



American Enterprise Institute for Public Policy Research

The OECD's Study on Health Status Determinant: Roles of Lifestyle, Environment, Health-Care Resources and Spending Efficiency: An Analysis

H.E. Frech III

University of California, Santa Barbara

AEI WORKING PAPER #145, FEBRUARY 6, 2009
www.aei.org/workingpapers
<http://www.aei.org/publication29340>
#23888

The OECD's Study on Health Status Determinant: Roles of Lifestyle, Environment, Health-Care Resources and Spending Efficiency: An Analysis*

Jan. 30, 2009

H.E. Frech III

**Economics Department
University of California, Santa Barbara
Santa Barbara, CA 93106**

frech@econ.ucsb.edu

*Thanks are due to the Office of the Assistant Secretary for Planning and Evaluation of the U.S. Department of Health and Human Services for partial support of earlier work on this project. The views expressed here are not necessarily those of the Department. Thanks are also due for helpful comments on earlier versions of this paper by John Drabek, John Hoff and Peter Zweifel.

I. Introduction

The OECD has undertaken an ambitious large-scale statistical analysis of the determinants of health and the relative efficiency of the health care systems of various OECD member countries (Journard, André, Nicq and Chatal 2008). The Report makes a useful contribution to a continuing stream of literature that focuses on health outcomes, rather than cost. While the primary emphasis of the Report is on new statistical analysis, it includes a valuable, though spotty, literature review.¹

The predominate technical approach is econometric, using panel regression. An operations research technique called data envelopment analysis (DEA) is also used to a limited extent. The regression technique is applied to annual data on 16 OECD countries from at 1981 to 2003²

The authors describe three major findings:

¹ To round out the literature review, see H. E. Frech III and Richard Miller (1999), Miller and Frech (2004), William Greene (2004), James Shaw, William Horrace and Ronald Vogel (2005), Zynep Or, Jia Wang and Dean Jamison (2005) and Alan Garber and Jonathan Skinner (2008).

² The issue of sample size is a bit confusing. The authors state at one point that the analysis is based on 23 countries from 1981 to 2003 (Journard, André, Nicq and Chatal 2008, p. 20), but then state in a footnote to that sentence (2008, fn. 6a, p. 21) that seven countries were excluded and that some countries' time series were not of full length because of data problems. This makes sense, since the largest reported sample size is 325 (2008, p. 23). Complete data on 23 countries for 23 years would generate a sample size of 526. The sample size for the DEA analysis is 29 countries, but it is for only a single year (2004), that is later than the panel regression years. It uses a smaller set of variables (2008, p. 37). Also, some variables may have been interpolated.

1. *Mortality/longevity indicators are imperfect indicators but remain the best available proxies for the population's health status.*
2. *Health care plays an important role in explaining health status changes over time and cross-country differences.*
3. *Health care spending is not producing the same value for money across countries* (Joumard, André, Nicq and Chatal 2008, p. 6, italics in original).

Since health status is inherently a subjective and arguable concept, point No. 1 is a judgment and a point of view. Points Nos. 2 and 3 are based on an interpretation of the statistical results in this study. The U.S. health care system comes out as apparently relatively inefficient in some, but not all, of the analyses. However, there are problems in the analysis that undermine the statistical arguments for the point No. 2 and especially for point No. 3. The measures of the relative efficiency of different health care systems, particularly the U.S. system, are not robust, are contaminated by measurement and estimation problems and depend crucially on a strong and unreasonable assumption.

This Report presents econometric and operations research estimates of the productivity of health care in producing good health. The focus is on two issues: 1) The productivity of health care in improving health and 2) The relative efficiency of the health care systems of different countries. Technically, the primary emphasis is on econometric (panel regression) methods, rather than the operations research technique of data envelopment analysis (DEA). The panel method uses dummy variables for each country to control for all time-invariant differences across countries. These are called unit-specific fixed effects. There are no time fixed effects, so this is not a full fixed-effects approach. Further, there are no time trend variables. The Report interprets these estimated country-specific coefficients as the main part of the measure of health care efficiency, even though the coefficients pick up all fixed differences across countries, not just efficiency differences.

A. The Production Function Approach

The OECD study takes the household production function approach to determining health status. In this approach, inputs are combined to produce the output, health.³ Logically, the inputs include various types of health care and external factors such as culture, lifestyle choices, genetic makeup, industrial structure, disease prevalence, the transportation system and pollution. This approach is called household production to differentiate it from ordinary commercial manufacturing which takes place in factories (Becker 1971, pp. 165-170). The most important distinguishing feature of *household* production is that productive decisions cannot be separated from the values and tastes of the consumers themselves. An important example is the time preference (impatience) of the consumer. Time preference has been shown to be related to a variety of health behaviors (Fuchs 1980; Robb, Huston and Finke 2008; Zhang and Rashad 2008). Because of the intermingling of values and production, even with identical resources, different households, groups or countries, make different choices and, therefore, end up with different health

³ The work of Michael Grossman (1972) is seminal in this approach. For an excellent presentation, including more recent research, see Peter Zweifel and Freidrich Breyer (1997, pp. 52-172).

outcomes. This fact places limits on how far one can press a policy that promotes equity in health outcomes, rather than one that promotes equity in access to some reasonable level of health care services.⁴

Household production of health is a bit abstract. A more concrete example, often used in teaching, is producing meals at home using various inputs purchased in the market and the time and capital resources of the cook. The skill with which the cook combines the inputs matters for the output. This is why educational achievement is ordinarily considered one of the inputs into household production, including production of health. Health itself lasts into the future and better health both enhances and extends one's life. Therefore, it is considered an element of human capital.

B. Problems with the OECD Analysis

1. Confounding Inputs and Omitted Variables

It is easy to estimate a statistical production relationship that is misleading. The estimates can either overstate or understate the true productivity of an input by confounding the true productivity of the input with other factors. Paradoxically, to avoid that confounding and, therefore, to estimate the productivity of one particular input, one must include all the other important inputs in the estimation process. For example, one might find a strong relationship between education and health if there were no other inputs in the model (i.e. in simple regression or inspecting a scatter plot of the data). But, education is closely related to other inputs, such as income, healthy lifestyle choices, certain types of culture and low pollution. The actual causation may be from these other variables, not education. The problem results from omitting one or more relevant, correlated variables from the analysis. Hence, it is called omitted variable bias.⁵

2. Matching the Concept to the Measure

Omitting confounding variables is an inevitable problem with and limitation of this type of research. At best, there are some factors for which either no data or only crude approximations are available. In this Report, the inherent omitted variable problems is exacerbated by a consistent approach of favoring variables that are widely available, even if they are not the best variables for the concepts. This problem, common in empirical economics, arises because there is a choice of which observable variables to use to represent a conceptually important factor.

