A Value-Added Based Measure of Health System Output and Estimating the Efficiency of OECD Health Systems

Dennis Petrie, Kam Ki Tang, and D.S. Prasada Rao School of Economics Discussion Paper No. 393, June 2009, School of Economics, The University of Queensland. Australia.

ABSTRACT:

Life expectancy at birth is the most commonly used measure for health system output. However, there are a number of reasons why it may be a poor proxy. First, life expectancy assumes a stationary population and thus does not take into account the current demographic structure of a country; and second, the output of a health system should be measured in terms of the value-added to the population's health status rather than health status itself. The paper develops a new measure of health system output, the Incremental Life Years to address these problems. The new measure is applied to study health system output, efficiency and total factor productivity in OECD countries for the years 2000 and 2004. The new measure provides different results compared to those based on the traditional life expectancy measure, and the differences are further accentuated when changes in efficiency and productivity are estimated.

EPrint Type: Departmental Technical Report

Keywords: Life expectancy, health system, efficiency, total factor productivity, OECD

Subjects:

ID Code: JEL Classification: C43, I12

Deposited By:

A Value-Added Based Measure of Health System Output and Estimating the Efficiency of OECD Health Systems

Dennis Petrie,[a], Kam Ki Tang[b, *, #] and D.S. Prasada Rao[b]

[a] University of Dundee, Economic Studies, School of Social and Environmental Sciences

[b] The University of Queensland, School of Economics

Abstract

Life expectancy at birth is the most commonly used measure for health system output. However, there are a number of reasons why it may be a poor proxy. First, life expectancy assumes a stationary population and thus does not take into account the current demographic structure of a country; and second, the output of a health system should be measured in terms of the value-added to the population's health status rather than health status itself. The paper develops a new measure of health system output, the Incremental Life Years to address these problems. The new measure is applied to study health system output, efficiency and total factor productivity in OECD countries for the years 2000 and 2004. The new measure provides different results compared to those based on the traditional life expectancy measure, and the differences are further accentuated when changes in efficiency and productivity are estimated.

JEL Classification: C43, I12

Keywords: Life expectancy, health system, efficiency, total factor productivity, OECD

[*] Kam Ki Tang would like to acknowledge the support by a grant from Australia Research Council (DP0878752).

[#] Corresponding author. The University of Queensland, School of Economics, QLD 4072, Australia. Tel: +617 3365 9796. Fax: +617 3365 7299. Email: kk.tang@uq.edu.au

1. Introduction

In the past four decades, the share of GDP being spent on health has steadily increased across OECD countries, from 3.8 percent in 1960 to 6.6 percent in 1980 and further to 9.1 percent in 2006.[1] With increasing shares of GDP being spent on health, health sector performance and efficiency have become important issues in the health policy arena (OECD 2002; 2004). In particular, given the large and rising health care costs, even a small improvement in health system efficiency can mean large savings. Moreover, policymakers are concerned with the consequences of rising health expenditures on government budgets, especially in the context of ageing populations (Roserveare et al. 1996). How to respond to the population ageing and associated ramifications on health care and pension outlays is still an open question (Elmendorf and Sheiner 2000). It is clear, however, that to meet the rising need for health care services without placing strain on the public purse, improved health system efficiency is critical.

There are three levels of efficiency analysis for health care services, namely the disease level, the sub-sector level, and the system level (Häkkinen and Joumard 2007). The system level takes the broadest perspective in that it examines the efficiency of a country's entire health system; the disease level takes the narrowest focus to analyze the efficiency in dealing with individual diseases; and the sub-sector level takes the middle ground by looking at individual components of the health sector, in particular hospitals, nursing homes, primary care facilities, or physicians.[2] The system approach has an advantage that it attempts to encompass all components of the health sub-sectors, and thereby includes linkages between sub-sectors (e.g. physicians and specialists) in providing health care services. Health system efficiency analysis, however, is much less common than sub-sector and disease level analyses. The methodological problems of efficiency analysis are a major reason for this; for efficiency

[1] The figures are based on OECD Health 2008 database. Figures for earlier years cover fewer countries.
[2] For a review of efficiency analysis in health care, see Hollingsworth (2003).

analysis to play a larger role in actual policy deliberation, much more work is needed (Hollingworth and Street 2006). The publication of the World Health Report 2000 and related studies (WHO 2000; Evans, Tandon, Murray and Lauer 2000, 2001) highlighted the importance of the topic, albeit the controversy surrounding it (see, for example McKee 2001 and Williams 2001). Following the WHO publication, there were a series of attempts to address various methodological problems in measuring health system efficiency, including Hollingworth and Wildmand (2003) and Greene (2004, 2005). This paper, however, aims to address a different issue in health system efficiency analysis that has not received much attention in the literature, namely the appropriateness of using life expectancy at birth (LE) as a health system output measure and measuring inputs in per capita terms (e.g. health expenditure per capita). Given that estimation methods can deliver useful results only if outputs and inputs are measured accurately, the measurement issue raised in this paper is arguably of primary importance.

Life expectancy at birth (LE) is by far the most commonly used measure of health system output. Although LE has been widely accepted as an average health status indicator, it can be a poor indicator of health system output and performance for the following reasons. First, even in a primitive society with no health system in place, LE would not be equal to zero. Therefore, using LE as an output indicator grossly overstates the value-added of the health system to lifespan. Furthermore, LE measurement assumes a stationary population whereby the population in any age-sex group is computed based on the current (or more precisely recent past) mortality rates of younger age groups. Therefore, LE does not capture the actual demographic structure of the country. By the same logic it is also inappropriate to compare health inputs in per capita terms unless the actual population structure is the same over time and across countries. For instance, a country with an aged population can have the same LE as a country with a young population; however, the health system that serves an aged population is likely to be spending more compared with a health system providing for a younger population. Measuring health input in per capita terms, therefore, wrongly indicates that the former is less cost efficient than the latter in achieving the same LE.

Once we start to consider demographic factors, it is obvious that the contributions of the health system to different age-sex groups are vastly different. For instance, the reduction in infectious disease since 1840 reduced mortality rates of infants and children more than those of adults[3] (Cutler, Deaton, and Lleras-Muney 2006). Not only does the contribution of the health system to reducing mortality rates vary across age and sex, but also the "average cost" of producing a certain health output per person (e.g. additional life year saved) may vary for each age-sex group.

Considering these limitations of LE, we propose a new measure of health system output, namely the *Incremental Life Years* (ILY) and its variant *Relative Incremental Life Years* (RILY). Simply put, ILY measures the value-added of the health system in terms of extending the longevity of its current population. Evans, Tandon, Murray and Lauer (ETML) (2000) also put forward a similar framework of using the observed health status above an estimated minimum level as a measure of the health system output.[4] Using ILY to replace LE obviously does not resolve all the difficulties in measuring health system output. For one, a significant portion of health services in developed countries are directed toward addressing morbidity problems, especially improving the quality of life for the elderly, without necessarily affecting their life spans. The methodology adopted here can be readily extended to include both mortality and morbidity in order to produce an output measure such as *Incremental Healthy Life Years*, though the lack of reliable and comparable morbidity data forces us to focus on mortality based measures in the current paper.

[3] Improvement of public health was believed to be a contributing factor.
[4] Their health status measure is disability adjusted life expectancy (DALE). They estimate the minimum level of DALE from a cross-section of 25 countries at around 1908 as their baseline health status, controlling for literacy rate. The exact method and data are not clear as details are provided in another study, Evans, Bendib, Tandon et al. (2000), which apparently was not available on the WHO website anymore (date of access: 7 April, 2008).

The rest of the paper is organized as follows. Section 2 further explains the rationale behind the new health system output measure and its actual construction. Section 3 applies the new output measures to 30 OECD countries and compares the new health output measure with LE. Focusing on only OECD countries can go some way to mitigate problems in relation to data availability and quality, as well as the heterogeneity problems inherent in the WHO study.[5] For illustrative purpose, in Section 4 we use the newly proposed output measures and data envelopment analysis method to analyze the efficiency and productivity of health systems of the OECD countries, and contrast the results with those obtained using LE. Here it should be emphasized that the objective of this paper is to contribute to the *methodology* of measuring health system output and input, additional work is needed before the results of cross country efficiency and productivity comparisons can provide conclusive *policy direction*. Lastly, Section 5 discusses the findings and offers some concluding remarks.

2. A Value-added Measure of Health System Output

2.1 Number of Lives Saved

In the national accounts, the contribution of a sector is measured by the value-added generated by that sector, i.e. the gross output net of intermediate inputs used. Similarly, the value-added of an operation or a medical intervention, ignoring the quality of life issue, may be measured by the number of life years the patient will have after the operation minus the number of years the patient would have lived without the operation. The same principle could be applied to measuring the output of a hospital or the entire health system. Using LE as a measure of health system output violates this basic principle. Life expectancy at birth encompasses the effects of natural health endowment, the health system, and environmental factors on *all* age groups. Here environment is used as a general term, referring to all factors that can affect health but are not part of natural health endowment or health resources, such

[5] Green (2004, 2005) explain that a key problem of the method used in WHO (2000) is that it does not account for the heterogeneity of the large sample of 191 countries.

as education, and law and order. The role of natural health endowment of human beings is clearly evident in the fact that, even in primitive societies with no health systems LE was well above zero. Population studies have found that LE was about 25 years in the earliest human societies (Acsádi and Nemeskéri 1970), about two-thirds of the LE in many Sub-Saharan African countries nowadays and close to one-third of the OECD average. Simply subtracting this natural LE from the current LE, however, is not an appropriate measure of the value added of the health system. This is because the contribution of the health system to the reduction of mortality risks varies across age and gender, and thus the demographic structure of a country can make a substantial difference to actual health system output and performance, however, LE assumes a particular demographic structure that is unlikely to hold.

