercept approaches are used [3]. Seeing the problem he proposes two solutions. First, the estimation of a country-specific disturbance term, which is to be added to an overall effect of an independent variable using hierarchical or multilevel models. Secondly, a latent class model, in which cases is assigned to classes, where the coefficients of an independent variable are identical for each case in a class, but differ among classes. Problems in his application arise from using the WHO panel data set. Because there are only few points in time per country, most of the variation occurs between countries but only little within each country.

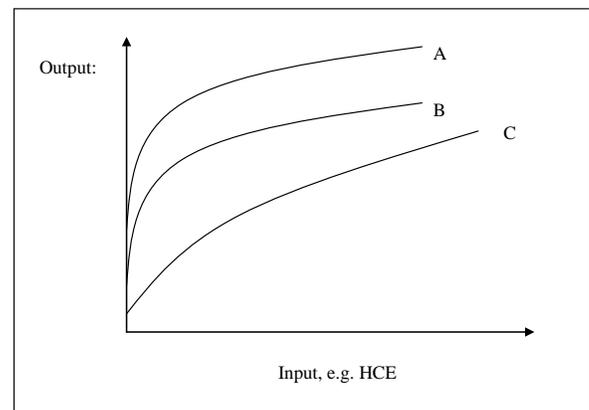The distinction among intercepts and slopes as efficiency indicators are elaborated in Fig. (**1**) below.

Fig. (1). HCS efficiency: intercept *vs* slope as measure of efficiency.

Assume that the three curves in Fig. (**1**) represent the increase in life expectancy, which is observable in each of the three countries as HCE increases over time.

Let country A is the country achieving the maximum health level that is currently possible for biological reasons. Country A's HCS would be the most efficient according to the intercept approach to efficiency. Its development over time basically describes the production frontier: as HCE increases, so does life expectancy, and in country A the health outcome obtained for a given level of HCE is always the highest. Country B would be less efficient, since it achieves constantly lower health levels. If it increases HCE over time, it will follow a curve which has the identical form

but is shifted downwards by the country-specific intercept. In the terms of Greene, countries A and B are heterogeneous in terms of health relevant factors, and health levels differ. But their HCS are of equal efficiency, because the rate at which they turn money into life years is equal. Both HCS have an identically shaped production function: for each additional dollar invested in health, both HCS produce the same increase in health output. The production function is only shifted parallel by different intercepts. The fact that A and B operate at different levels might be due to other reasons, e.g., life style.

The production function of country C is different. Country C starts at a lower health level for reasons unknown. But C's HCS is more efficient: for an increase in HCE it produces more additional life years than either A or B. Its production function is steeper. In particular, it is steeper in the later stages, which would cover the period of industrialization, where the improvements in health status obtainable from investing in health diminish. In statistical terms the slope coefficient b estimated for HCE in the production function of health using data from C is larger than the slope estimated using data from either A or B.

Estimating a uniform production function, an identical slope coefficient b for HCE for all three HCS, one forces the efficiency of turning HCE into additional health output to be constant. Based on this argument, the appropriate measure for country it's HCS efficiency is not the a_i, but the b_i, a country-specific slope of HCE.

DATA, ESTIMATION AND RESULTS

I will use the OECD Health Database to estimate these country-specific slopes as efficiency indicator [28]. Just like in the intercept-based approaches, the estimates of slopes must be controlled for health relevant variables, in particular changes thereof, to avoid attributing the impact of these to the HCS. Since I am interested only in the magnitude of the slope coefficients, I will use a conventional OLS approach. An alternative method would be a statistical model which takes into account the multilevel structure inherent to the data by simultaneously estimating country-specific slopes and modeling the system level factors influencing these coefficients [29, 30].

To estimate the country-specific slopes I regress life expectancy on production factors, which were found to be relevant in previous studies on life expectancy. I constructed a panel data set using the OECD Health Database. To estimate the health production function of an individual country, information on health levels and HCE for several points in time is needed to track how the health level reacted to changes in HCE. Further, to preclude omitted variable bias, data on health relevant factors are required. Contrary to a larger sample like all WHO members, such information about several health-relevant factors outside of the HCS is available for the OECD countries.

There are several open issues in analyzing health levels, which have created a substantial literature, most notably the GDP-health and the GDP-HCE relationship. In the frame of this paper, I will focus only on the issue at hand: the contribution of HCE to health status in different countries. I will not go into a detailed discussion of the effect of control variables.

Variables and Models

The produced outcome and dependent variable of the regression is the life expectancy at birth, as the unweighted average for both sexes, not adjusted for quality of life or disabilities, denoted LE; see the documentation in the OECD Health Data for details on calculation [28].