3. Measurement Problems

One cannot estimate the productivity of an input unless it is measured well. Nonsystematic measurement error in the output variable, perhaps surprisingly, does not bias the estimates, though it does reduce statistical power. Systematic measurement error in either inputs or outputs causes obvious problems. The

⁴ For a strong advocacy of the inequity-in-health-outcome viewpoint, see WHO (2008) and a commentary on the WHO Report (*Economist* 2008).

⁵ For a textbook exposition of omitted variable bias, see Robert Pindyck and Daniel Rubinfeld (1998, pp. 184-195).

productivity of an input can be seriously underestimated or overestimated. An important and counterintuitive result from econometric analysis is that even unsystematic measurement error in an input biases estimates of its productivity towards zero. Further, measuring one input badly implies that one has not actually held it constant. The consequences are similar to the situation where that variable is omitted altogether from the analysis.⁶ Thus, poor measurement of one variable creates an omitted variable bias in the estimates of other coefficients.

4. Implications for the OECD Study

In our context, estimation of health care productivity and the efficiency of different countries' health care systems, requires adequate statistical controls for other determinants of health. This is necessary to avoid confounding health care use or the efficiency of a country's system with other variables that operate in the society or the economy that are largely outside the health care system, such as healthy lifestyles, favorable cultures, high income, low pollution and good education. It is also necessary to measure variables well. As we will see, this study has problems with both omitted variables and systematic and unsystematic measurement error. Some of these problems could be mitigated with OECD or other available datasets and some could not. The net effect of omitted variables and systematic measurement errors is to bias upward the estimated apparent inefficiency of the U.S. health care system and probably to bias upward the apparent productivity of health care. Further, because many external factors cannot be measured, country-specific health care efficiency probably cannot be isolated. The contrary assumption in the Report is far too sweeping. These issues are taken up below. For ease of interpretation, this paper roughly follows the structure of the Report itself.

II. Measuring Health

A. Measuring Health At the Conceptual Level

As discussed above, health is an aspect of human capital that is produced by household production. According to Michael Grossman (1972, p. 223), "Health care be viewed as a durable capital stock that produces an output of healthy time." Health is the unobservable capital good that produces the healthy time that consumers actually value. The demand for health care is a derived demand, resulting from health care's productivity in producing health.⁷ Importantly, there are many other inputs that help produce health, such as lifestyle and the environment.

Health care is not directly valued, independently of the health that it produces. Indeed, consuming most health care is unpleasant. Health care is one step away from the good that people actually value. Health insurance is removed by another step, since it depends on the productivity of health care in producing health and also on the consumer's subjective and idiosyncratic attitudes towards risk. The fact that health is

⁶ It is easiest to think of this at the extreme. Suppose that a measure of some input was so inaccurate that it was random noise. Such a variable would be given almost no weight in the statistical procedure—in effect, it would be ignored.

⁷ For excellent textbook expositions, see Charles Phelps (2003, pp. 10-12; 59-97) and Sherman Folland, Allan C. Goodman and Miron Stano, (2001, pp. 120-140). For a more advcd treatment, see Peter Zweifel and Fredrich Breyer (1997, pp. 116-124).

not observable creates a problem for studies of and for policy towards health. Health must be proxied or imperfectly measured by something that can be observed.

Luckily, the basic definitions point towards some possible measures. Health produces healthy time. Healthy time cannot exceed total time. Therefore, LE represents the maximum expected healthy time for an individual or a group. Indeed, an important line of theoretical work, closely associated with Isaac Ehrlich (1999), views the consumer as choosing a health level so as to optimally choose his life expectancy (LE). In principle, one might also adjust downward this measure to account for time in poor health. Further, one might try to directly measure healthy time.

B. Equity in Health

The OECD authors note that they do not use any measure of equity in health within populations (Joumard, André, Nicq and Chatal 2008, p. 18). Whether any situation is equitable is, of course, a subjective value judgment. Further, there is an additional problem because health is a result of household production. Health depends on the preferences, values and choices of different individuals and different groups. Nonetheless, there is some degree of agreement that equity concerns equality across particular groups (not typically across individuals) in health. For example Braveman and Gruskin define equity in health as “the absence of systematic disparities in health...between groups with different levels of underlying social advantage (2003, p. 255).” Thus, it is an aspect of equality. The OECD authors state that there is no systematic internationally comparable data of health outcome inequity. But, Eddy van Doorslaer, Christina Masseria and Xander Koolman have studied intergroup differences for an *input*, physician visits. They find that there are income-related differences in physician visits that favor higher income consumers in about half of the OECD countries (van Doorslaer, Masseria and Koolman 2006). Conceptually, this idea of input inequality may be more useful than health inequality.

While studies of inequality within countries are rare, there is an extensive literature on inequality in health *across* countries, including both poor and rich countries. It tells a clear story of greatly increasing equality across countries, measured by LE, over the last 40 years or so. For example, LE in North America improved from 70 to 77 years over the period 1960 to 2000, while it improved much more radically in East Asia, from 42 to 71 years. This remarkable historical convergence has been recently halted by the AIDS epidemic in Africa, starting in the decade of the 90s (Becker, Philipson and Soares 2005, pp. 278, 282; Cutler, Deaton and Lleras-Muney 2006, p. 98). This rapid increase in equality of LE over this longer time period indicates that improving income and technology of the poor countries has led to large health improvements. The amount of the improvement no doubt varies according to varying choices in household production. Because of incomplete and inconsistent data, there are few international comparisons of internal (within-country) health status equity. But, in an important recent study, June O’Neil and David O’Neil find a slightly stronger relationship between income and health status in Canada than in the U.S. (2008, p. 35).

C. Health Measures Considered in the OECD Study

Several observable measures were considered by the OECD authors. All of them focus on some derivative of mortality. All of them are outcome measures that are meant to proxy health, rather than the consumption of health services.

1. Mortality and Life Expectancy

a. Raw (Unadjusted) Measures

The OECD authors describe raw mortality/longevity as including the basic, widely-available measures such as LE at various ages, infant, neonatal and perinatal mortality and premature mortality (PYLL). As they note, these measures are the most commonly available across many countries (Joumard, André, Nicq and Chatal 2008, p. 7) for fairly long time periods. On the other hand, these crude measures do not take account of the quality of life.

One common measure is LE at birth. In the literature, this is often analyzed separately (i.e. in separate regression equations) for males and females, partly because the measured impact of health care on health status in most of the literature is greater for females than for males. Surprisingly, the sex difference in the OECD study is reversed, but it is fairly small (Joumard, André, Nicq and Chatal 2008, pp. 22, 23). Not small at all is the gross difference in LE; female LE is longer than the male LE by years. So, if the data are pooled, there should be a dummy variable controlling for this difference (Miller and Frech 2004, p. 39).^{8,9} If sex is not controlled for by a dummy, varying proportions of females would be confounded with other inputs.