To address the aforementioned problems, we propose a new measure—Incremental Life Years (ILY)—as a measure of health system output. To understand the rationale behind ILY, we first start with a simpler concept—the number of lives saved. Suppose we divide the population of country *i* into a number of age-sex groups. The total number of lives saved by the health system in a given year is equal to

$$LS_i = \sum_x (p_{ix} - \hat{p}_{ix}) n_{ix} \tag{1}$$

where p_{ix} is the *observed* survival probability of age-sex group x in country i, $\hat{p}_{ix}$ the counterfactual survival probability of the group if the health system ceased to function throughout the year, and n_{ix} the *actual* population size of the group.

Obviously $\hat{p}_{ix}$ is not observable. To circumvent the problem, we use proxy life tables as a baseline measure (a number of different proxy life tables are considered with details given in Section 2.2). The baseline survival probabilities are assumed to be country invariant, i.e. $\hat{p}_{ix}$ becomes $\hat{p}_x$.[6] Because the baseline probabilities are common to all countries, the value of LS_i will depend not only on the value-added of country *i*'s health system, but also on the socio-economic conditions present in the country. Consequently variation in health system efficiency between countries will be affected not only by the differences in their health care policy and institutional settings, but also by the heterogeneity in their environment. The efficiency effects of policy, institutional and environment elements can be controlled by including those elements into the efficiency model; alternatively, they can be isolated from the "raw" efficiency scores using regression methods to derive efficiency scores conditional on those elements. However, since the main purpose of the paper is to introduce the new health output measure, we will abstract from this issue. Focusing on the relatively homogenous OECD countries helps mitigate this problem to some extent.

2.2 Incremental Life Years (ILY)

In equation (1), it is implicitly assumed that the value of the life saved of a young person is the same as that for an old person.[7] Since a young person is expected to live longer than an old person, it is more reasonable to measure health system output in terms of the number of *expected life years saved* rather than the number of lives saved. Here we use the total number of life years (expected to be) saved, or simply Incremental Life Years (ILY), which is given by

$$ILY_i = \sum_x (p_{ix} - \hat{p}_x)\,\hat{e}_x\, n_{ix} \qquad (2)$$

Here $\hat{e}_x$ is the baseline LE of the group *x* and is equal to the number of years a person is expected to live after being saved in the current year without any assistance from the health

[6] In ETML (2000), the baseline health status is country specific because it is conditional on the country's literacy rate. Since our sample covers only the OECD countries, not controlling for the most basic level of education should not be a cause for concern.
[7] Because an young person saved is assumed to live longer than an old person saved, using lives saved implies a weighting system where a year lived by an older person is of greater value than that lived by a younger person.

system in the future. The baseline LE instead of country *i*'s life expectancy (i.e. e_{ix}) is used for the following reason. Life expectancy is calculated using mortality rates of all age groups. Therefore, if we were to use age-specific LE of the country itself to calculate ILY for those who aged, 10-14, then we would implicitly also be counting the benefit of future health system resources that this age group may need to receive in order to face the same mortality risks that their older cohorts currently enjoy. As a result, the ILY of this age group would have taken into account the resources received by those aged 10-14, as well as future health resources needed when this cohort was aged 15-19, 20-24,…etc. Obviously this is not desirable because we only want to consider the direct benefit of the current health provision and not the benefit of future healthcare provision. Using the baseline LE can mitigate this problem because it is assumed to reflect the additional life years gained due to current health care that could be expected if no health resources were provided in subsequent years. Consequently, there will be no double-counting of the contribution of the health system to the older cohorts.

The value of $\hat{e}_x$ is calculated using the information of $\hat{p}_x$ across the entire age spectrum using standard life table methods. In essence, we weight the number of lives saved for each age-sex group by the baseline LE at the age of intervention. This involves value judgments about the relative values of life for different age-sex groups. The current specification of ILY means that lives of younger people are valued more than those of elderly, as they will live for longer, but each of the years lived is valued the same. Whether these judgments are appropriate is an important question of its own. For instance, in a recent study, Kniesner, Viscusi and Ziliak (2004) show that because people prefer to delay consumption till old age, the value placed on life for the elderly actually is higher than that for the young. Although we are not suggesting that the simple explicit judgment currently made in ILY is the most appropriate one, it is nevertheless transparent, allowing open discussion and even modifications. In comparison, value judgments are also implicitly made when using LE as the output measure. Additionally, it should be emphasized that the alternative weightings will not change the nature of ILY as a value-added based output measure. More importantly, the proposed weighting can be further justified by the fact that the resulting ILY measure can be interpreted objectively as the expected number of life years saved if there is a one-off increase in health system expenditure that increases the survival probabilities for just one year.[8]

2.3 Average and Relative Incremental Life Years (AILY and RILY)

If we want to compare health system performance across countries, we need to control for the size of the population or the number of life years that could *potentially* be saved by the health system. Otherwise we will risk overly crediting those countries that have a relatively large population and therefore could easily save a lot of life years at the "lower end" of the cost spectrum even though they may have a large amount of unmet demand at the "higher end".[9] To make this point clearer, consider Mexico and Austria which have a ten-fold difference in population size. Since the cost of saving a life is not constant but contingent on the circumstance, there will be a distribution (i.e. a spectrum) of intervention costs from very low to very high. To save, for example, 1000 lives out of its population of 8.4 million, Austria probably needs to work on cases of low intervention costs (e.g. diseases that can be prevented with low cost immunization) as well as of high intervention costs (e.g. diseases that requires complicated operations). On the contrary, to save 1000 lives out of its population of 108 million, Mexico could just focus on cases of low intervention costs and ignore the high cost ones. If the average cost-effectiveness of intervention is independent of age and sex, this problem could be easily addressed by normalizing ILY by the population size, however, this

[8] In reality, increases in survival probabilities in a given year are also likely to some extent be due to health investment in previous years.
[9] "Lower end" and "higher end" refer to the relative costs necessary to save a life.

is unlikely the case and therefore ILY needs to be normalized by an index which takes into consideration both the population size and the relative difficulty/ease of saving life years in each age-sex group.

To address this issue, we gauge the health performance of a country against some benchmark that automatically adjusts to the demographic structure of the country in concern. To that end, we propose to use the Average Incremental Life Years (AILY):

$$AILY_i = \sum_x (\overline{p}_x - \hat{p}_x)\,\hat{e}_x\, n_{ix} \qquad (3)$$

The newly constructed measure, AILY, is calculated in the same way as ILY except replacing p_{ix} with $\overline{p}_x$, which is the median survival probability observed amongst all the countries in the dataset (i.e. OECD in our case) in a given, base year.[10] AILY therefore represents the total number of life years that would have been saved in a country that had the same demographic profile as country *i* but with standardized, average mortality rates $\overline{p}_x$ in the based year. Note that $\left(\overline{p}_x - \hat{p}_x\right)\hat{e}_x$ is the weight applied to each age-sex group and represents the difficulty (on average experienced by countries) for saving life years per population in each age-sex group, with a high value indicating that countries on average save a large number of life years per person in that age-sex group compared to other age-sex groups. In a way, $\left(\overline{p}_x - \hat{p}_x\right)\hat{e}_x$ can also be considered as a proxy for the demand for life years to be saved per person in each age-sex group, where the demand is assumed to be proportional to the number of life years saved per person in the "average" country in the base year.

Normalizing ILY by AILY yields another output measure called the Relative Incremental Life Years (RILY):

[10] It is possible that $\overline{p}_x$ is measured based on a year different from that of p_{ix} or $\hat{p}_x$.

$$RILY_i = \frac{ILY_i}{AILY_i} = \frac{\sum_x (p_{ix} - \hat{p}_x)\,\hat{e}_x\, n_{ix}}{\sum_x (\bar{p}_x - \hat{p}_x)\,\hat{e}_x\, n_{ix}} \tag{4}$$

Since RILY controls for the demographic structure and because it is an index measure defined relative to average levels, it is useful for comparing health system output across countries. Furthermore, as AILY can be interpreted as an index for total demand upon the health system,[11] using RILY to measure health system output has another important merit over LE in that it provides an index as to what extent the country has successfully met the demand for health services.

Since RILY is equal to the actual number of life years saved relative to some benchmark value, it may also be considered as a type of efficiency measure reflecting the extent to which a health system is able to meet the demand. However, it does not take into account the inputs used by the country to achieve this outcome. It should also be pointed out that although $\hat{e}_x\, n_{ix}$ enters both AILY and ILY, it does not cancel out the effect of demography on RILY, it merely moderates it. Additionally, if a country has higher than average survival probabilities in all the age groups, its RILY will be greater than one.

3 Computations of ILY and RILY

3.1 Baseline Survival Probability

To implement the new measures ILY and RILY, we need to specify the baseline survival probability $\hat{p}_x$. We have experimented with the survival probabilities of four different populations, respectively denoted as the ancient population, US Black population in 1901, low income and low-middle income populations as a proxy for the baseline survival

[11] This implies that some countries will fall short of meeting the "average" demand while others will exceed it. However, whether a country falls short or exceed average demand will depend on the amount of health resources allocated and how efficiently these are used.

probability. In general, the survival probabilities progressively increase from the ancient population to the low-middle income population.

The survival probability of the ancient population is a measure of the natural survival probability of human species. It has been well established in population studies that for a population to be sustainable, LE must have been above 20, and for it to exceed 30, some kind of modern medical knowledge is required (Hopkins 1966). A LE of 25—the estimated figure for hunter-gatherers—is widely considered the upper limit a population can achieve without modern medicine (Acsádi and Nemeskéri 1970). Coale et al. (1983) has published a series of life tables pertaining to different development stages of human societies. At the "level 3" development classified by their study,[12] the LE for males and females are equal to 22.9 and 25.0 years, respectively. We select these life tables to provide us with the data for the natural survival probability of human species.

The second baseline population is the US Black population in 1901. This is the earliest obtainable life table for the Black population that is compatible with other life tables used in this study. Though slavery in the US was abolished in 1865, it is reasonable to assume that the Black population still received little health care services in the late 1800s, making their survival probability a good proxy for the natural survival probability. The LE for US Black males and females in 1901 are 32.5 and 35.0 years, respectively. These figures are comparable to the LE of present bottom African countries that are severely affected by the HIV/AIDS epidemic.