The list of control variables encompasses the factors found to be relevant in the literature on life expectancy: income, education, and lifestyle [4]:

Education: The positive effect of education on life expectancy has much supportive evidence. The underlying mechanisms are that more educated people are better aware of behavior conducive or detrimental to health, and adapt their behavior accordingly [5, 8]. While literacy is a good indicator to compare educational levels in developing and industrialized countries, the variation in literacy among OECD countries is only small. To capture differences in educational achievement at an already high level, I have chosen the attainment with regard to the ISCED 5A/6 levels, i.e., university degrees and Ph.D. programs. The percentage of the population that has achieved this educational level is used as indicator for education.

GDP: Wealth, measured here as GDP per capita, is a strong proxy for many health-related living conditions like diet, hygiene, housing etc. [9, 11, 24]. The relationship between GDP and life expectancy is characterized by strongly diminishing effects, an effect that can be found even for the selection of advanced industrialized societies included in the analysis. To fit in this curvilinear relationship, the log of GDP was used.

Over 65: This demographic variable is the percentage of the population that is 65 year old or older. An aging population is a driver of HCE, albeit the mechanisms are disputed [31, 32]. HCE increases if the population gets older, but this is unrelated to HCS efficiency. I interpret it as a factor in capturing the impact of the health levels already achieved: it is for medical reasons more expensive to increase life expectancy from 78 to 79 years, than to increase it from 40 to 41 years. So if there are already many old people in the population, it is more expensive to produce additional life years for them than it would be if the average age of the population were lower.

Alcohol: As a first lifestyle variable, alcohol consumption in liters per head for the population aged more than 15 years was included.

Tobacco: Tobacco consumption as a second lifestyle variable is measured as a percentage of the population which are daily smokers.

Both indicators cover the issue, whether the HCS is seemingly underperforming, while it is actually struggling with offsetting the effects of unhealthy lifestyles.

Coverage: HCE can be used to intensify the care provided for a small group or to expand coverage to people that are up to now without access to health care. In the first case, additional HCE has less effect on life expectancy, since the recipients of care are already well supplied and the diminishing returns arguments holds. In the second case, HCE invested might yield more health output by providing services to people who had no access before. The dataset contains two proxies for this: first, the percentage of the population with coverage for inpatient treatment (**cov_inpat**). The second proxy used is the percentage of the population with coverage for outpatient care, i.e., primary care treatment (**cov_outpat**). A problem is that the data is not paying heed to the actual access, which might be limited by factual availability and waiting lists.

The production factor which is of the most interest is expenditure for health, measured in total health expenditure in 1000 international Dollars per head. The absolute figure was used because if one uses the HCE/GDP-ratio, one has not only the financial input but also the economic development as an explanatory variable, which is less valid. GDP is subject to substantial fluctuations during booms and recessions. While the bivariate relationship among HCE and life expectancy is characterized by diminishing returns -just like the GDP/life expectancy relationship - the effect of HCE on life expectancy was in supplementary diagnostics found to be linear, once GDP and the other independent variables are controlled for. HCE is thus used in natural units, presuming a linear additive contribution in health production.

Regarding the specification of the regression equation, a much debated issue is the GDP-HCE relationship: is this relationship actually existing or spurious? If existing, what is the direction? But also, which of both actually influences life expectancy? [33-36]. There is however consensus that when analyzing health status, both should be included to avoid that the effect of both factors are merged [37- 39]. In the models estimated here, I found in consistence with the literature reviewed above that including both simultaneously results in a strong positive effect of GDP and a small but still positive overall effect for HCE. This also holds true if only the share of GDP not spent for health is used as an indicator. Excluding GDP from the model did not lead to a change in the other variables, since HCE takes over the role of GDP.

With regard to the impact of producing beyond-health outcomes in addition to health outcomes, it would be interesting to include an indicator thereof, e.g., the WHO's responsiveness measure [40, 41]. Regrettably, this information is available for one year only and cannot be extrapolated.

The availability of data varies among countries and availability is usually better for west European countries and the US. Hungary, Korea, Mexico, the Slovak Republic and Turkey were excluded altogether, since most data were unavailable. The remaining dataset is an unbalanced panel, encompassing about 18 data points for most countries in the full model. The effective time period covered is 1984 till 2003. With reference to Fig. (**1**), one can say that the countries included have left the section of the production function, where the slope is very steep. Being industrialized, they

entered a period, where the production function is quite flat because additional improvements in health come at a higher price and are of smaller magnitude.

Modeling and Estimation

The aims of this paper are twofold. First, to describe the efficiency of a country's HCS in producing health, defined as the rate by which it turns HCE into life years, controlling for other factors relevant for health status. Secondly, to capture residual heterogeneity among the countries with regard to the health outcome level achieved; i.e., unknown features of the country, which influences the level of health outcome, but are not due to differences in income, life style, health expenditure etc. [3, 42]. The effect of the control variables per se is not of interest here.