LE at older ages, such as 40, 60 or 65, provides a different measure. It is less affected by the measurement, lifestyle and cultural problems inherent in infant, neonatal and perinatal mortality and in LE at birth. Using LE at later ages reduces, but does not eliminate, the confounding of lifestyle choices, culture and other inputs with health care inputs. Further, these are the ages where a great deal of health care spending is focused. So, these measures are potentially more sensitive to health care inputs.

Premature mortality (PYLL) is another useful measure. The analysis can be separate by sex or dummied to account for sex differences. One advantage stressed by the authors of the OECD Report is that PYLL can be easily and naturally defined by cause of death.¹⁰ This allows the analyst to reduce the confounding of some other external causes with health care inputs and with country-specific effects. Specifically mentioned in the Report are transport accidents, accidental falls, assaults and suicides (Joumard, André, Nicq and Chatal 2008, p. 8). Note that lifestyle and other external causes that raised deaths would also raise health care costs. This is much harder to adjust for and the Report does not attempt it.

Infant, neonatal and perinatal mortality are also commonly-used measures. These measures are nicely described in the Report (Joumard, André, Nicq and Chatal

⁸ A dummy variable takes on a value of one or zero, allowing the constant term in the equation to take on a different value. In this situation, it would allow the constant to differ for LE of men versus women, thereby controlling statistically for sex.

⁹ It is not clear whether the equations in the Report for combined male and female LE at birth and PYLL include a dummy variable for sex.

¹⁰ While there is an easy and natural way to adjust for cause of death, such an adjustment is not perfect for the purposes of this study. For example, if more people with generally risky lifestyles die from accidents, the survivors have better than average lifestyles. Further, as mentioned in the text, risky lifestyles directly raise health care use as well as raising PYLL.

2008, pp. 47, 48). They are typically expressed as the number of deaths per 1,000 live births or in the case of perinatal mortality, per 1,000 possible births. Infant mortality is the number of deaths in the first year per 1,000 live births. Perinatal mortality is the number of deaths in the first week, plus fetal deaths that meet or exceed the minimum standard of 28 weeks of gestation or a weight of 1,000 grams. Neonatal mortality is the number of deaths in the first 28 days per 1,000 live births. Following Nixon and Ullmann (2006), the OECD authors state that these measures are “less influenced by factors not related to the health care system such as education or tobacco consumption” (Joumard, André, Nicq and Chatal 2008, p. 8). As is discussed below, there are strong arguments that the opposite is true—that infant, neonatal and perinatal mortality are especially strongly related to environmental and cultural influences and to lifestyle issues of the parents, especially the mother. Further, as is discussed below and recognized in the Report, basic definitions are not consistent, not even across the rich countries.

b. Adjustment for Quality of Life

Mortality measures can be adjusted for morbidity or quality of life. In practice, the adjustments amount to a weighting on LE such that time spent with a lower quality of life is adjusted downward. For example, a year spent with a migraine headache might be counted as equivalent to only a month of healthy time. So, the year with the migraine would be weighted by a weight of one twelfth or 0.083. The common terms for these adjusted measures are quality-adjusted LE (QALE), disability-adjusted LE (DALE) and health-adjusted LE (HALE). The adjustment are fundamentally a matter of the values (or utility functions) of the individual consumers and thus differ person by person. In practice, some sort of opinion survey of consumers or experts is the source of average weights (Miller and Frech 2004, pp. 20-21). In the QALE and DALE, time lived in poor health of some kind is downweighted or discounted by some weight that is less than 1.0. The HALE is an extreme version, where time lived in poor health is not counted at all. In effect, a weight of 0.0 is used. QALE and DALE are crude adjustments at best, because they ignore differences in consumer values across individuals, groups and countries. Still, they probably get closer to measuring health in the sense that consumers value it than the unadjusted measures do. Further, following consumer values, much health care is directed at quality of life, rather than quantity of life.

c. Adjustment for Disease Prevalence

The OECD authors also state that mortality should ideally be adjusted for disease prevalence (Joumard, André, Nicq and Chatal 2008, p. 6). But, unlike for PYLL, where a partial adjustment is available, there is no natural way to do this sort of adjustment for LE. Any adjustment would require statistical modeling and would be controversial because other adjustments would be defensible and would give different results. Further, the cost or input side should be adjusted as well as the outcome side. A higher prevalence of disease leads to both higher health care use and worse outcomes. Probably, a better way to deal with varying disease prevalence would be explicitly in the overall production function analysis, rather than adjusting the health measures themselves.

The lack of an adjustment for prevalence in the Report biases upward the apparent inefficiency of the U.S. health care system. Recent analysis by Kenneth

Thorpe, David Howard and Katya Galactionova (2007) show that prevalence of ten of the most costly disease conditions is much higher in the U.S. than in Europe. Some of this higher prevalence is due to lifestyle causes, such as obesity or smoking (e.g. diabetes, heart disease, circulatory disease), but some is due to the greater emphasis on preventative screening and related more aggressive treatment of early stages of disease¹¹ Even if there were good data on disease prevalence and treatment across countries, there is a problem with at least some of this data for adjustment of the health outcome variable or for including as inputs in the model. To some extent, disease prevalence is an endogenous result of the health care system, not only an exogenous burdens on it.

2. Other Possible Indicators of Health

The OECD authors consider and wisely reject two other measures that have very different conceptual bases.

a. Sick Leave

The first measure is the amount of sick leave taken by workers. This is a measure of unhealthy time, the reverse of the concept of healthy time that is the most fundamental good, produced by health and other inputs. Thus, the conceptual basis for this is quite solid. However, sick leave, in practice, corresponds to actual sick time very poorly. As the OECD authors carefully note, variation in sick leave across countries is highly influenced by the generosity of sick pay, the type of labor contracts and other institutional issues. To this I would add culture and tradition. A study by Donald Winkler (1980) showed that sick leave taken by U.S. school teachers was strongly affected by reporting and verification policies. A planned policy experiment in Tennessee reduced sick leave taken by state employees by 35 percent from 1980 to 1981, based on a bonus for not using sick leave (Turner 1982). Perhaps time spent on sick leave is best viewed as simply a type of leisure time. The OECD authors note that sick time is very poorly correlated with LE across countries (Joumard, André, Nicq and Chatal 2008, p. 16).

b. Public Satisfaction

The authors also briefly consider survey-based measures of public satisfaction with the health care system as a health outcome measure. They note that public satisfaction is influenced not only by experience with the health care system, but by expectations. Expectations vary across countries and across time. Further, they cite an interesting study by Eddy Adang and George Born (2007) showing that changes in public satisfaction and changes in the following variables (a measure of apparent health care system efficiency, LE, infant mortality and health care expenditures) were not statistically significantly related to public satisfaction. Note that some of the