The third baseline population is based on low income countries in year 2000 as classified by the World Bank, excluding those involved in military conflicts, or with a

[12] Besides the level of development, the authors also distinguish four different regions in the construction of human life tables. The four regions are "North", "East", "West" and "South". The data of natural survival probability is drawn from the "West" model life tables. The "West" model life tables are preferred to the other three because "they were based on mortality experience recorded in populations known to have relatively good vital statistics, and not showing a persistent systematic pattern of deviations from the preliminary model tables." (Coale et al., 1983, p.12)

HIV/AIDS prevalence over 1 percent or where no data were available. The list of the 20 low income countries that form this baseline population is provided in the Appendix. Year 2000 is chosen because this is the earlier one of the two years, 2000 and 2004, under analysis. For each age-sex group, the median survival probability from the 19 countries is taken as the baseline survival probability for the low income group. The LE of males and females for this baseline population are respectively 58.5 and 61.1 years.

The fourth baseline population is based on the low income countries plus the lower-middle income countries in 2000 as classified by the World Bank. Again, after excluding countries in military conflicts, or with a high HIV/AIDS prevalence, or those where no data were available, the median survival probabilities from the remaining 63 countries are used as the baseline measures. The male and female LE's for this baseline population are respectively 64.0 and 69.0 years. This group is labeled "low-middle income" group.

Each of the four selected baseline populations has its limitations. Using the ancient population is likely to overstate the valued added of modern day health systems by attributing to it the effects of improvement in nutrition, basic education and eradication of certain types of diseases in the past few decades (Culter et al. 2006). Even if we define the health system very broadly, the ILY based on the ancient population will still measure the cumulative valued added of the health system over decades rather than over a single year. The US Black population has a similar limitation, but to a lesser extent. However, it is based on a single country data. The low income populations are less country specific than the US Black population and takes into account at least to some degree the eradication of diseases and improvements in basic education, but African and Middle Eastern countries are overly represented in this group. By including more countries from other continents, the low-middle income population increases the degree of generalization compared with using just the low income countries. Nevertheless, since a functioning health system may be in place in many low-middle income countries, their survival probabilities may not truly indicate the counterfactual survival probabilities in the absence of health systems.[13] On balance, the low income population is our preferred baseline population. Despite this, we still report the results based on other baseline populations in order to demonstrate the robustness of the RILY measure to the choice of the baseline population.

3.2 Estimates of Health System Output for OECD Countries

We estimate health system output using the new measures proposed in the study, ILY and its variant RILY, for 30 OECD countries for the years 2000 and 2004. Two years are chosen in order to examine efficiency changes in these countries over time. To make the output measures of the two years commensurable, we need to apply the same median survival probability observed amongst all the countries, $\overline{p}_x$, in calculating AILY for both years. Either the median values of year 2000 only, or year 2004 only, or both years can be used, but, given the two years are not that far apart and the rate of decline in mortality figures is quite stable across age-sex groups, the choice will make little difference. In this paper, we use the median values of year 2000, but, will also provide the result using the median values of year 2004 as a robustness test. Data for the OECD countries are drawn from the OECD Health Data 2008.[14] Mortality data for the US Black population are drawn from the Human Life-Table Database.[15] Mortality data for low and low-middle income countries are drawn from the WHO Statistical Information System (WHOSIS).[16]

The estimates of ILY and RILY are shown in Table 1, alongside with LE. The variation in LE amongst the OECD countries, while it exists, is limited. For instance, in year 2000, the coefficient of variation (i.e. standard derivation divided by the mean) is merely equal to 0.03, with the highest LE (Japan) being equal to 1.18 times that of the lowest one

[13] We have made an adjustment for this in terms of health expenditure, see section 4.2.
[14] http://www.oecd.org/health/healthdata
[15] http://www.lifetable.de/
[16] http://www.who.int/whosis/en/

(Turkey). As explained before, since LE is a measure of health status rather than the value-added of the health system, these figures clearly understate the gap between health system outputs amongst countries. When RILY(ancient) and RILY(US Black) are used as health system output measures, the result remains largely the same. However, when RILY(low income), and RILY(low-middle income) are used instead, the picture changes dramatically. The coefficient of variation has increased to 0.12 and 0.23 respectively. Although Japan and Turkey are also of the highest and lowest RILY values, respectively, the ratios now increase to 1.80 (RILY(low income)) and 4.12 (RILY(low-middle income)). The ratio of 1.80 seems more plausible than the alternatives, supporting our choice of the low income population as the preferred baseline.

Despite the differences between LE and the four RILY measures, they are in general highly correlated, as illustrated in Figure 1. Notwithstanding the high correlation, one should be cautious not to jump to the conclusion that the new output measures are merely a rescaling of the LE and therefore will have no further implication. As will be seen next, the new output measure (together with the corresponding new input measure) leads to very different outcomes in terms of efficiency and productivity assessment.

4. Measuring Efficiency and Productivity of Health Systems

In this section, we demonstrate how ILY and RILY constructed in the previous section can be used to analyze the efficiency and productivity of OECD health systems. The discussion starts with an analytical framework, followed by the modeling method and specification, and then the empirical results.

4.1 Analytical Framework

Efficiency, in standard production economics, refers to the output produced, relative to what could have potentially been produced with a given set of inputs and production technology.

Thus, measuring health system efficiency implies a production function that transforms health inputs into health output. Suppose a health system production function can be represented with the following general form:

$$Health\,System\,Output_{it} = f(demography_{it}, health\,resources_{it}, technology_{t}, efficiency_{it}) \quad (5)$$

Demography is a key factor in driving the demand for health services in that ill people are a necessary “input” in order to save life years. In the current setting, technology is considered global and shared by all OECD countries but may change over time. In the case where LE is used as a measure of health output per capita, the output of the whole health system will be equal to LE multiplied by the population, or total life expectancy (TLE). For the purpose of exposition, we assume a Cobb-Douglas function form for $f(.)$[17]:

$$TLE_{it} = A_t\,\theta_{it}\,(HR_{it})^{\alpha}\,(POP_{it})^{\beta} \quad (6)$$

where A_t represents the technology available to all health systems at time *t*; and HR_{it}, POP_{it} and θ_{it} respectively represent the level of health resources, the population size, and the efficiency of country *i* at time period *t* ($0 \le \theta_{it} \le 1$).

Equation (6) implies that for a given demographic structure, technology and efficiency level, there are diminishing returns to health spending. This is because, to maximize the effects of health spending, typically the life years that are the cheapest to save will be saved first; as a result, the cost of saving an extra life year will increase as each additional life year is saved. The equation also implies for a given technology, efficiency and health resources, *TLE* increases with diminishing returns to *POP*. Suppose the population increases by duplicating the original population (twice as many “sick people”) but the total health resources is held constant. If the authority shifts some of the resources previously used to treat the relatively more expensive cases of the incumbent residents (all the cheap cases of

[17] The actual estimation does not rely on a specific production function form.

the incumbents are still dealt with) to the cheaper cases of the new arrivals, then the additional cheap cases that can be treated will be more than the expensive cases (that are now neglected), and therefore the total number of life years saved will increase. However, since some of the expensive cases are no longer being treated, the total increase in *TLE* will be less than double; hence population has diminishing returns to life years saved.[18]

If the production function is of constant returns to scale (CRS), i.e. $\alpha + \beta = 1$, then (6) can be easily made to resemble the per capita form commonly seen in the literature:

$$LE_{it} = TLE_{it} / POP_{it} = A_t \, \theta_{it} (HR_{it} / POP_{it})^{\alpha} \tag{7}$$

In other words, the use of (7) in the literature actually implicitly assumes the health production technology is of CRS with respect to health resources and population size. This assumption, however, has rarely been tested. Thus, our empirical analysis will be based on (6) instead, though the applicability of (7) will be tested.

When *ILY* instead of *TLE* is used to measure health system output, the production function becomes:

$$ILY_{it} = B_t \, \phi_{it} \left(HR_{it} \right)^{\gamma} \left(AILY_{it} \right)^{\phi} \tag{8}$$

where B_t is a counterpart of A_t and ϕ_{it} is a counterpart of θ_{it}.

$AILY_{it}$ is used instead of the total population because it takes into account the fact that the proportion of "ill" people and therefore demand for health resources is likely to be different for each age-sex group. Here $AILY_{it}$ represents the number of life years that would have been saved for the population in country *i* at time *t* if the country was like an "average" country in terms of health resources, health system performance and the available technology at some base year (2000 in our case). Therefore, $AILY_{it}$ is a function of demography for

[18] To double the total *TLE* it is required to provide treatment to both the cheap and expensive cases of the incumbents and new arrivals to the extent the incumbents originally received.

country i in the current time period t, the global technology in the base year, and average health resources and average efficiency for countries in the base year. It should therefore be noted that the only reason $AILY_{it}$ changes is due to differences in the demographic structure of the population both over time and across countries and, as such, any technology growth is captured within B_t and any efficiency differences both across countries and over time are captured within ϕ_{it}. This implicitly assumes that the technology growth and efficiency growth is the same for all age-sex groups which may be an appropriate approximation in the short-run but may be less appropriate in the long-run. We will show later that the choice of the base year average technology, health resources and efficiency within $AILY_{it}$ over our 4 year window has little effect on our results.

The reason why $AILY_{it}$ is used as an input can be understood in the case of a CRS production function, in which (8) can be simplified into

$$RILY_{it} = ILY_{it} / AILY_{it} = B_t \, \phi_{it} (HR_{it} / AILY_{it})^{\gamma} \qquad (9)$$

This equation indicates that the number of life years that are saved relative to what would have been expected to be saved in the "average" country if it had the identical demography to country i in time period t . This will be a positive function of health resources used per expected life year to be saved in the average country for the given demography, or, in other words, health resources used per unit of demand.