Regarding the specification of the models, HCE per se has an effect on life expectancy, which is common to all cases. But the question is, whether a country achieves more with the money invested for health production than the average or a reference country? As a first variant, I used the dummy-interaction specification proposed by Steenbergen and Jones [30]. I generated a multiplicative interaction variable between HCE and the country dummy. This interaction variable captures the impact of HCE in a particular country. While HCE has a certain effect, the effect of HCE spent in this particular country may differ. To capture this, I estimated the following equation:

$$LE_{it} = a + b(X) + bHCE + b_{1i} HCE*Country + b_{2i} Country \quad (3)$$

where X is the set of control variables, selected on the basis of the literature on determinants of health levels reviewed above: GDP, lifestyle factors, demographics, education and coverage. According to the argument, remaining health-relevant heterogeneity among countries will be captured by the country dummy variables, the b_{2i} coefficients obtained for all countries. The country's efficiency will be captured by the country-specific slope, b_{1i}. The reference category, both for the country dummies and the country slopes, is Norway.

As a second variant, to contrast the slope approach to the standard approach of efficiency as the distance to a single productivity frontier, the same analysis was conducted but including only country dummies in the regression model.

With regard to the functional form of the production function two alternative forms were tested. The idea underlying models using logarithms of the variables is that health production is best described by a Cobb-Douglas-function. This assumes a multiplicative combination of all production factors and implies zero output if any input factor is zero. Given that there is health outcome even in the absence of an HCS, this might not be the appropriate functional form. The health level might be rather the sum than the product of factors, suggesting a linear-additive-model.

Regarding the necessity of transforming variables, I found that only GDP has a curvilinear relationship with life expectancy. The other independent variables, including HCE, have a linear bivariate relationship with life expectancy, once GDP is taken into account. I hence used a mix-

ture, using the logarithm of GDP, but linear additive effects and natural units for the remaining independent variables. As a second variant a model using the logarithms of all variables was estimated. The results obtained do not differ substantially.

Since basically only the additional contribution of HCE to the health level is of interest, focus is put on the linear additive model.

A further question of interest is, whether and how the country intercept/dummy and the country slope respond to changes in the selection of control variables. The argument was that the impact of omitted control variables is included in the country effects, biasing them. The full model, including all available control variables, used gdp, education, alcohol, tobacco, over65, HCE, **cov_inpat**, and **cov_outpat**. The reduced model includes only gdp, education, and HCE.

Results

The estimates of HCS heterogeneity and productive efficiency obtained using the different approaches are given in Tables **1** and **2** below. The models given in Table **1** use a linear-additive production function, while those reported in Table **2** use a log-model, implying a multiplicative production function. The estimation was done using four different model variants. Model 1 tests the full model, including country slopes and intercepts and the full set of control variables, in particular variables capturing country characteristics such as lifestyle. Model 2 omits all but the most common control variables, and the idea is to compare, how the coefficients obtained for the country-slopes and country dummies change. Model 3 omits the country slopes, and model 4 omits country slopes and most of the control variables. Again, the idea is to check, whether the coefficients obtained when omitting crucial variables become biased.

I interpret briefly the fixed coefficients of the production function, which are identical for all cases, and then I turn to the country-specific effects.

Looking at the obtained overall production function the findings for the control variables are consistent with the research cited previously: GDP is a strong factor for life expectancy. Its effect remains approximately constant in all models. For the sample of advanced industrialized countries education is no longer of significant impact on the health level, indicating that the health-increasing effect of education already took place at lower levels of overall development.

More important are the lifestyle factors tobacco and alcohol, which are significant and have the excepted directions. This is also true for the demographic structure. Inpatient coverage, which shows a larger variation than outpatient coverage, affects the level of life expectancy significantly: the coefficient found is negative; indicating that efforts used to expand coverage cannot be invested in the production of life years. HCE has an increasing impact on life expectancy even if GDP is controlled for. Its effect is roughly stable with regard to the selection of control variables capturing lifestyle and demographics.

Table 1. **Heterogeneity and Efficiency in Health Production: Linear Model**

Model	Linear 1	Linear 2	Linear 3	Linear 4
log_gdp	2.193***	2.195***	3.248***	3.092***
education	0.002	0.009	0.030***	0.026**
alcohol	-0.195***		-0.259***	
tobacco	-0.029**		-0.005	
over65	0.182***		0.011	
hce1000	0.544***	0.424**	0.384***	0.572***
cov_inpat	-0.138**		-0.340***	
cov_outpat	-0.037		0.007	
Country Slopes				
hce_daul	0.259	0.983***		
hce_daus	1.250***	1.721***		
hce_dbel	-0.183	0.550***		
hce_dcan	-0.293	0.592***		
hce_dcze	3.639***	4.174***		
hce_dden	-0.041	0.543**		
hce_dfin	1.674***	2.322***		
hce_dfra	0.008	0.904***		
hce_dger	1.270***	1.652***		
hce_dgre	-1.690***	-0.366		
hce_dice	0.696**	0.689**		
hce_dire	-0.057	-0.108		
hce_dita	0.002	2.208***		
hce_djap	-0.243	1.232***		
hce_dlux	0.408*	0.589**		
hce_dnet	-0.243	-0.464*		
hce_dnz	1.782***	2.315***		
hce_dpor	-0.07	1.122***		
hce_dspa	-1.205**	0.633**		
hce_dswe	0.827***	1.061***		
hce_dswi	0.797***	0.958***		
hce_duk	0.656***	0.671***		
hce_dusa	-0.336*	-0.125		
Country Intercepts				
oecdd_aul	1.746***	-0.992**	2.128***	0.909***
oecdd_aus	-1.707***	-3.625***	0.843***	-0.434**
oecdd_bel	0.496	-1.476***	0.874***	-0.271
oecdd_can	2.322***	-0.605	1.419***	0.630***
oecdd_cze	-3.464***	-5.523***	0.546*	-1.026***
oecdd_den	-0.135	-2.738***	0.178	-1.542***
oecdd_fin	-2.448***	-4.232***	0.376*	-0.457**
oecdd_fra	1.549**	-1.385***	2.109***	0.595***