¹¹ Obesity, in common language, simply means extreme fatness. Laboratory measurement of body fat is very costly. Therefore, the idea of obesity is typically made operational as a body mass index (BMI) over 30. The BMI is mass in kilograms, divided by the square of the height in meters. A BMI of 30 is actually not so extreme. It corresponds to a person who is five feet, five inches tall and weights 180 pounds or to a person five feet, ten inches tall who weighs 207 pounds. In a similar way, overweight is defined as a BMI between 25 and 30. Other measures have been proposed, but they are not commonly available. See Charles Baum and Christopher Ruhm (2008, p. 6).

simple correlations were not so low (the highest being 0.376 between change in infant mortality and changes in public satisfaction), so the lack of significance may be due to a lack of statistical power, rather than well-estimated nearly zero relationships.¹² Along the same line, the authors perform a related, but simpler, analysis themselves. They present a scatter diagram of public satisfaction and health-adjusted LE for 2003, that shows an impressive lack of correlation (Joumard, André, Nicq and Chatal 2008, p. 17). Even if there was a statistical relationship with other output measures, public satisfaction is not conceptually a reasonable measure of health output. It depends critically on expectations, which are influenced by culture and history, the news media, health politics and recent events.

III. Discussion of the OECD Choices of Measuring Health

The actual measures used in the OECD production function are LE at birth for males, females and in total, at age 65 for males and females, PYLL for males, females and in total, and infant mortality. HALE is used in some analyses, but not in the production function study that is the main focus of the work. PYLL is adjusted for certain external causes of death.¹³ Sticking with raw LE, rather than a morbidity-adjusted version seems to follow from the choices of the authors of the Report to use variables that are available for many countries over many years, even if they are conceptually inferior. The Report stresses results on LE at birth. The reported correlations, for 2003 only, among the raw LE measures are fairly high, but correlations with PYLL, adjusted mortality and infant mortality are quite a bit lower (Joumard, André, Nicq and Chatal 2008, p. 12). As we shall see, results are quite different, depending on the measures used.

A. PYLL Explained

Potential years of life lost (PYLL) is a measure of premature mortality—death that occurs before some benchmark of a potential life span. Mortality before the benchmark potential age causes lost years that, in a simple idealized world would not be lost. As the name suggests, it is measured in years of life lost. Aside from a benchmark expected life span, the concept requires a reference denominator. In principle, one could measure the potential years of life lost per person or per million people.

In this data, that benchmark of potential life is set at 70. Thus, any life beyond age 70 is ignored. PYLL can also be defined with different assumptions of potential life, such as 65 years. In this data, following traditional definitions, the PYLL is calculated per 100,000 population, for a year. The PYLL results from adding the missing years from the deaths before the assumed potential life span. Arithmetically, the measure is constructed as follows:

¹² A relationship can fail to be statistically significant by being small and tightly estimated. If so, one can reasonably rule out any important relationship. Alternatively, a relationship can fail to be statistically significant because it is very loosely estimated. In this alternative case, one cannot rule out an important relationship. The later seems to be the situation here.

¹³ Exactly how the adjustment for external causes was made was unclear, but, as discussed above, it seems that the analysts excluded deaths from road accidents, accidental falls, assaults and suicides (Joumard, André, Nicq and Chatal 2008, p. 8).

$$(1) \quad PYLL = \sum_{a=0}^{l-1} (l-a)(d_{at} / p_{at})(P_a / P_n) * 100000,$$

where:

a = age,

l = the age limit,

d_{at} = the number of deaths at age a ,

p_{at} = the number of persons aged a in country i at time t ,

P_a = the number of persons aged a in the country,

P_n = the total number of persons aged 0 to $l-1$ in the country.¹⁴

To illustrate with a simplified example, suppose that there were 2,400 people in the country, 1,000 aged 20, 800 aged 50 and 600 aged 80. Five people died during the year, one from the youngest group, two from the middle group and three from the oldest group. Thus, the PYLL equals

$$(70-20)(1/1,000)(1,000/1,800)(100,000) \\ + (70-50)(2/800)(800/1,800)(100,000) = 5,000.$$

This calculation ignores those who are over 70. As a reference, the median PYLL for the OECD countries was 3,158 (2,330 for women and 4,008 for men) in 2003 (Joumard, André, Nicq and Chatal 2008, p. 13).

As the Report notes, the PYLL has an advantage over other measures, such as LE, in that it can be adjusted by cause of death, to eliminate some of the causes of death that are due to other factors external to the health care system, such as accidents and violence. There is a natural way to do this using PYLL data, because the cause of death is recorded. One simply calculates a PYLL for deaths due to causes of death that are at least arguably sensitive to health care. One can also calculate PYLL for categories of diseases and analyze the effect of the health care system and other variables on PYLL by category, as is done if Miller and Frech (2004) for the respiratory, circulatory and cancer categories and in Or, Wang and Jamison for heart disease (2005, p. 545).

Further, the PYLL is contaminated by infant mortality, just as is LE at birth. This contamination varies by cause of death. The PYLL from cancer and heart disease are less contaminated by infant mortality than the general PYLL because infant deaths from these causes are fairly rare. PYLL by respiratory disease may be even more contaminated than LE at birth, because respiratory disease is a major problem for infants. The Report adjusts PYLL to eliminate some external causes of death. It gives examples of excluded causes of death: land transport accidents, accidental falls, suicides and assaults, but it is not clear if this list is exhaustive (Joumard, André, Nicq and Chatal 2008, pp. 19, 47).¹⁵

¹⁴ See Joumard, André, Nicq and Chatal (2008, pp. 47, 48) and Miller and Frech (2004, pp. 21-23) for more discussion on PYLL.

¹⁵ Figure 2 (Joumard, André, Nicq and Chatal 2008, p.10) shows premature mortality from “external causes” (i.e. lifestyle and social), Japan is very high, even higher than the U.S. since about 1997. This seems to be a mistake. Japan is generally low and the U.S. high in these causes of death. Robert Ohstfeldt and John Schneider (2006, p. 19) show Japan to be much below the U.S. in homicide and transport accidents, about a tenth and a half of the U.S., respectively. This data that seems incorrect for Japan calls into question the adjustments to PYLL that are meant to remove these non-health care causes and also the use of Japanese data generally. The U.S. is well above Japan in deaths due to external causes over the entire period. By 2003, the U.S. is 70 percent higher. Japan is substantially

While there is a natural and reasonable adjustment available for PYLL, it is not perfect. Those who die from an external cause differ from the general population in ways related to disease and to health care systems (e.g. risk-taking behavior, poor drug compliance). Thus, the survivors are different in a country with many deaths due to external causes. Further, accident and assault victims may die, after a lag, from some related disease. For example, an accident victim may not be able to exercise, thus have a shorter LE. Also, accident and assault victims use health care resources, especially if they do not die. This is not accounted for. Further, many deaths that are caused by external factors, but which are mediated by a disease (such as obesity and circulatory disease, or pollution and respiratory disease) cannot be eliminated by adjusting PYLL. Thus, health care and country-specific variables are still subject to confounding with omitted external factors, even when an adjusted PYLL is used.