As explained before, RILY is a weighted per capita measure, and as such is similar to LE. Again, since we cannot presume CRS, we use (8) for our empirical work and compare the results with (6).

4.2 Data Envelopment Analysis and Model Specification

In total we estimate five main models (plus additional models for sensitivity tests). The specifications for each model are listed in Table 2. Since the purpose of this paper is to illustrate how the new output measure can affect performance assessment, we use the simplest model specification possible. The results using alternative model specifications will be discussed later. Besides health expenditure, the other input is either the total population or AILY, depending on the output measure used. There are five different output measures, namely TLE (corresponding to the total population input measure) and the four alternative measures of ILY which are constructed using different proxies for baseline mortality (corresponding to different AILY input measures).

Health outputs are measured for years 2000 and 2004. The corresponding health expenditures of the same years are measured in purchasing power parity, constant 2000 international dollars. The measure includes both public and private expenses, and is defined as the sum of expenditure on activities in relation to health promotion, prevention, curing, caring, and administration.[19] We use real expenditure per capita which has been adjusted for price level differences across countries and for movements over time. Different health input figures are used according to which health output measure is used. When TLE is used as the health output measure, health input is measured simply in total health expenditure, and the size of the population is used as a proxy for the demand of health services. When ILY is used as the output measure, the health expenditure needs to be modified in some cases. In the case of the ancient and US Black populations, it is reasonable to assume their health expenditure would have been negligible. The same assumption cannot be made for the low and low-middle income populations though. To correct for this, we deduct the median per capita health expenditure of low (and low-middle) income countries from the per capita health

[19] For detail definition, see OECD Health Data 2008 dataset.

expenditure of each of the OECD countries, and then multiply the resulting value with the population of the country to obtain net total health expenditure.

Estimates of efficiency of health systems depend on the underlying production frontier of the systems. The production frontiers can be identified using both the non-parametric, data envelopment analysis (DEA) method, and the parametric, stochastic frontier analysis (SFA) method. The SFA method has the advantage of controlling for random noise in data. However, the SFA estimation results show that the random noise component is very small compared to the inefficiency component[20]. Therefore, we adopt the DEA method because it does not require prior specification of the functional form of the production technology. The DEAP program (Coelli et al., 2005) is used for the computations.

Before considering the results we quickly review the basic terminology used in the standard efficiency and productivity measurement literature (see Coelli et al. 2005 for a more detailed discussion). There are a number of different measures of efficiency and this paper focuses on those most commonly used concepts in the health literature, namely the output orientated technical efficiency, which represents the proportion by which outputs of a country can feasibly be expanded given the technology and the level of inputs. When we examine health system efficiency over time, we can measure a number of aspects associated with health system performance. First, we can measure technical efficiency change (EFFCH) which provides an indication of the performance of the system vis-à-vis the technology in different periods. Technical efficiency change (EFFCH) can be decomposed into pure technical efficiency change (PECH) and scale efficiency change (SECH): $EFFCH_i = PECH_i \times SECH_i$ where the scale efficiency change due to changes in the scale of production. If the production technology exhibits *constant returns to scale* (CRS), SECH will

[20] The gamma value, which measures the size of the inefficiency component relative to that of inefficiency plus noise, is over 0.95 in all estimations, and mostly over 0.97. The SFA estimation is based on the assumption of a Cobb-Douglas production function with time dummies.

be equal to one, i.e. there is no efficiency gain or loss by alternating the scale of production. In addition, it is also possible to measure *technical change* (TC) when there is a shift in the production frontier. Technical change measures the degree by which output can be increased by an efficient firm using the same inputs but with the new technology available in the new period. The components EFFCH and TC are combined to obtain the *Malmquist productivity index* (MPI) that measures the change in total factor productivity (TFP) over a given period of time.

4.3 Efficiency Scores

The top panel of Table 2 presents various models with different combinations of inputs and outputs used in the study. The bottom panel of Table 2 shows the average efficiency scores, averaged over all the countries for a given year[21]. The country specific efficiency scores, under different models, are presented in Table 3. It can be seen that the efficiency scores from different models can vary substantially, especially for Hungary and Turkey. Using total life expectancy (TLE) as an output measure tends to generate the highest efficiency scores and ILY(low-middle income) the lowest efficiency scores. This result is due to the fact that TLE measures the gross health status of countries rather than value-added of their health system and, thus, tend to compress the differences in performance between countries. Amongst the four ILY measures, ILY(low-middle income) is based on the highest baseline mortality rates and thus tends to magnify the performance differences between the countries.

Based on our preferred baseline population—low income population, the health systems of Japan, Korea and Mexico are the most efficient (with an efficiency score of 1.00) in both years and are considered as "peer" countries that define the best performance for other countries. The finding of Mexico as being one of the most efficient OECD countries in providing health care services may raise questions for some. Here it is important to remember

[21] Unless and otherwise stated, all the averages are computed as unwieghted geometric means of the country-specific measures.

that, due to Mexico's input mix (i.e. a low level of health expenditure), it is only being compared with similar countries, such as Czech Republic, Hungary, Poland, Portugal, Slovakia and Spain; thus it can only be concluded that it is the best performing country within this group. We also need to be aware that this is a limitation of having a small sample and thus care must be taken when interpreting the results (in particular for TC and the TFP). The health system of Hungary is ranked the most inefficient in almost all models for both years, followed by the US.[22]

In order to examine differences between models it is more useful to analyze the relative performance of the countries. To this end, the correlations of the efficiency scores across the models are shown in Table 4. From Panels I and II, it can be seen that the correlations between the efficiency scores of the model that uses LE as health output measure and those that use ILY vary from 0.97 to as low as 0.54. Figure 2 shows a scatter plot of the efficiency scores from the ILY-based models against those from the TLE-based model.

Besides the levels of technical efficiency in year 2000 and 2004, we have also estimated technical efficiency change (EFFCH) across the two years. Again, Table 2 shows the average results across countries. Detailed country-specific results are presented in Table 5 where estimates of EFFCH, TC and TFP change are presented for different output measures. Efficiency change figures less than unity indicate a drop in efficiency when the country is assessed against the frontier production function in each time period. An EFFCH of 0.999 in the TLE model means that efficiency falls by 0.1 percent over the two periods. Figures in Table 2 show that on average there is little scale efficiency change (SECH). In fact, not only is the average SECH close to one, but also the SECH for each individual country, implying a CRS production technology for OECD countries.[23]. Assuming CRS means that we can focus on technical efficiency change rather than pure technical efficiency change. It can be seen

[22] For the year 2000 ILY(low-middle income) model, Hungary is the second most inefficient country after Turkey and the US the third. Also, for the year 2000 ILY(low income) model, the US comes third after Turkey.
[23] Estimation results from SFA also indicate that the health production function is of constant returns to scale.

that on average, changes in technical efficiency are not that large; however, the details results reported in Table 5 show that there are substantial variations of EFFCH across countries. In particular, a few countries register large improvement in efficiency for the ILY(low income) model, including Turkey, Italy, Portugal, Poland and Hungary. On the other hand, Slovakia, Spain, New Zealand and Czech Republic show the biggest declines in their efficiency scores from 2000 to 2004.

In the bottom panel of Table 2, we also report summary measures of technical change and TFP change over the periods 2000 to 2004 averaged across the countries. Both of these measures show, on average, a decline over the period when either the ancient or the US Black are used as the baseline population. However, when low and low-middle income populations are used, on average, a TFP growth of 0.2 and 1.5 percent are reported respectively. In contrast the technical change is of an order of 0.3 and 0.4 percent when low income and low-middle income populations are used as the reference group.

In assessing TC and TFP growth it is also important to examine country-specific growth rates. In Table 5, we find significant variation in the performance across countries. In general, most of the developed or high-income countries seem to perform quite well irrespective of which reference population is actually used. An interesting point to note here is that the performance of these countries is significantly higher when low and low-middle income countries are used as the baseline population. For example, the United States shows a technical change of 3.9 percent and 5.5 percent in the ILY(low income) and ILY(low-middle income) models respectively compared to 1.3 and 1.1 percent in the TLE and ILY(ancient) models.

Examining the correlation between results for efficiency change based on TLE against the alternative efficiency change measures based on ILY, we find that these correlations are quite low ranging from 0.29 to 0.48 (see Panel III in Table 4). In contrast the correlations for efficiency change within the different measures of ILY are quite strong ranging from 0.89 to 0.99. Very diverse correlation coefficients are also observed for TFP changes (see Panel IV in Table 4). Figures 3 and 4 are scatter plots of EFFCH and TFP changes of models based on ILYs against that based on TLE. The figures illustrate the use of ILY-based output measures can substantially change the relative performance of health systems across OECD countries as compared to LE-based measure. The reason why TFP changes are also highly diverse amongst the ILY-based measures is that the sample size is small and that all countries increase health expenditure over the time period. In particular, Turkey and Mexico act as outliers because while they form part of the frontier in both periods, for most specifications considered, in the second time period they lie inside the original frontier, This can be seen in Figure 4 which is a diagrammatic illustration of the DEA frontier for the low income baseline for both years 2000 and 2004 assuming CRS. This may therefore suggest either that there has been technical regression in this part of the production function (which is unlikely) or that these countries were not fully efficient in the year 2000 and were still not fully efficient in the year 2004 (which is more likely)[24]. For this reason caution should be exercised when interpreting the results of TE, TC and TFP especially for the low expenditure countries.