(Table 1) contd…..

Model	Linear 1	Linear 2	Linear 3	Linear 4
Country Intercepts				
oecdd_ger	-4.466***	-4.500***	-2.311***	-0.815***
oecdd_gre	4.788***	1.913***	3.370***	2.088***
oecdd_ice	0.426	-0.166	1.408***	1.270***
oecdd_ire	-1.351	-0.854**	1.616	-0.542***
oecdd_ita	1.44	-2.711***	2.285***	1.219***
oecdd_jap	4.053***	0.825**	3.615***	2.823***
oecdd_lux	-0.902	-3.229***	-0.095	-2.260***
oecdd_net	-4.130*	0.963**	-11.130***	0.174
oecdd_nz	-0.627	-2.847***	1.557***	0.494**
oecdd_por	0.989	-1.929***	2.336***	0.106
oecdd_spa	4.256***	0.820*	3.582***	2.098***
oecdd_swe	-0.651	-0.677	1.630***	1.437***
oecdd_swi	-0.432	-1.900***	2.136***	0.428**
oecdd_uk	-0.461	-1.343***	0.985***	-0.12
oecdd_usa	-13.917***	-2.363***	-27.284***	-3.338***
constant	71.099***	54.621***	78.499***	45.049***
Adj. R^2	.98	.98	.96	.94
N	386	408	386	408

Of interest are the country-specific effects, which represent differences in heterogeneity and efficiency among the countries relative to Norway. Many country-specific effects are significant in the full model (*p<0.05; **p<0.01; ***p<0.001). Even more of them become significant in the models which do not take into account other health-relevant features. All country-specific effects react strongly to the omission of control variables.

The entries oecdd_aul (for Australia) to oecdd_us (for USA) are the country dummies, the country-specific intercepts. They capture the unexplained heterogeneity among countries, and become more significant, if control features are excluded. At the same time, the range of heterogeneity shrinks, if control variables are left out. This is mostly due to some countries, most notably the US and the Netherlands, which stand out only if other health-relevant features are taken into account. Their health level is lower than one would expect, and it is even lower when one takes into account all the factors which should actually increase their health level.

The entries such as hce_daul (for Australia) are the country-HCE interaction effects, i.e., the country-specific slopes for HCE. These slopes differ from country to country, indicating significant differences in HCS efficiency. Some HCS are using the money more productively than others. The Czech Republic and Finland have the most productive, Greece and Spain the least productive HCS. The slopes react strongly to the choice of control variables for some countries

Table 2.   Heterogeneity and Efficiency in Health Production: Log-Production Function

Model	log1	log2	log3	log4
log_gdp	0.034***	0.038***	0.036***	0.025***
log_education	0	0.002	0.003	0.002
log_alco	-0.022***		-0.025***	
log_tobac	-0.012***		-0.006	
log_over65	0.047***		-0.004	
log_hce	0.012*	0.003	0.015***	0.025***
log_cov_in	-0.073*		-0.189***	
log_cov_out	0.002		0.053**	
Country Slopes				
loghce_aul	-0.003	0.020***		
loghce_aus	0.020***	0.032***		
loghce_bel	-0.010*	0.011**		
loghce_can	-0.018**	0.016***		
loghce_cze	0.028***	0.041***		
loghce_den	-0.003	0.012*		
loghce_fin	0.020***	0.039***		
loghce_fra	-0.005	0.020***		
loghce_ger	0.036***	0.046***		
loghce_gre	-0.034***	-0.007		
loghce_ice	0.022**	0.022**		
loghce_ire	-0.009	-0.007*		
loghce_ita	-0.013	0.041***		
loghce_jap	-0.022**	0.016***		
loghce_lux	0.01	0.020**		
loghce_net	-0.017***	-0.014**		
loghce_nz	0.025***	0.037***		
loghce_por	-0.011*	0.010**		
loghce_spa	-0.028***	0.002		
loghce_swe	0.013**	0.025***		
loghce_swi	0.030***	0.032***		
loghce_uk	0.005	0.010*		
loghce_usa	-0.007	0.004		
Country Intercepts				
oecdd_aul	0.037***	-0.001	0.025***	0.011***
oecdd_aus	-0.002	-0.024***	0.011**	-0.006**
oecdd_bel	0.014**	-0.012***	0.015***	-0.005*
oecdd_can	0.041***	-0.002	0.017***	0.007***
oecdd_cze	0.008	-0.016***	0.005	-0.012***
oecdd_den	0	-0.029***	0	-0.022***
oecdd_fin	-0.003	-0.025***	0.004	-0.006**
oecdd_fra	0.026***	-0.006*	0.025***	0.005*

(Table 2) contd…..