B. Infant Mortality and External Factors

The Report (Joumard, André, Nicq and Chatal 2008, p. 8) claims that infant mortality is less influenced by non-health care causes than LE. But, there are strong reasons to believe the opposite. Infant mortality has two major problems as a health care outcome measure. First, it is affected by data definition problems and differences in common health care practice. For example, U.S. physicians (and also those in some other countries) are more likely to resuscitate very small premature babies, who later die. This U.S. practice raises measured infant mortality (and neonatal but not perinatal mortality). Pushing in the same direction, babies who die before their births are recorded are more likely to be classified as stillbirths in other countries, especially Japan and France. In the U.S., nonviable births are often recorded as live births, making the U.S. infant mortality rate misleadingly appear high. In a detailed study of medical records and birth and death certificates in Philadelphia, Gibson et. al. (2000) found that infant mortality had been overstated by 40 percent by the recording of nonviable births as live births alone. These errors are systematic, tending to make the health care system in the U.S. and other countries with similar medical and record-keeping traditions appear less efficient. The differences can be important quantitatively.¹⁶ In a comparison reported by Korbin Liu and Maryln Moon (1992, p. 109), a more inclusive measure (combining infant mortality and stillbirths) moved the U.S. up from 18th to 15th and moved Japan from first to third.^{17, 18}

There is another problem with infant mortality as a health output. Additional effective health care may improve the odds of a live birth of a baby with poor survival chances. If so, additional health care may actually make measured infant mortality worse, rather than better. This would make that the country that provided this additional health care appear to both spend more on health care and have poorer outcomes.

higher in suicide, presumably for cultural and religious reasons, but the difference in suicides is not nearly great enough to overcome the U.S. high rates in accidents and homicide (OECD 2007b).

¹⁶ For more on the measurement problems involved in infant mortality, see (Joumard, André, Nicq and Chatal 2008, pp. 47-49; Gibson et al 2000, pp. 1303 and Frech and Miller, 1999, pp. 28-29).

¹⁷ Liu and Moon do not report the total number of countries.

¹⁸ While in the rich countries, life expectancy is probably better measured than infant mortality, this relationship reverses in the poor countries. In those countries, life expectancy is generally derived from infant mortality applied to model life tables, not any actual count of age-specific mortality (Prichett and Summers 1986, pp. 858, 859).

Second and probably even more important, infant (and perinatal and neonatal) mortality are strongly and quickly influenced by other external influences, especially the mother's behavior and lifestyle, such as obesity, tobacco use, excessive alcohol use and recreational drug use (Liu and Moon 1992, p. 113; O'Neil and O'Neil 2008, pp. 8-12). Infant mortality is strongly linked to birthweight, itself largely a result of lifestyle choices (and cultural and environmental influences). The role of genetic variation across populations is controversial, but it clearly plays a role at the individual level.¹⁹ Teenage mothers are more likely to have low-birthweight babies. Mortality rates for infants born to unwed mothers were about two times as high as for infants born to married women in the U.S. (Liu and Moon, 1992, p. 112).²⁰ The mortality rates for infants born to U.S. teenage mothers is from 1.5 to 3.5 times as high as the rate for infants born to mothers aged 25-29 (Liu and Moon 1992, p. 112). The U.S. rate of births for teenage mothers is very high, 2.8 times Canada and 7 times Sweden and Japan. If the U.S. had the higher birth weights of Canada, its infant mortality would be slightly lower than Canada's, 5.4 v. 5.5 per 1,000 (O'Neil and O'Neil, 2008, p. 10).²¹ Further, apart from worsening the infant mortality statistics, the low birthweights of the U.S. lead directly to higher health care utilization and total spending because health care for low birthweight babies is costly.

Since infant mortality is an important component of LE at birth, these problems imply that LE at birth is inferior to LE at later ages for analyzing the productivity of health care. As Martin Neil Baily and Alan Garber put it:

Life expectancy (at birth) is heavily influenced by neonatal mortality, which is higher in the United States than in the other two countries (the United Kingdom and Germany). Although impaired access to health services and a lack of productivity could contribute to less favorable birth outcomes in the United States, neonatal mortality is heavily influenced by social and economic factors, along with individual health behaviors that are not strongly related to health care delivery. Overall life expectancy at birth, then, may be an unsuitable measure of health outcomes for the purpose of measuring productivity of health services (Baily and Garber 1997, pp. 188-189).

C. Adjusting the Measure for Non-Health-Care Causes

The Report argues that deaths unrelated to the health care system, such as transport accidents confound mortality estimates (Joumard, André, Nicq and Chatal 2008, p. 4). The argument should be extended to include more factors, such as

¹⁹ For a study of the importance of genetic factors at the individual level, see David Stevenson and John Carey (2004). For an argument that genetic factors are probably not important in comparing infant mortality and low birthweight across black and white American mothers, see Richard David and James Collins (2007). David and Collins are skeptical that the "social, economic political and historical effects of racial discrimination" can be adequately controlled for with observable measures (p. 1192). The same issue arises in comparisons across countries.

²⁰ This relationship has likely weakened since the 1980s. Further, it is probably far weaker in Europe, where unmarried fathers more often live with their children.

²¹ For a more detailed analysis using slightly older data that performs similar calculations in comparisons to many other countries see Liu and Moon (1992).

violence and cultural and lifestyle variables. As noted above, there is a simple and natural, if imperfect, way to adjust partially only for PYLL. For LE, there are two possible ways to adjust for non-health-care causes of death. Both require statistical modeling and, therefore, judgment calls and controversy. The two ways to adjust for non-health-care causes of death are discussed below in Section III. 3. First, partial but simpler approaches will be discussed.

1. Life Expectancy at Later Ages

One can somewhat reduce the problem of confounding variables by focusing on LE at later ages. This helps because infant mortality is highly contaminated with external factors like lifestyle and with measurement problems that make the U.S. health care system appear less efficient. LE at later ages, such as 40, 60 or 65, of course, eliminates the people who died in the first year of life. Further, many of the lifestyle choices that lead to bad outcomes are more heavily concentrated among younger consumers and affect LE more at younger ages. For example, in 2005 U.S. data, transport deaths peak at 29.10 per 100,000 population per year for ages 20-24 and never reach that level again at any age. Similarly, the all-injury death rate has an early peak at 73.75 per 100,000 per year at ages 20-24. After that, the all-injury death rate does not catch up to that level until ages 75-79 (CDC 2008).