4.4 Sensitivity tests

In order to assess the robustness of our general findings, we conduct a number of sensitivity tests. First, AILY can be constructed using the median survival probability of low income countries in year 2004 instead of year 2000. We find that changing the base year for the construction of AILY has very little effect on every measure. Second, we have added education expenditure as an additional input. For instance, for our preferred output measure, ILY(low income), the new and original efficiency scores have a correlation of 0.74 for 2000 and 0.92 for 2004, and the correlation of efficiency change is 0.50. This result is mainly

[24] Another possible explanation may be that the index used to adjust health care expenditure so that it was in constant dollars may not have been an accurate reflection of the actual increases in the price of healthcare inputs.

driven by the fact that one or two more countries that were previously close to the frontier are now on the frontier, which is not surprising because with a small sample adding additional inputs in the DEA can increase the number of peer countries on the frontier and this substantially change the findings. We also prefer not to include education expenditure in the models because in the SFA estimations undertaken, the education expenditure variable is highly correlated with health expenditure and is not always significant for all model specifications.

In SFA estimation, the model using ILY(low-middle income) does not converge[25] with and without education expenditure. Also, while the models based on TLE, ILY(ancient) and ILY(US Black) do not reject the hypothesis of a CRS Cobb-Douglas production function at the 1 percent level, the model based on ILY(low income) does not reject the hypothesis only at the 5 percent level.

Overall, it can be concluded that the estimation results do exhibit a degree of sensitivity towards estimation methods and model specifications. The choice of baseline population also appears to have significant influence on some of the results. This type of sensitivity, however, is not uncommon in efficiency analysis. And it should be reiterated that, the objective of the paper is to improve on the methodology of measuring health system output and inputs, not the methodology of efficiency analysis itself. We believe that the results in particular for the EFFCH based on the DEA methodology are robust and provide interesting insights into the change in performance of health systems for the OECD countries.

5. Discussion and Conclusion

Measuring health system efficiency is a highly desirable yet difficult task. This paper has illustrated the importance of considering a country's demographic structure when measuring both health output and health input. Firstly, a new, improved health system output measure,

[25] This means that the program (Stata) fails to find the maxima of the likelihood function.

namely the Incremental Life Years (ILY), and its variant Relative Incremental Life Years (RILY) that takes demographic structures into consideration were constructed. Second, a related health system input measure, namely the Average Incremental Life Years (AILY) that arguably is a better proxy for health services demand than total population size, was also constructed. These new measures were then applied to estimating health system efficiency of OECD countries to demonstrate the differences compared with those results obtained from the regularly used LE output measure and health expenditure per capita as the input measure.

A fundamental conceptual difference between ILY and the traditional output measure—life expectancy at birth—is that ILY measure the value-added of the health system rather than the health status of population. This is achieved by subtracting baseline survival probabilities from those observed. Measuring against some baseline is a common practice in many economic indicators such as price indexes or real GDP. The empirical results reported in sections 2 and 3 show that correcting for the baseline survival probabilities can result in significant differences, not only in the measures of health output and health system efficiency, but also in the gauging efficiency and productivity changes.

The difference in the efficiency scores (and their correlations) and change in efficiency scores across the ILY models and TLE model has demonstrated that large errors could result when LE (gross output) rather than ILY (value-added based measures) are used to measure health system output. Simply adjusting LE by subtracting a baseline value from it, instead of doing so for each individual age-sex group as in the case of ILY, is not advisable. While it may seem to be a convenient short cut, it fails to account for differences in demographic structures—another advantage of ILY. As some countries face rapidly aging populations these differences in demographic structures play an increasingly important role in terms of explaining differences in health system expenditure and outputs across countries; LE and health expenditure per capita do not capture these demographic factors.

While the ILY and RILY measures constructed in this paper are arguably better health system output measures than LE, the proxy nature of baseline survival probabilities means that their values are likely to be subject to some errors. In comparison, using LE as a health output measure appears to be clean and easy to understand, but this simplicity masks the fact that it is equivalent to setting the baseline survival probabilities to zero, using the current mortality of older age groups in the countries to weight each life saved and assumes a particular demographic structure, all of which are unrealistic and extreme cases compared to the four baselines and methods used in this paper. Simply put, the approach proposed in this paper, in Warren Buffett's famous words, is that "it is better to be approximately right than precisely wrong."

It should be stressed that the development of ILY and RILY in this paper should be considered a step on the journey towards achieving more accurate measures of health system output and efficiency measurement, rather than the finishing line. Its primary function is to establish the importance of measuring health output as value-added, accounting for demographic factors in output measurement and, therefore, making the assumptions underlying health system output measurement explicit rather than implicit. If these elements are properly acknowledged, then efforts can focus on better measures of counterfactual survival probabilities and more appropriate assumptions and weightings, which will eventually lead to more accurate measures of health system output and efficiency.

Acknowledgements

The authors wish to thank Geertje Klein Goldwijk for her invaluable suggestions regarding data and references sources, and acknowledge the comments from the participants of seminars at the University of Queensland, the Australian National University, the University of Groningen and the Swiss Economic Institute for many useful comments and suggestions. All remaining errors are ours.

Reference

Acsádi György Y, and János Nemeskéri, *History of Human Life Span and Mortality*, Budapest: Akademiai Kiado, 1970.

Coale, Ansley J., Paul Demeny, and Barbara Vaughan (1983). *Regional Model Life Tables and Stable Population*. *Studies in Population*, Academic Press, New York. 2nd edition.

Coelli, Timothy J, DS Prasada Rao, Christopher J O'Donnell, and George E Battese (2005) *An Introduction to Efficiency and Productivity Analysis*, 2nd edn., Springer, New York.

Cutler, David, Angus Deaton, and Adriana Lleras-Muney (2006) "The determinants of mortality" *Journal of Economic Perspective* 20(3): 97-120.

Elmendorf, Douglas W, and Louise M Sheiner (2000) "Should America save for its old age? Fiscal policy, population aging, and national saving." *Journal of Economic Perspectives* 14(3): 57-74.

Evans, DB, L Bendib L, A Tandon et al. (2000) "Estimates of income per capita, literacy, education attainment, absolute poverty, and income Gini coefficients for the World Health Report 2000" World Health Organization, *GPE Discussion Paper*, No.29.

Evans, David B, Ajay Tandon, Christopher J L Murray, Jeremy A Lauer (2000) "The comparative efficiency of national health systems in producing health: An analysis of 191countries" World Health Organization, *GPE Discussion Paper*, No.29.

Evans, David B, Ajay Tandon, Christopher J L Murray, Jeremy A Lauer (2001) "Comparative efficiency of national health systems: cross national econometric analysis" *BMJ* 323: 307-310.

Green, William (2004) “Distinguishing between heterogeneity and inefficiency: stochastic frontier analysis of the World Health Organization’s panel data on national health care systems” *Health Economics* 13(10): 959-980.

Green, William (2005) “Reconsidering heterogeneity in panel data estimators of the stochastic frontier model” *Journal of Econometrics* 126: 269-303.

Häkkinen, Unto, Isabelle Joumard (2007) “Cross-country analysis of efficiency in OECD health care sectors: options for research” OECD Economics Department Working Papers, No. 554.

Hollingsworth, Bruce (2003), “Non-parametric and parametric applications measuring efficiency in health care” *Health Care Management Science* 6(4): 203-218.

Hopkins, Keith (1966) “On the probable age structure of the Roman population” *Population Studies* 20(2): 245-264.

McKee, Martin (2001) “Measuring the efficiency of health systems” *BMJ* 323: 295-296.

OECD 2002, *Measuring Up: Improving health system performance in OECD countries*. The OECD Health Project, Paris.

OECD 2004, *Toward High-Performing Health Systems*. The OECD Health Project, Paris.

Rimm, E.B., E.L. Giovannucci, W.C. Willett, G.A. Colditz, A. Ascherio, B. Rosner, and M.J. Stampfer (1991) “Prospective study of alcohol consumption and risk of coronary disease in men” *Lancet* 338(8765): 464-468.

Roseveare, D. et al. (1996) “Ageing populations, pension systems and government budgets: simulation for 20 OECD countries” OECD Economics Department Working Papers, No. 168, OECD Publishing.

Williams, Alan (2001) “Science or marketing at WHO? A commentary on ‘World Health 2000” *Health Economics*, 10(2): 93-100.

Table 1 Life Expectancy at birth (LE), Incremental Life Years (ILY), Relative Incremental Life Years (RILY)*

2000

		Level 3		US Black		Low income		Low-middle income	
Country	LE	ILY	RILY	ILY	RILY	ILY	RILY	ILY	RILY
Australia	79.80	14623875	1.014	11666271	1.023	4659231	1.065	2960303	1.111
Austria	78.50	5761841	1.006	4532145	1.010	1848280	1.027	1197111	1.045
Belgium	77.80	7493368	0.999	5871921	0.999	2368280	0.996	1498713	0.994
Canada	79.10	22893213	1.011	18183073	1.018	7242970	1.050	4615269	1.084
Czech Rep	75.05	6718914	0.970	5262605	0.951	1874648	0.868	1070474	0.782
Denmark	76.85	4000922	0.992	3125660	0.987	1223082	0.965	745160	0.941
Finland	77.60	3733754	0.996	2943116	0.994	1160965	0.985	725343	0.976
France	79.10	46844296	1.010	37107174	1.017	15170863	1.042	9692909	1.070
Germany	77.90	57468772	1.000	44960056	1.000	18407273	1.002	11925623	1.004
Greece	78.05	7413655	1.002	5825608	1.001	2380000	1.009	1541366	1.016
Hungary	71.55	6525868	0.927	4963405	0.883	1467166	0.665	617770	0.443
Iceland	79.60	223850	1.013	179315	1.021	70831	1.065	44107	1.115
Ireland	76.40	2866409	0.989	2272891	0.982	831195	0.952	479651	0.917
Italy	79.00	40949942	1.012	32162685	1.019	13731847	1.053	9187367	1.085
Japan	81.15	91493362	1.030	73627023	1.051	32366620	1.131	22408766	1.213
Korea	74.65	34454172	0.978	27384499	0.966	9060183	0.891	4769536	0.804
Luxembourg	78.15	334017	1.002	263357	1.004	104729	1.011	65332	1.020
Mexico	74.20	87052353	0.965	69182593	0.948	20315683	0.794	8236843	0.596
Netherlands	78.05	11885789	1.003	9390490	1.004	3733009	1.015	2337220	1.026
New Zealand	78.25	2971287	1.003	2357796	1.005	899233	1.012	542536	1.020
Norway	78.65	3468533	1.007	2726308	1.011	1109405	1.031	702814	1.053
Poland	73.85	25582975	0.961	20028103	0.938	6573167	0.815	3422986	0.686
Portugal	76.40	7415047	0.985	5810535	0.975	2206343	0.930	1318754	0.885
Slovakia	73.25	3589367	0.958	2803643	0.932	900732	0.799	453645	0.657
Spain	78.75	28219099	1.008	22284223	1.014	9214257	1.037	6028675	1.061
Sweden	79.70	6501167	1.019	5129715	1.030	2216403	1.084	1492231	1.136
Switzerland	79.80	5343307	1.018	4232934	1.029	1777516	1.078	1177140	1.130
Turkey	68.95	50598444	0.933	39370057	0.897	9699595	0.629	2527311	0.295
United Kingdom	77.00	43575670	0.993	33924702	0.987	13283212	0.968	8165843	0.946
United States	76.75	214701499	0.991	169521602	0.986	63167806	0.951	36751929	0.914
Average	77.13	28156825	0.99	22236450	0.99	8302151	0.96	4890091	0.93
Coefficient of variation	0.03	1.53	0.02	1.53	0.04	1.54	0.12	1.57	0.23