Model	log1	log2	log3	log4
Country Intercepts				
oecdd_ger	-0.033***	-0.045***	-0.006	-0.014***
oecdd_gre	0.046***	0.023***	0.044***	0.025***
oecdd_ice	0.013	0	0.016***	0.016***
oecdd_ire	0.025	-0.011***	0.068***	-0.005**
oecdd_ita	0.026***	-0.007*	0.028***	0.014***
oecdd_jap	0.060***	0.028***	0.049***	0.038***
oecdd_lux	-0.008	-0.043***	-0.004	-0.025***
oecdd_net	0.003	0.010**	-0.052***	0.002
oecdd_nz	0.025***	-0.006*	0.020***	0.008***
oecdd_por	0.015*	-0.006**	0.024***	0.001
oecdd_spa	0.046***	0.021***	0.049***	0.029***
oecdd_swe	0.003	0.002	0.020***	0.017***
oecdd_swi	-0.005	-0.021***	0.025***	0.004
oecdd_uk	0.008*	-0.008***	0.014***	0
oecdd_usa	-0.106***	-0.036***	-0.217***	-0.042***
constant	4.277***	3.967***	4.666***	4.073***
Adj. R^2	.98	.98	.95	.93
N	386	408	386	408

only. In most cases the exclusion of health-relevant variables which change over time, e.g., life style, systematically biases the country-specific slopes for HCE, which are then overestimated in both magnitude and significance.

As was to be expected, country slopes and the country dummies obtained correlate negatively: to overestimate the slope leads to an underestimated intercept. The functional form chosen, i.e., linear *vs* logarithmic, influences the results only slightly, but does not change the basic findings.

To interpret the results, Fig. (**2**) is the real life equivalent to Fig. (**1**), illustrating the results for selected cases. It shows the change in life expectancy if HCE is increased by 1000$, using the coefficients for HCE, the country-HCE interaction and country dummies which are reported for the full linear model; cf. column 1 in Table **1**. An increase of 1000 $ in HCE roughly represents the change which occurred in most countries in the sample during the period from the mid 1990s to the mid 2000s. The figure indicates the productivity of the HCS in that period, after controlling for changes in other health-relevant factors.

The country-intercepts represent the country-specific heterogeneity in the level of life expectancy, which is not due to controlled factors like GDP, education, or lifestyle.

The country slopes are the total change in life expectancy for a country, i.e., b (HCE) + b (HCE*Country), and represent country-specific productive efficiency in the usage of HCE to produce additional life years.

The intercepts as well as the changes are relative to Norway, which is the reference case. If a country has a steeper total slope, i.e., b (HCE*Country) > 0, it is more efficient in producing health than Norway. If the country-specific slope is zero, it is of equal efficiency as Norway. If it has a lower intercept than Norway, there are features unknown which cause life expectancy to be lower in this country than in Norway.

Looking at Fig. (**2**), the most striking feature is that differences in life expectancy levels are mainly due to residual heterogeneity. This heterogeneity remains, even if several features of potential relevance for the health level are controlled. The production of life expectancy using HCE does increase the former, but the variation in productivity among countries is limited and does not usually change the ranking of the countries. There is no country, where the "default" life expectancy is lower than in another country, but where the HCS is so efficient in producing additional life years, that the ranking is inverted.

Table 3.   **Comparison of Efficiency Estimates**

	SlopeLinear	SlopeLog	InterceptLinear	InterceptLog	GrossEfficiency	WHOScore1	WHOScore2
SlopeLinear	1.000	0.851	0.076	-0.012	-0.502	-0.552	-0.515
SlopeLog	**0.851**	1.000	0.032	-0.068	-0.406	-0.497	-0.374
InterceptLinear	0.076	0.032	1.000	0.970	-0.679	0.398	0.352
InterceptLog	-0.012	-0.068	**0.970**	1.000	-0.555	0.445	0.401
GrossEfficiency	-0.502	-0.406	-0.679	-0.555	1.000	-0.007	0.042
WHOScore1	-0.552	-0.497	0.398	0.445	-0.007	1.000	0.952
WHOScore2	-0.515	-0.374	0.352	0.401	0.042	0.952	1.000

Remarks

Correlations among different indicators of HCS efficiency:

SlopeLinear efficiency measured as slope of health expenditure, linear model; full set of control variables.