2. Birthweight-Specific Infant Mortality

As discussed above, birthweight is highly sensitive to lifestyle choice (and also to social and environmental issues) and strongly affects infant mortality. Thus, holding birthweight constant would eliminate some of the confounding effects of lifestyle and other influences.²² The result of doing so is dramatic. For example, comparing the U.S. to Canada, O'Neil and O'Neil (2007, pp. 21, 22) show that the total infant mortality is lower in Canada (5.50 v. 6.85 per 1,000), while the birthweight-specific mortality rates are lower in the U.S. for smaller babies and only slightly higher for the larger babies. This is why, as is discussed above, virtually all of the difference between the two countries can be explained by differences in birthweight. Looking across many more countries, Liu and Moon (1992, p. 115) show that most of the difference between the U.S. and these other countries is explained by the difference in the distributions of birthweights.

In the context of this study, a simple, but imperfect, adjustment would be to replace the overall infant mortality rate with a birthweight-specific rate. The results would probably differ slightly depending on the exact choice. Alternatively, one could form an index by picking some distribution of weights to multiply by the birthweight-specific infant mortality rates. Here again, the results may differ slightly depending on which weights were used.²³ Natural possibilities would include the OECD average weights, OECD less the U.S. average weights and U.S. weights.

²² On the other hand, medical care has some effect on birthweight, so holding birthweight constant may overcontrol to some limited extent.

²³ This is an example of the index number problem. An index is a type of weighted average. The index number problem arises when, as is often the case, there are several candidates for the weights. The consumer price index is a good example. The weights are the expenditure proportions as of some date. The results for the measured change in the CPI depend on the choice of date as a basis for the expenditure weights.

3. Life Expectancy and Non-Health-Care Causes of Death

As mentioned above, there are two possible ways to adjust LE for non-health-care causes of death.

a. Adjusting the Life Expectancy Variable

The LE variable can be adjusted for some external causes by a statistical process to produce a standardized LE. A standardized LE is the LE after it has been purged of the influences of the external causes. One may also think of it as the LE one would expect if the country had the average level of deaths by external causes, rather than its own actual level. For the U.S., it would be the expected LE in the U.S. if it had the average level of deaths by external causes, rather than the higher level it had in fact. A somewhat generalized version of this approach is taken by Ohstfeldt and Schnider (2006, pp. 5-33). It is generalized in that it also standardizes for GDP *per capita* and for year, not merely external, non-health-care causes of death.

Ohstfeldt and Schnider estimate a model explaining country-level LE at birth by year dummy variables, GDP *per capita*, injuries in transport and fall, homicides and suicides.²⁴ The dataset is a panel of OECD countries from 1980 to 1999. This model explains an impressive amount of the cross-country variation in LE, more than 79 percent.²⁵ The equation resulting from that estimation can be used to create a standardized LE. The residual (difference between a country's actual LE and the LE predicted by the model) is a measure of over or underperformance of that country. This residual (over or underperformance) can be added to the expected LE for the average OECD country. The result is the expected LE if that country had the mean level of the external causes (and *per capita* income). This expected LE is the predicted value from the equation, evaluated by setting all the independent variables to their means. The result is the standardized LE.²⁶ It has been purged of the effects of these non-health care causes of death.

To parallel the OECD study, the next step would be to estimate a production function with this adjusted LE.²⁷ Ostfeld and Schneider instead do something less formal. They compare the average LEs over this time period for the OECD countries, raw versus standardized. The differences are dramatic. In the raw LE measure, the U.S. LE at birth averaged 75.3 years, which is less than 76.6 years for France, 78.7 years for Japan and 77.7 years for Sweden.²⁸ In the adjusted LE measure, the U.S. LE at birth averaged 76.9 years, which is more than 76.0 years for France and Japan and 76.1 years for Sweden (Ostfeld and Schneider 2006, pp. 21, 22). In fact, the U.S. does the best of all the OECD countries on this measure. Note that this analysis controls for deaths by injury, but it does not control for many lifestyle, cultural and environmental variables. The difference in the LEs generally, and especially the large effect on the U.S., show that these non-health-care factors are heavily confounded in

²⁴ The natural logs are taken of the continuous variables.

²⁵ The adjusted R^2 is 0.79, implying that the ordinary R^2 is higher. The percent of the variation explained is equal to the ordinary R^2 .

²⁶ There are many mathematically equivalent ways to get to this estimate of the standardized LE. Ostfeld and Schneider (2006) are not explicit about exactly how they did it.

²⁷ To be exactly parallel to the OECD study the adjustment would have to be slightly simpler; excluding year dummies and GDP.

²⁸ In the adjusted LE measure, the range of variation is compressed.

the raw LE measures and that the confounding makes the U.S. health care outcomes appear worse than they are.

There are other specifications that might give somewhat different results. There is no natural or dominant approach. Also, just as in the PYLL adjustments, the people who die of injuries are not a random selection, so that the survivors might be healthier. That would make countries with high injury rates, like the U.S., appear more efficient than they are. Injuries lead to more use of health care, which has the opposite bias: causing the U.S. to appear less efficient than it is. A recent study by Charles Roehrig, George Miller and Craig Lake shows that expenditures due to injuries and poisoning account for 7 percent of U.S. health care spending and that trauma is the third most costly condition, ranking ahead of cancer (2008, pp. 21-22). Further, if one wants to go beyond showing the importance of the confounding problems and get the best quantitative estimates of the effect of health care and of country-specific health care efficiency, there is a better method.

b. Directly Controlling for Non-Health Care Causes of Death

The alternative method to adjust for at least some non-health-care causes of death is to add the relevant variables to the production function itself. Applying this method to the injury causes discussed above, one approach would be to add the relevant injury death rates as new independent variables to the final health production function equation. If getting the most accurate final estimates is the goal, this is a more efficient and more direct method. This is an example of the general strategy of augmenting the estimating equation with important variables that are now omitted.

IV. Specification of the Panel Data Regressions

A. The General Model

In the panel data approach, a single-equation production function is estimated in the Report (Joumard, André, Nicq and Chatal 2008, pp. 19-22). The equation is specified as

$$Y_{it} = \alpha_i + \beta HCR_{it} + \gamma \cdot SMOK_{it} + \phi \cdot DRINK_{it} + \theta \cdot DIET_{it} + \delta * AIRPOL_{it} + \sigma \cdot EDU_{it} + \lambda \cdot GDP_{it} + \varepsilon_{it}$$

where:

Output

Y = health status, variously measured.

Inputs

HCR = health care resources *per capita*, measured two ways.

SMOK = tobacco consumption in grams *per capita*.

DRINK = alcohol consumption in litres *per capita*.

DIET = consumption of fruit and vegetables *per capita* in kgs.

AIRPOL = emissions of nitrogen oxide (NOx) *per capita* in kgs.