2004

		Level 3		US Black		Low income		Low-middle income	
Country	LE	ILY	RILY	ILY	RILY	ILY	RILY	ILY	RILY
Australia	80.75	15119967	1.023	12113460	1.038	4998874	1.104	3278685	1.177
Austria	79.30	5738031	1.015	4524173	1.024	1914342	1.064	1281535	1.105
Belgium	78.45	7532311	1.006	5917583	1.009	2452980	1.026	1593157	1.042
Canada	80.20	23713938	1.021	18956217	1.034	7835199	1.093	5165166	1.154
Czech Rep	75.90	6839720	0.978	5356533	0.963	1997350	0.904	1190825	0.844
Denmark	77.65	4005050	0.999	3142395	0.997	1258038	0.996	790014	0.993
Finland	78.70	3767114	1.007	2993043	1.012	1228710	1.029	797644	1.048
France	79.70	48169532	1.015	38256981	1.027	15957956	1.067	10400165	1.112
Germany	79.10	57574302	1.013	45243223	1.021	19411073	1.057	13117310	1.092
Greece	79.20	7636187	1.014	6014830	1.021	2574561	1.061	1726048	1.100
Hungary	72.80	6581228	0.941	5033177	0.905	1627646	0.736	793042	0.564
Iceland	80.65	229250	1.021	184245	1.033	74693	1.101	48076	1.177
Ireland	77.95	3067440	1.002	2447269	1.003	938225	1.012	569018	1.022
Italy	80.65	41941296	1.029	33131451	1.048	14857622	1.123	10345129	1.196
Japan	82.20	92954258	1.041	74512849	1.071	34301231	1.176	24516510	1.279
Korea	76.75	33216540	0.993	26550662	0.988	9419248	0.965	5515935	0.940
Luxembourg	78.50	348127	1.006	274833	1.009	110579	1.027	70057	1.045
Mexico	74.40	88878256	0.969	70711477	0.953	21369796	0.818	9252877	0.648
Netherlands	79.20	12103917	1.013	9626474	1.020	3944452	1.059	2561158	1.099
New Zealand	79.65	3090546	1.013	2469032	1.021	976584	1.057	611698	1.098
Norway	79.65	3463892	1.016	2743096	1.027	1146536	1.075	751557	1.125
Poland	74.90	25539029	0.970	20095399	0.952	7009113	0.862	3917345	0.769
Portugal	77.55	7611536	0.996	5962710	0.993	2373900	0.982	1488881	0.970
Slovakia	74.15	3648773	0.965	2859419	0.943	968701	0.838	518224	0.726
Spain	80.10	29312514	1.021	23170165	1.036	10024013	1.092	6779466	1.150
Sweden	80.55	6628865	1.027	5267579	1.044	2329014	1.118	1599967	1.192
Switzerland	80.80	5401473	1.027	4303476	1.045	1858869	1.121	1267417	1.199
Turkey	71.00	52181100	0.948	40918650	0.920	11247346	0.713	4115887	0.463
United Kingdom	78.65	43991332	1.009	34705118	1.014	14285371	1.039	9263626	1.064
United States	77.60	224390507	0.997	177944558	0.997	67910165	0.982	40709911	0.968
Average	78.22	28822534	1.00	22847669	1.01	8880073	1.01	5467878	1.01
Coefficient of variation	0.03	1.55	0.02	1.55	0.04	1.54	0.11	1.55	0.20

* The estimations of RILY are based on the median survival probability observed in year 2000.

Table 2 Model specification and average results

	Models				
	TLE	ILY(ancient)	ILY(US Black)	ILY(low income)	ILY(low-middle income)
Inputs					
Total Health Expenditure*	X	X	X	X	X
Total population	X				
AILY (ancient)		X			
AILY (US Black)			X		
AILY (low income)				X	
AILY (lower-middle income)					X
Output					
TLE	X				
ILY (ancient)		X			
ILY (US Black)			X		
ILY (low income)				X	
ILY (lower-middle income)					X
Average Results					
Technical efficiency (TE) 2000	0.937	0.979	0.964	0.914	0.863
Technical efficiency (TE) 2004	0.972	0.977	0.961	0.913	0.868
Technical efficiency change (EFFCH)	0.999	0.998	0.997	0.999	1.011
Technological change (TC)	0.996	0.995	0.998	1.003	1.004
Pure technical efficiency change (PECH)	0.998	0.999	0.998	1.000	1.010
Scale efficiency change (SECH)	1.000	1.000	0.999	0.999	1.001
Total Factor Productivity (TFP) change	0.995	0.993	0.995	1.002	1.015

* Total health expenditure for models D and E are net of the total expenditures calculated based on the median expenditure of the corresponding income group.

Table 3 Technical Efficiency (TE) Scores based on Data Envelopment Analysis

	Year 2000					Year 2004				
	TLE	ILY(ancient)	ILY(US Black)	ILY(low income)	ILY(low-mid income)	LE	ILY(ancient)	ILY(US Black)	ILY(low income)	ILY(low-middle income)
Australia	0.983	0.985	0.974	0.941	0.916	0.982	0.982	0.969	0.939	0.920
Austria	0.967	0.977	0.962	0.908	0.861	0.965	0.974	0.956	0.905	0.863
Belgium	0.959	0.970	0.951	0.881	0.820	0.954	0.966	0.943	0.872	0.815
Canada	0.975	0.981	0.969	0.928	0.894	0.976	0.980	0.966	0.929	0.902
Czech Rep	0.982	0.979	0.963	0.925	0.902	0.973	0.973	0.952	0.898	0.852
Denmark	0.947	0.963	0.939	0.853	0.776	0.945	0.959	0.931	0.847	0.776
Finland	0.970	0.979	0.965	0.916	0.866	0.965	0.974	0.959	0.898	0.840
France	0.975	0.980	0.968	0.921	0.882	0.970	0.975	0.959	0.907	0.869
Germany	0.960	0.971	0.952	0.886	0.828	0.962	0.973	0.954	0.899	0.853
Greece	0.962	0.972	0.953	0.892	0.838	0.964	0.973	0.954	0.903	0.860
Hungary	0.942	0.941	0.902	0.733	0.546	0.937	0.939	0.900	0.744	0.588
Iceland	0.981	0.984	0.971	0.942	0.920	0.981	0.980	0.965	0.937	0.920
Ireland	0.950	0.972	0.953	0.871	0.768	0.948	0.963	0.937	0.861	0.798
Italy	0.974	0.982	0.970	0.930	0.895	0.981	0.988	0.979	0.955	0.935
Japan	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Korea	0.989	0.999	0.996	1.000	1.000	1.000	0.999	0.996	1.000	1.000
Luxembourg	0.963	0.973	0.955	0.894	0.841	0.955	0.966	0.943	0.873	0.817
Mexico	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Netherlands	0.962	0.974	0.956	0.897	0.846	0.964	0.972	0.953	0.901	0.859
New Zealand	0.984	0.994	0.988	0.965	0.921	0.980	0.987	0.977	0.934	0.868
Norway	0.969	0.977	0.962	0.912	0.869	0.969	0.976	0.959	0.914	0.880
Poland	0.990	0.987	0.976	0.968	1.000	0.996	0.989	0.980	0.983	1.000
Portugal	0.962	0.972	0.953	0.885	0.817	0.966	0.975	0.957	0.900	0.843
Slovakia	0.981	0.983	0.969	0.945	0.945	0.971	0.975	0.956	0.899	0.833
Spain	0.996	0.996	0.991	0.992	1.000	0.982	0.987	0.977	0.948	0.918
Sweden	0.982	0.989	0.981	0.958	0.937	0.980	0.986	0.975	0.951	0.931
Switzerland	0.983	0.988	0.979	0.953	0.932	0.983	0.986	0.976	0.954	0.937
Turkey	1.000	0.967	0.946	0.792	0.510	1.000	1.000	0.987	0.907	0.772
UK	0.956	0.972	0.952	0.883	0.811	0.957	0.969	0.947	0.884	0.832
US	0.946	0.961	0.938	0.841	0.754	0.944	0.957	0.931	0.835	0.756

Table 4 Correlation of results across models

Panel I: Technical Efficiency (TE) 2000

	TLE	ILY(ancient)	ILY(US Black)	ILY(low income)	ILY(low-middle income)
TLE	1.00				
ILY(ancient)	0.82	1.00			
ILY(US Black)	0.82	1.00	1.00		
ILY(low income)	0.70	0.97	0.96	1.00	
ILY(low-middle income)	0.54	0.87	0.87	0.97	1.00