SlopeLog efficiency measured as slope of health expenditure, log model; full set of control variables.

InterceptLinear efficiency as intercept in a linear model; full set of control variables but no country slopes.

InterceptLog efficiency as intercept in a log-model; full set of control variables but no country slopes.

GrossEfficiency Gross Efficiency proxy of costs per life year gained: ratio of change in Health Expenditure to change in Life Expectancy of 1999 relative to 1990.

WHOScore1 WHO's score of HCS efficiency based on production of Disability Adjusted Life Expectancy (DALE) only; estimates for 1997 WHO (2000 Annex 10).

WHOScore2 WHO score of HCS efficiency based on the HCS' overall Performance (composite output measure encompassing DALE, fairness and responsiveness); estimates for 1997 WHO (2000 Annex 10).
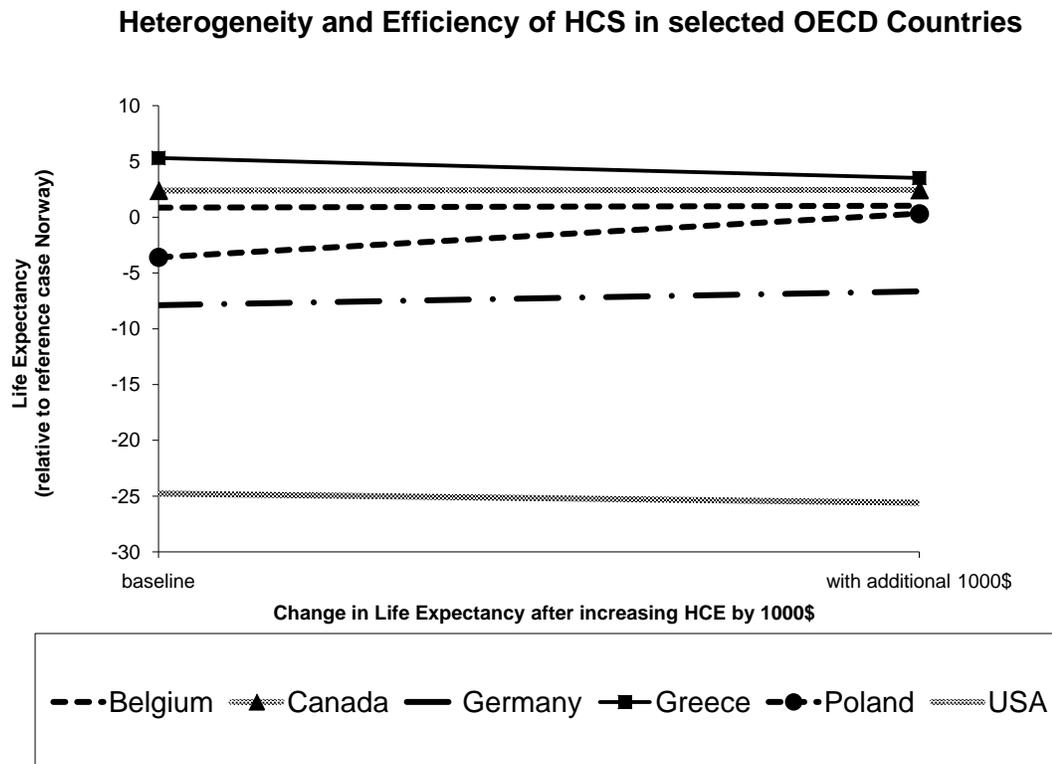
Heterogeneity and Efficiency of HCS in selected OECD Countries

Fig. (2). Heterogeneity and Productive Efficiency of selected HCS.

In order to relate the obtained efficiency indicator (the country-specific slope for HCE) to existing measurements of HCS efficiency, the efficiency scores obtained using the full models of each functional form (first column in Tables **1** and **2**) were correlated with alternative measures of efficiency. Table **3** gives the correlations between the various efficiency scores.

The first reference is the "efficiency as intercept" concept of efficiency. Two versions were used: first, the country-specific intercept as obtained in a linear model, using the full set of control variables but no country slopes (country intercepts reported in column 3 of Table **1**). The second version is identical, but uses the log-model (column 3 in Table **2**).

The WHO efficiency indicators constitute the first external reference. The first concerns the efficiency in producing only DALE (i.e., only Disability Adjusted Life Expectancy). The second one is broader, concerning the efficiency in coming close to an ideal HCS, i.e., high-performing, fair and responsive HCS. Both scores are from the World Health Report [43]. Next, I calculated a rather intuitive efficiency indicator based on the ratio of increase in HCE relative to the increase in life expectancy for the period 1990 to 1999; dHCE/dLifeExpectancy. It is a gross indicator, since it basically assumes that all increase in life expectancy is due to increased investments in the health system.