EDU	=	share of the population (aged 25 to 64) with at least upper secondary education.
GDP	=	GDP <i>per capita</i> ,
ε_{it}	=	the error term, accounting for all omitted factors and randomness.

The equation also includes a set of dummy, zero-one, variables for country fixed effects. This allows the constant term in the equation to be different for each country. The model is not a full fixed effects model, because there are no dummy variables for the years. All continuous variables are in natural logs. The subscript it , refers to country i and year t , e.g. Y_{it} is health status in country i and year t . Health status, Y , is measured by LE at birth, for males, females and in total, LE at 65, for males and females, potential years of life lost (PYLL) (adjusted for external causes as discussed above), for males, females and in total, and infant mortality. Health resources are measured alternatively by total spending or by a weighted count of certain health practitioners (Joumard, André, Nicq and Chatal 2008, pp. 19, 20). This is a reasonable general approach, but not the only possible one. Or, Wang and Jamison (2005) use a more flexible alternative that allows the (slope) coefficient on health care to vary across countries, as well as the constant. Another possibility is the stochastic production function method that apportions the error into apparent inefficiency versus other country-specific effects (Greene 2004).

Logging all the variables imposes a particular functional form on the data. This functional form is called a log-log, double-log or constant elasticity form. The log-log form incorporates and imposes diminishing returns to the inputs.²⁹ There are two margins for increases in health care resources. First, at the extensive margin, more health care inputs increase the number of consumers treated. Second, at the intensive margin, more health care inputs increase the intensity of treatment for the same number of consumers.

One would expect diminishing returns in health production on both margins because of a tendency to allocate health care where it has the largest effect. On the extensive margin, one would expect the consumers with the most ability to benefit from the care to be the first ones to get care. Subsequent consumers to get care would be less likely to benefit. This type of rationing across consumers is called triage. On the intensive margin, the first type of care would be the most productive. A similar argument could be made for the other inputs. Health production data typically support diminishing returns (Baily and Garber 1997, pp. 147-148; Frech and Miller 1999, pp. 80-81; Fuchs 2004, p. VAR-105; Garber and Skinner 2008).

The existence of diminishing returns implies that countries with heterogeneous populations (i.e. different consumers choosing different levels of health care) will appear, falsely, to be less efficient. In this context, Alan Garber and Jonathan Skinner (2008, pp. 31-35) point out that the U.S. is likely to be especially heterogeneous for two reasons. First, the health care insurance system is more varied. Second, regional variation in health care utilization (most of which cannot be explained by variation in health insurance) is more pronounced in the U.S. than in other rich countries.

²⁹ The log-log functional form exhibits diminishing returns if the estimated coefficients are less than 1.0 in absolute value. That is clearly the range of possible values here. The largest estimate of the effect of health care on any health measure, for infant mortality, is -0.572. The largest estimate for any form of life expectancy is 0.061 (Joumard, André, Nicq and Chatal 2008, p. 32).

Using this log-log functional form, the estimated coefficients are elasticities, giving the percentage impact of a 1.0 percent increase in the variable. Thus, an estimate of 0.04 for β would imply that a doubling of health care resources would increase health status by 4 percent. This would be a large effect.

Time, Trends and Simultaneous Equations Bias

The OECD Report's model uses partial fixed effects, with dummy variables entered only for countries only, not for years. Thus, all the effects are estimated by changes over time in the independent variables. Time-invariant cross-sectional variation is absorbed by the country dummy variables. Time is not picked up by a year fixed effect, nor by a time trend. Since all the economic effects come from changes over time, this specification causes the estimated effects to be confounded with the passage of time. In the health care sector, the main concern is the rapid pace of technological change. Health care has apparently become much more productive over time.³⁰ There are two possible ways of dealing with the problem. First, one could make the analysis a full fixed effects model, by adding a dummy variable for each year. Those year dummy variables would account for general shocks that affect all OECD countries, such as technological progress. That solution uses up a lot of degrees of freedom, hence statistical power, because it requires the estimation of about 20 more coefficients. A partial solution that would be less costly in degrees of freedom would be to introduce a linear, or perhaps quadratic, time trend.³¹

Also a time issue, the explanatory variables are contemporaneous with the health outcomes; there are no lags. This is a problem because it leads to measurement error and also the possibility of simultaneous equations bias. In terms of measuring the inputs into health production, using lags makes economic sense because it takes years for the effects of some variables, especially lifestyle ones, to take full effect. Not using lags will bias down the effects of observed and included lifestyle variables. Because the incorrect lag implies that the variable is not fully controlled for, it will introduce measurement error into the variable. This biases upwards the apparent inefficiency of the U.S. system, because the U.S. lifestyles are relatively unhealthy. Most of the prior literature uses lags. For example, in cross-sectional analysis, Comanor, Frech and Miller (2006, pp. 13, 14), Miller and Frech (2004, p. 63) and Zweifel and Ferrari (1992) use lags of about six to 10 years. In a panel of OECD data that is similar to what is used in the Report, Peter Zweifel, Lukas Steinmann and Patrick Eugster (2005, p. 136), test lags of differing lengths and report that a lag of 10 years seems to be the best. The only lag to be tested experimentally in the Report is on GDP (Joumard, André, Nicq and Chatal 2008, p. 21). The conceptual argument for lagging GDP is probably weaker than for many other variables.

The simultaneous equations problem arises because of possible reverse causation. A country may use many health care resources because its population is in

³⁰ For a discussion of the trend in health care productivity, see (Cutler and McClellan 2001; Murphy and Topel 2003; Cutler 2004).

³¹ Another model, called the random effects model, is also commonly used for panel data. It is an adjustment for heteroskedasticity only—allowing the error term to differ by country and by year. Random effects models assume that there is no correlation between the country-specific effects and the explanatory variables (i.e. that the fixed effects, if any, are uncorrelated with the independent variables). Here, that seems clearly to be incorrect. Random effects coefficients are as vulnerable to omitted variable bias as ordinary least squares estimated coefficients are. It is possible to use random effects and fixed effects in the same model, but that is rarely done.

poor health. That is, health outcomes may influence health care resources used, the reverse of what the Report's authors are trying to estimate. This effect would bias the apparent productivity of health care downward. While this is a new area of research, there is some evidence for this reverse causation in OECD data (Zweifel and Ferrari 1992, Zweifel, Steinmann and Eguster 2005). The use of lags would reduce concern about this issue. It is less likely that health outcomes in 2000 could have influenced health care spending ten years earlier in 1990 than that health outcomes could have influenced health care spending in the same year.