Panel II: Technical Efficiency (TE) 2004

	TLE	ILY(ancient)	ILY(US Black)	ILY(low income)	ILY(low-middle income)
TLE	1.00				
ILY(ancient)	0.97	1.00			
ILY(US Black)	0.96	0.99	1.00		
ILY(low income)	0.91	0.93	0.96	1.00	
ILY(low-middle income)	0.80	0.83	0.87	0.97	1.00

Panel III: Technical Efficiency Change (EFFCH)

	TLE	ILY(ancient)	ILY(US Black)	ILY(low income)	ILY(low-middle income)
TLE	1.00				
ILY(ancient)	0.43	1.00			
ILY(US Black)	0.48	0.99	1.00		
ILY(low income)	0.47	0.97	0.96	1.00	
ILY(low-middle income)	0.29	0.92	0.89	0.96	1.00

Panel IV: Total Factor Productivity (TFP) Change

	TLE	ILY(ancient)	ILY(US Black)	ILY(low income)	ILY(low-middle income)
TLE	1.00				
ILY(ancient)	0.95	1.00			
ILY(US Black)	0.95	0.99	1.00		
ILY(low income)	0.76	0.75	0.81	1.00	
ILY(low-middle income)	0.11	0.09	0.18	0.68	1.00

Table 5 Technical Efficiency Change (EFFCH), Technical Change (TC) and Total Factor Productivity Change (TFPCH) over 2000 to 2004

	TLE			ILY(ancient)			ILY(US Black)			ILY(low income)			ILY(low-middle income)		
	EFFCH	TC	TFPCH	EFFCH	TC	TFPCH	EFFCH	TC	TFPCH	EFFCH	TC	TFPCH	EFFCH	TC	TFPCH
Australia	0.999	1.013	1.012	0.998	1.008	1.006	0.995	1.014	1.009	0.998	1.039	1.037	1.004	1.055	1.059
Austria	0.997	1.013	1.010	0.997	1.011	1.008	0.995	1.019	1.014	0.997	1.039	1.036	1.002	1.055	1.058
Belgium	0.995	1.011	1.007	0.996	1.008	1.003	0.991	1.014	1.005	0.990	1.032	1.022	0.994	1.053	1.047
Canada	1.001	1.013	1.014	0.999	1.011	1.010	0.997	1.019	1.016	1.001	1.039	1.041	1.009	1.055	1.065
Czech Republic	0.991	0.994	0.985	0.993	0.997	0.990	0.989	0.995	0.985	0.971	0.978	0.950	0.945	0.962	0.908
Denmark	0.998	1.013	1.010	0.996	1.008	1.004	0.992	1.014	1.005	0.992	1.034	1.026	1.000	1.055	1.055
Finland	0.995	0.998	0.993	0.995	1.000	0.995	0.994	1.000	0.993	0.981	0.995	0.977	0.970	0.995	0.964
France	0.995	1.013	1.008	0.995	1.009	1.004	0.991	1.016	1.006	0.986	1.039	1.024	0.985	1.055	1.039
Germany	1.002	1.013	1.015	1.002	1.011	1.013	1.002	1.019	1.021	1.015	1.039	1.055	1.031	1.055	1.088
Greece	1.002	1.003	1.005	1.001	1.004	1.005	1.001	1.008	1.009	1.012	1.014	1.026	1.026	1.018	1.045
Hungary	0.994	0.994	0.988	0.998	0.997	0.996	0.998	0.996	0.994	1.016	0.977	0.993	1.077	0.961	1.035
Iceland	1.000	1.013	1.013	0.997	1.011	1.008	0.994	1.019	1.012	0.995	1.039	1.034	1.000	1.055	1.055
Ireland	0.999	1.002	1.001	0.990	1.004	0.994	0.983	1.007	0.990	0.988	1.008	0.996	1.039	1.006	1.046
Italy	1.008	1.005	1.013	1.006	1.005	1.011	1.009	1.010	1.019	1.026	1.015	1.042	1.044	1.018	1.063
Japan	1.000	1.002	1.002	1.000	1.003	1.003	1.000	1.006	1.006	1.000	1.008	1.008	1.000	1.006	1.006
Korea	1.011	0.994	1.005	1.000	0.998	0.998	1.000	0.996	0.996	1.000	0.977	0.977	1.000	0.961	0.961
Luxembourg	0.992	1.013	1.004	0.993	1.011	1.004	0.987	1.019	1.006	0.977	1.039	1.015	0.971	1.055	1.025
Mexico	1.000	0.904	0.904	1.000	0.864	0.864	1.000	0.864	0.864	1.000	0.858	0.858	1.000	0.849	0.849
Netherlands	1.002	1.010	1.012	0.999	1.006	1.005	0.997	1.011	1.008	1.004	1.029	1.033	1.016	1.055	1.071
New Zealand	0.995	0.997	0.993	0.993	0.996	0.989	0.989	0.994	0.984	0.968	0.990	0.958	0.943	1.001	0.944
Norway	1.000	1.013	1.013	0.999	1.011	1.010	0.997	1.019	1.016	1.003	1.039	1.042	1.013	1.055	1.068
Poland	1.007	0.974	0.980	1.002	0.998	1.000	1.004	0.997	1.000	1.016	0.977	0.992	1.000	0.920	0.920
Portugal	1.004	0.992	0.996	1.003	0.996	0.999	1.004	0.994	0.998	1.017	0.978	0.994	1.032	0.961	0.991
Slovak Republic	0.989	0.977	0.967	0.991	0.998	0.989	0.987	0.996	0.983	0.950	0.979	0.931	0.881	0.929	0.819
Spain	0.986	0.999	0.985	0.991	1.000	0.991	0.986	1.002	0.987	0.956	0.998	0.954	0.918	0.997	0.915
Sweden	0.998	1.010	1.008	0.997	1.008	1.005	0.995	1.014	1.009	0.993	1.030	1.023	0.994	1.047	1.041
Switzerland	1.000	1.013	1.013	0.999	1.011	1.010	0.997	1.019	1.016	1.000	1.039	1.040	1.006	1.055	1.061
Turkey	1.000	0.885	0.885	1.034	0.861	0.891	1.043	0.860	0.897	1.145	0.845	0.968	1.512	0.831	1.256
United Kingdom	1.001	1.002	1.003	0.996	1.004	1.000	0.995	1.007	1.002	1.001	1.008	1.009	1.026	1.007	1.032
United States	0.998	1.013	1.011	0.996	1.011	1.007	0.992	1.019	1.011	0.993	1.039	1.032	1.003	1.055	1.058

Figure 1 Relative incremental life years (RILY) of four baseline populations against life expectancy at birth, pooled data for 2000 and 2004

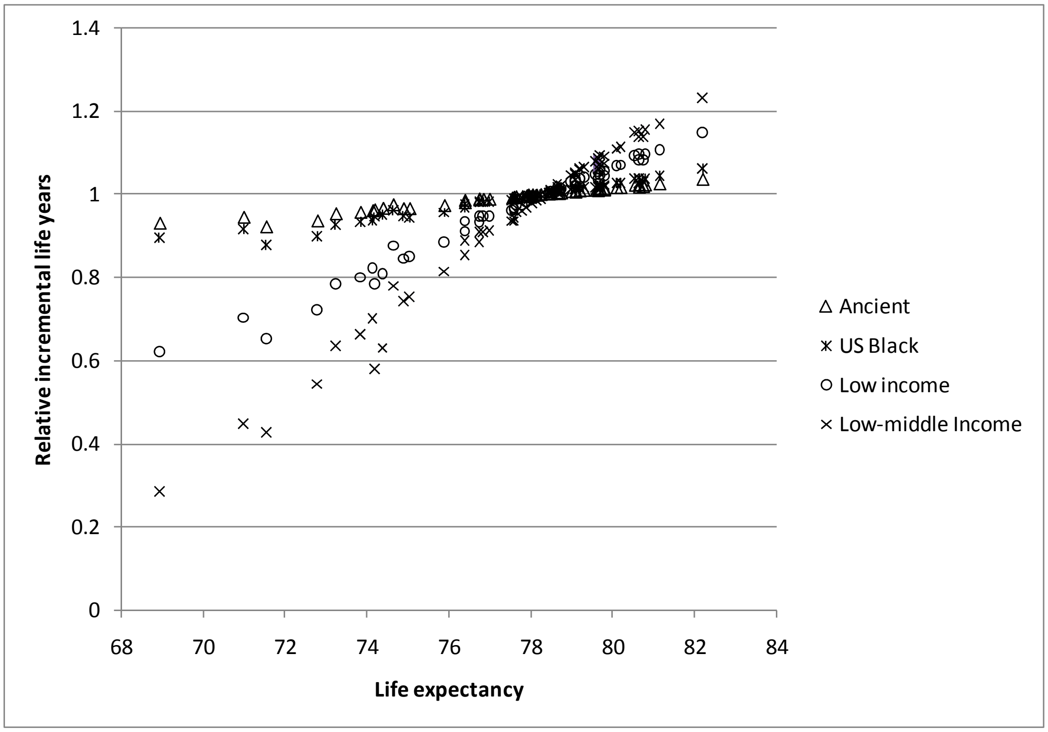


Figure 2 Technical efficiency score of model based on ILYs against that based on total life expectancy, pooled data for 2000 and 2004