The results using different functional forms (linear *vs* log version) for the production of health but the same concept of efficiency (slope *vs* intercept) correlate strongly. The low correlation among different approaches indicates that a different concept of efficiency leads to a fundamentally differ-

ent evaluation of the HCS. The results obtained by the WHO correspond to some degree with the results obtained using the conceptually equivalent intercept approaches. But they correlate negatively with efficiency estimates based on the slope concept. HCS, like Greece and Japan, are evaluated as highly efficient according to the intercept approach, but are evaluated as only average (Japan) or even under-performing (Greece) according to the slope approach.

DISCUSSION AND CONCLUSION

The main result of the paper is that the distinction among heterogeneity and efficiency is an empirically important one and countries differ in both respects. For the relatively homogeneous sample of the OECD countries, the differences in HCS efficiency, albeit existing, are relatively small and dominated by heterogeneity.

The major differences in achieved life expectancy are due to country-specific features. Some of these are known and can be controlled for. But substantial variation remains, due to factors unknown and hence uncontrollable. Thus, an efficiency measurement based on outcome levels is heavily influenced by features which are unknown. Whether they reflect HCS efficiency cannot be inferred. They may reflect efficiency; they may also reflect something outside of the HCS, which can be anything, from regulations on safety at the workplace to the quality of the roads. Changes in health levels over time are more likely due to changes in variables like lifestyle, than changes in HCS efficiency.

The similar levels of productive efficiency as defined and found here, might reflect the usage of the same set of medi-

cal knowledge in the countries. The variation in HCS efficiency, and thus the leeway for efficiency improvement by reforms which change the HCS institutionally, seems to be overestimated.

Regarding the evaluative character of efficiency measurements, the comparison of efficiency concepts (intercept-based *vs* slope-based) and their implementation reflect that the evaluation and ranking of HCS is to a large degree dependent on the concept of efficiency one has in mind. Even within an approach, the results are strongly influenced by the choice of control variables included in the statistical model.

By way of an outlook, I would further argue that the public's preferences substantially affect input/output-based efficiency measurements. Firstly, investing in health has diminishing returns: at a certain output level, the additional costs of creating more output increase dramatically. This will negatively influence any efficiency measurement which is based on an overall input/output-ratio. It is a question of preferences, not of efficiency of production, whether a society is willing to pay for this additional increase in life expectancy or quality of life. Secondly, one has to take into account everything the HCS is actually producing. Citizens increasingly demand beyond-health outputs, like high accommodation standards in hospitals, and more resources are consumed to produce these. This implies that evaluating the HCS only with regard to biological health outcomes is to some degree misleading, because it assumes that all HCE is spent to produce health, which may not be true. It seems to me to be necessary to extend the framework of efficiency evaluation from a single-outcome to a multiple-outcome framework, capturing both health and beyond-health outcomes. While the WHO's overall performance indicator captures multiple outcomes, like DALE, fairness and responsiveness, I cannot apply the proposed slope-approach to the multiple-outcome setting, since the responsiveness-data is survey based and only available for one year. As far as the preference for beyond-health outputs is constant over time, its effect will be included in the country intercept, e.g., if citizens of a country constantly demand more beyond-health output, this will lower the country-intercept, because a share of HCE is spent on this output rather than on health production. This would leave the estimated slope efficiency unbiased. A bias occurs only if the preference for beyond-health output changes over time. If an increasing share of HCE is spent on beyond-health output, the slope of HCE will be smaller, and the HCS will appear as less efficient.

Regarding the implications for health policy, it seems that the role of HCS efficiency is overestimated, and therefore the potential for efficiency improvements is too. Differences in life expectancy are basically due to factors outside of the HCS. Tackling these might be more effective than changing the HCS.

REFERENCES

[1]   Evans DB, Tandon A, Murray CJL, Lauer JA. The comparative efficiency of national health systems in producing health: an analysis of 191 countries. Geneva: World Health Organization Global Programmed on Evidence for Health Policy Discussion Paper Series: No. 29 2000.

[2]   Retzlaff-Roberts D, Chang CF, Rubin RM. Technical efficiency in the use of health resources: a comparison of OECD countries. Health Policy 2004; 69(1): 55-72.

[3]   Greene WH. 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(10): 959-80.

[4]   Cutler DM, Deaton AS, Lleras-Muney A. The determinants of mortality. New York/Stanford: NBER Working Paper No. 11963, 2006.

[5]   Elo IT, Preston SH. Educational differences in mortality: united States, 1979-85. Soc Sci Med 1996; 42(1): 47-57.

[6]   Feinstein JS. The relationship between socioeconomic status and health: a review of the literature. Milbank Q 1993; 71(2): 279-322.

[7]   Filmer D, Pritchett L. The impact of public spending on health: does money matter? Soc Sci Med 1999; 49(10): 1309-23.

[8]   Leigh JP. Direct and indirect effects of education on health. Soc Sci Med 1983; 17(4): 227-34.

[9]   Meer J, Miller DL, Rosen HS. Exploring the health-wealth nexus. J Health Econ 2003; 22(5): 713-30.

[10]   Nolte E, McKee M. Measuring the health of nations: analysis of mortality amenable to health care. Br Med J 2003; 327: 1129-33.

[11]   Pritchett L, Summers LH. Wealthier is healthier. J Hum Res 1996; 31(4): 841-68.

[12]   Mooney G. Agency in health care: what is the maximand? Aberdeen: HERU Discussion Paper 1991; 10/91.

[13]   Mooney G. Beyond health outcomes: the benefits of health care. Health Care Anal 1998; 6(2): 99-105.

[14]   Anderson GF, Reinhardt UE, Hussey PS, Petrosyan V. It's the prices, stupid: why the United States is so different from other countries. Health Aff 2003; 22(3): 89-105.

[15]   Berndt ER, Cutler DM, Frank RG, Griliches Z, Newhouse JP. Price indexes for medical care goods and services: an overview of measurement issues. Cambridge, Mass, NBER Working Paper No. w6817, 1998.

[16]   Newhouse JP. Medical care costs: how much welfare loss? J Econ Perspect 1992; 6(3): 3-21.

[17]   Newhouse JP. Medical care price indices: problems and opportunities. Cambridge, Mass: NBER Working Paper No. w8168 2001.

[18]   Cornwell C, Schmidt P, Sickles RC. Production frontiers with cross-sectional and time-series variation in efficiency levels. J Econom 1990; 46(1-2): 185-200.

[19]   Dorfman JH, Koop G. Current developments in productivity and efficiency measurement. J Econom 2005; 126(2): 233-40.

[20]   Farrell MJ. The measurement of productive efficiency. J R Stat Soc A 1957; 120(3): 253-82.

[21]   Kumbhakar SC, Lovell CAK. Stochastic frontier analysis. Cambridge: Cambridge University Press 2000.

[22]   Schmidt P, Sickles RC. Production frontiers and panel data. J Bus Econ Stat 1984; 2(4): 367-74.

[23]   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(6): 493-504.

[24]   Frijters P, Haisken-DeNew JP, Shields MA. The causal effect of income on health: evidence from German reunification. J Health Econ 2005; 24(5): 997-1017.

[25]   Tengs TO. Dying too soon: how cost-effectiveness analysis can save lives. Dallas/Washington: National Center for Policy Analysis, Policy Report S207, 1997.

[26]   Huang HC. Estimation of technical inefficiencies with heterogeneous technologies. J Productiv Anal 2004; 21(3): 277-96.

[27]   Tsionas EG. Stochastic frontier models with random coefficients. J Appl Econom 2002; 17(2): 127-147.

[28]   OECD Health Data. Statistics and indicators for 30 countries. Paris: OECD 2005.

[29]   Snijders TA, Bosker RJ. Multilevel analysis: an introduction to basic and advanced multilevel modeling. London: Sage 1999.

[30]   Steenbergen MR, Jones BS. Modeling multilevel data structures. AJPS 2002; 46(1): 218-37.

[31]   Seshamani M, Gray A. Time to death and health expenditure: an improved model for the impact of demographic change on health care costs. Age Ageing 2004; 33(6): 556-61.

[32]   Zweifel P, Felder S, Meiers M. Ageing of population and health care expenditure: a red herring? Health Econ 1999; 8(6): 485-96.

[33]   Bhargava A, Jamison DT, Lau LJ, Murray CJL. Modeling the effects of health on economic growth. J Health Econ 2001; 20(3): 423-40.

[34]   Gerdtham UG, Löthgren M. On stationarity and cointegration of international health expenditure and GDP. Stockholm: Stockholm School of Economics Working Paper, 232, 1999.

[35]   Gerdtham UG, Löthgren M. New panel results on cointegration of international health expenditure and GDP. Appl Econ 2002; 34(13): 1679-86.

[36]   Jewell T, Lee J, Tieslau M, Strazicich MC. Stationarity of health expenditure and GDP: evidence from panel unit root tests with heterogenous structural breaks. J Health Econ 2003; 22(2): 313-23.

[37]   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.

[38]   Navarro V. The new conventional wisdom: an evaluation of the WHO report Health Systems: improving performance. Int J Health Serv 2001; 31(1): 23-33.

[39]   Williams A. Science or marketing at WHO? A commentary on 'World Health 2000'. Health Econ 2001; 10(2): 93-100.

[40]   Darby C, Valentine NB, Murray CJL, de Silva A. Strategy on measuring responsiveness. Geneva: World Health Organization Global Programme on Evidence for Health Policy Discussion Paper Series No. 23, 2000.

[41]   Valentine NB, de Silva A, Murray CJL. Estimating responsiveness level and distribution for 191 countries: methods and results. Geneva: World Health Organization Global Programme on Evidence for Health Policy Discussion Paper Series No. 22, 2000.

[42]   Grubaugh SG, Santerre RE. Comparing the performance of health care systems: an alternative approach. South Econ J 1994; 60(4): 1030-42.

[43]   WHO (World Health Organization) The world health report. Geneva: WHO 2000.