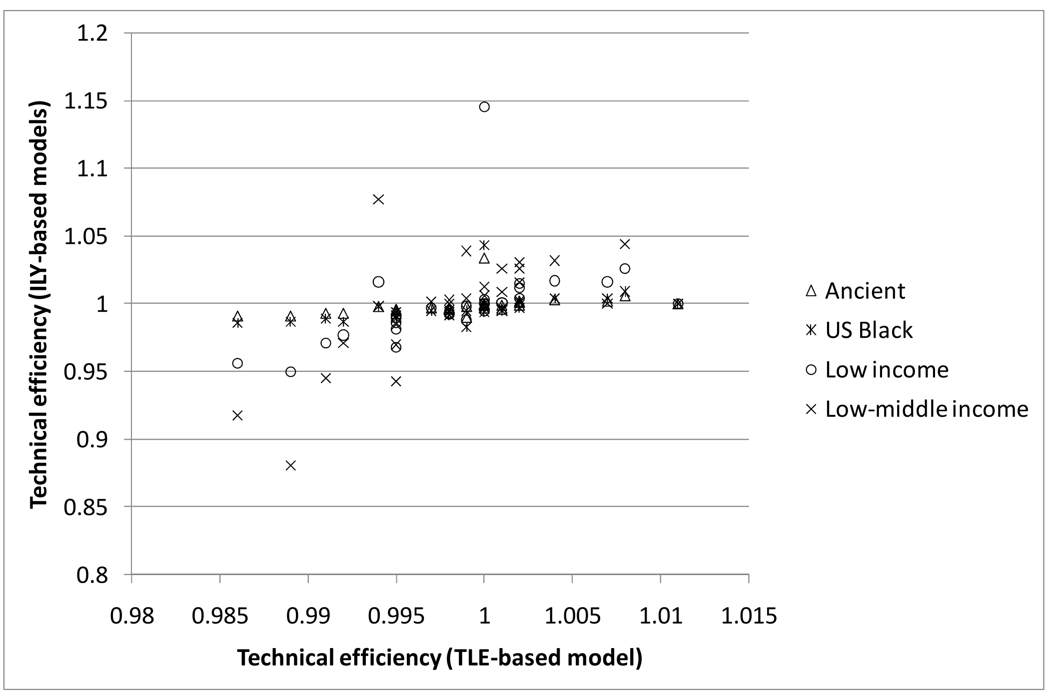


Figure 3 Technical efficiency change of model based on ILYs against that based on total life expectancy, pooled data for 2000 and 2004

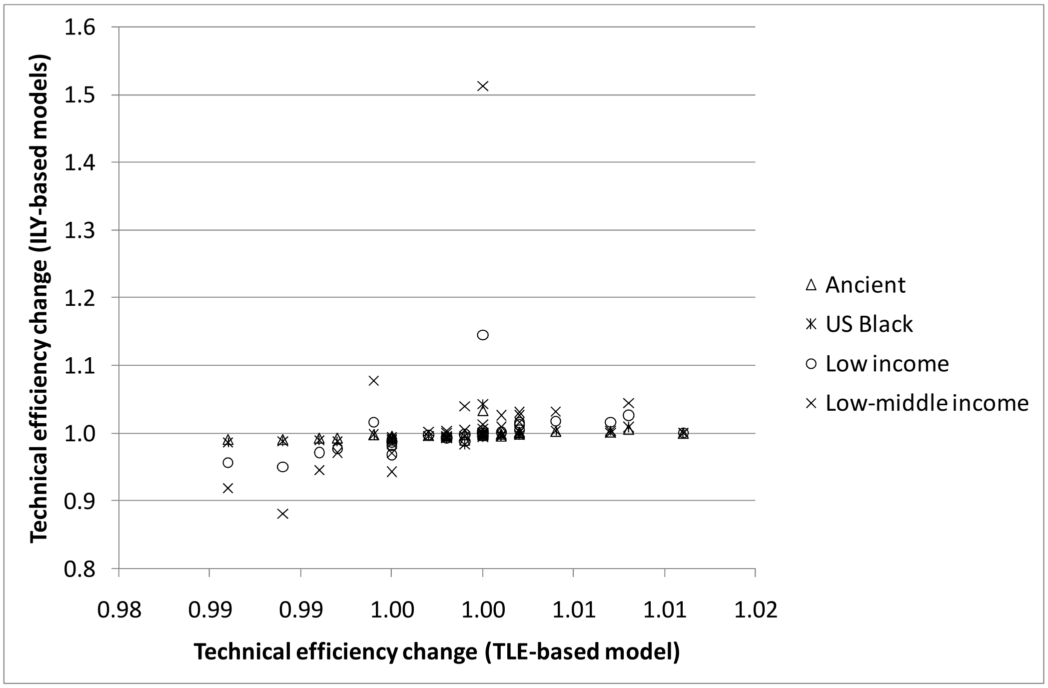


Figure 4 Total factor productivity change of model based on ILYs against that based on total life expectancy, pooled data for 2000 and 2004

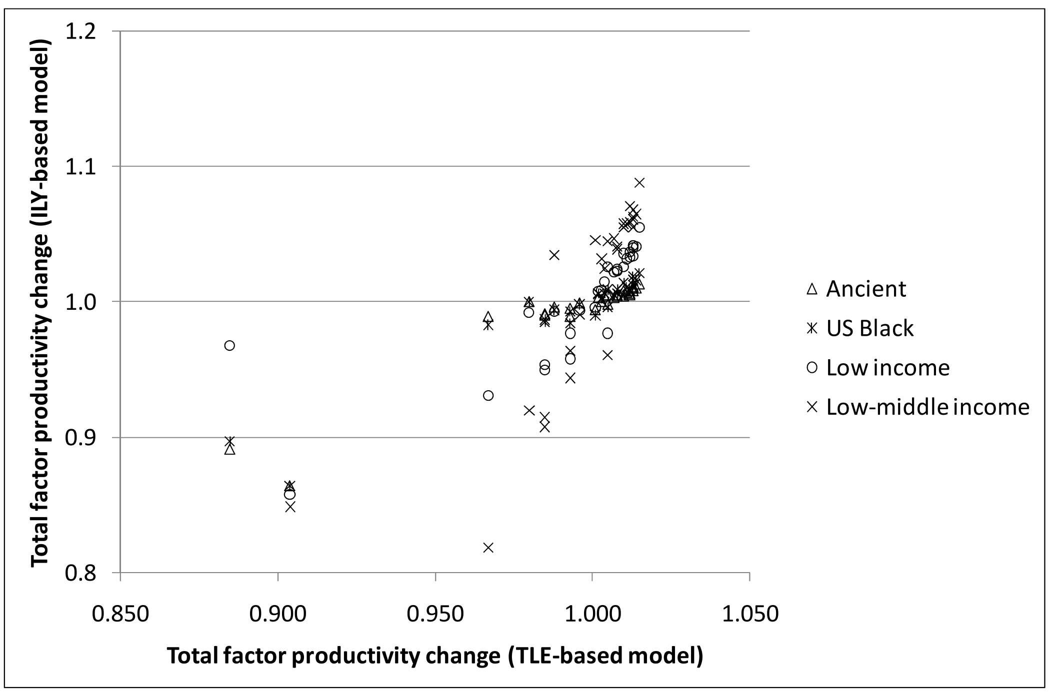
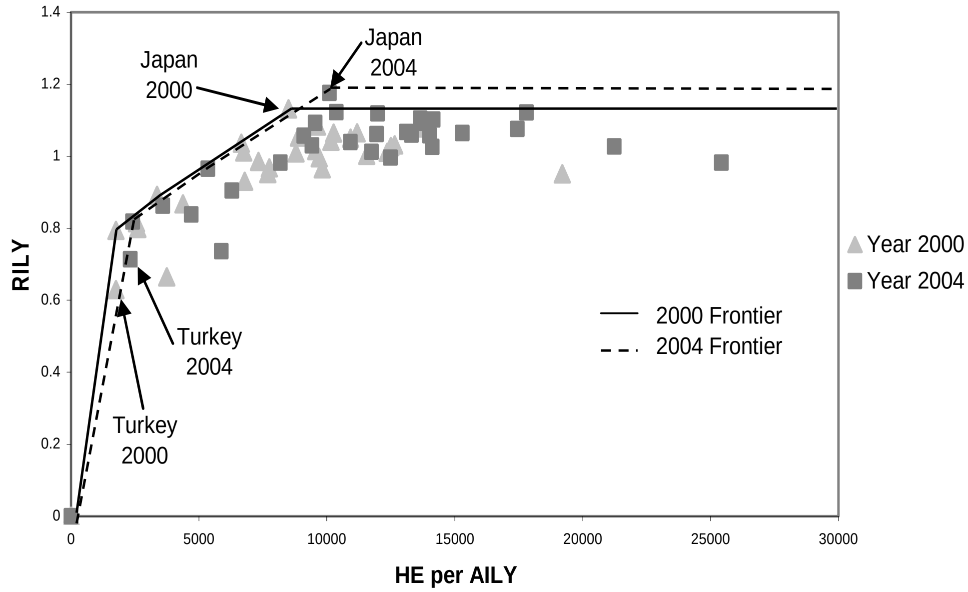


Figure 5 DEA frontier for the years 2000 and 2004 assuming CRS using Low income baseline population



Appendix A. List of low income countries used to construct the "low income" baseline population

Bangladesh
Comoros
India
Korea, Dem Rep.
Kyrgyz Republic
Lao PDR
Liberia
Madagascar
Mauritania
Mongolia
Myanmar
Nepal
Niger
Pakistan
Sao Tome and Principe
Senegal
Tajikistan
Uzbekistan
Vietnam
Yemen, Rep.

Appendix B. List of lower-middle income countries used to construct the "low-middle income" baseline population[a]

Albania	El Salvador	Nicaragua
Algeria	Fiji	Paraguay
Armenia	Georgia	Peru
Azerbaijan	Guatemala	Philippines
Belarus	Indonesia	Samoa
Bhutan	Iran, Islamic Rep.	Sri Lanka
Bolivia	Iraq	Syrian Arab Republic
Bosnia and Herzegovina	Jordan	Thailand
Cape Verde	Kiribati	Tonga
China	Macedonia, FYR	Tunisia
Colombia	Maldives	Turkmenistan
Cuba	Marshall Islands	Ukraine
Dominican Republic	Micronesia Fed. Sts.	Vanuatu
Ecuador	Moldova	
Egypt, Arab Rep.	Morocco	

(a) The list of countries plus those in Appendix A are used jointly to construct the "low-middle income" baseline population.