The model is estimated by a Generalized Least Squares method that corrects for heteroskedasticity (expected errors differing across observations) and serial correlation (errors being correlated over time). The correction for serial correlation is flexible: allowing the serial correlation to differ among countries (Joumard, André, Nicq and Chatal 2008, pp. 20, 21). This correction for serial correlation may avoid the problem of spurious correlation that can overstate the relationship between variables that move together over time.³²

B. Included Input Variables

The output, health status variables have been analyzed above. The categories of input variables included are based on generally sound concepts, but the variables actually used in the econometric analysis are often questionable. At best, the input variables are limited and incomplete.

1. Health Care Resources

a. Total Spending

There are two very different measures of health care resources used in the Report. The total spending variable is aggregated over the entire health care system. This could create problems if, as previous research suggests, the productivity differs for different types of care (e.g. spending on pharmaceuticals versus other spending, public versus private spending). The coefficient on the aggregate version captures a type of weighted average effect.³³

Perhaps most important, the health spending variable is converted to a common currency by the general purchasing power parity (PPP) exchange rate (Joumard, André, Nicq and Chatal 2008, p. 31) exchange rate, not the health care PPP exchange rate.³⁴ Recall that health spending is meant to represent real resources devoted to health care. Using an incorrect exchange rate introduces measurement error into health spending. The error is systematic because of health care prices are known to be high in the U.S.

As a practical matter, there are three possible exchange rates that might be used: market exchange rates, PPP exchange rates for the general economy and PPP

³² Spurious correlation is caused by what is called the unit root problem. Asymptotically (as the sample size grows large), the serial correlation correction avoids the problem (Hamilton 1994, pp. 557-562). The unit root problem is likely to be present in health care time series data. See Miller and Frech (2004, pp. 12-14).

³³ Specifically, the coefficient measures a weighted average where the weights are the sample variances of the independent variables.

³⁴ The data are originally collected from each country in that country's currency. Thus, some exchange rate is necessary to convert the data into a common currency.

exchange rates specifically calculated for health care. Market exchange rates are obviously flawed for our purpose, as they are strongly affected by financial flows and inflationary expectations. These rates can be volatile and clearly fail to represent real resource use. For example, the exchange rate for changing U.S. dollars into euros (dollars/euro) was 0.95 on Jan. 1, 2001. Seven years later, it was 1.47 on Jan. 1, 2008 (Board of Governors of the Federal Reserve System 2008). That is a 55 percent increase. So, if domestic health spending in euro-block countries remained static in terms of domestic currency, it would appear to have risen by 55 percent in U.S. dollars, based on the change in this exchange rate. The point is put well by Ian Castles and David Henderson,

An exchange-rate-based conversion of the money GDP of two countries in a particular year takes no account of price differences between them. *It therefore does not yield a measure of comparative output.* Only by eliminating price effects, and thus valuing each country's GDP at a common set of prices, is it possible to derive a valid measure of differences in real GDPs (emphasis in original) (Castles and Henderson, 2005, p. 9).

The goal here is a valid measure of comparative use of health services resources.

PPP exchange rates are based on the ability to purchase goods with one unit of the base currency (here the U.S. dollar). The idea is that it might take €0.85 in France purchase the same bundle of goods as \$1.00 in the U.S. Therefore, multiplying the French spending in euros by 1.18 (1/0.85) converts it into U.S. units so that the measure corresponds to real resources. One can define a PPP exchange rate for the overall economy, usually called the Gross Domestic Product (GDP) PPP exchange rate, or one can define a PPP exchange rate for a sector, such as health care or pharmaceuticals. Using the overall Gross Domestic Product PPP exchange rate, as the OECD Report and some other literature does, would be correct only if the GDP and health care PPP exchange rates were proportional. That is, it would be correct to use the GDP PPP exchange rate only if the relative price of health care to other goods was constant across countries. Perhaps for some sectors that are comprised of internationally traded and standardized goods, one could expect this constant relative price to be approximately correct.

But, the relative price of health care services varies a great deal across countries, so the health PPP exchange rate varies greatly from the GDP PPP exchange rate. Table 1 and Figure 1 show the values for 1990 for the health PPP exchange rate, the pharmaceutical PPP exchange rate, the GDP PPP exchange rate and some ratios of the health and pharmaceutical PPP exchange rates to the GDP PPP exchange rates. The PPP exchange rates here are the number of units of other currency necessary to purchase a U.S. dollar. So, a GDP PPP exchange rate of 1,421 for Italy means that it takes 1,421 lire to purchase \$1.0. The ratios of these rates indicate the extent of understatement of real resources used in the other countries that results from using the GDP PPP exchange rate. The mean ratio for health spending is 0.67, while the mean for pharmaceutical spending is 0.70. Thus, health resources consumed in these OECD countries, measured by the GDP PPP exchange rate is about 30 percent lower than the health resources measured by the health PPP exchange rate. Alternatively put, to obtain estimates of real resources used in another country in U.S. dollar terms, one would have to multiply the health care spending as measured by the GDP PPP

exchange rate by the inverse of the ratio shown in the table. That inverse is the ratio of the GDP PPP exchange rate to the health care PPP exchange rate.³⁵ This has large effects on the apparent health resources used in the production of health.

As one can see in Table 2 and Figure 2, the use of the health PPP exchange rate increases the estimates of real spending on health care in the other OECD countries substantially. They go from an unweighted average of 50 percent to 78 percent of U.S. expenditures. The difference is 56 percent (28 percentage points). Interestingly, when the health PPP exchange rate is used, the U.S. is no longer the highest country. France and Norway exceed the U.S. in real health care consumption.

This overstatement of U.S. health care resources consumed relative to those of other countries occurs because health care prices are much higher in the U.S. The numbers in the Table 2, using the GDP PPP exchange rate (and the shorter bars in Figure 2) are the ones one typically sees in comparisons of health spending across countries (e.g. Reinhardt, Hussey and Anderson, (2003, p. 170; Huber and Orosz 2003, p. 11). Further, even within Europe and within particular narrowly defined episodes of care, using GDP PPP exchange rates, health care PPP exchange rates and carefully-constructed, episode-specific PPP exchange rates leads to substantially different estimates of real expenditures (Schreyogg, Tiemann, Stargard and Busse 2008, p. S100).

³⁵ The ratios can be confusing. To fix ideas, consider the following example. Suppose that health care spending in the U.K. for some year was £2000 and the GDP PPP exchange rate is 1.5 \$/£. U.K. health care spending in \$ is then,

$$£1500 * 1.5 \$/£ = \$2250.$$

Now, suppose that, because health care prices are lower in the U.K., the health care PPP exchange rate is 2.0\$/£. The real U.K. health care spending, reflecting resources use, is

$$£1500 * 2.0 \$/£ = \$3000.$$

Starting with the \$2250 from the GDP PPP calculation, one can also arrive at the correction amount for real spending by multiplying by the ratio of the health PPP exchange rate to the GDP exchange rate

$$\$2250 * (2.0 \$/£) / (1.5 \$/£) = \$3000.$$

This is the procedure used in the text and in Figure 2.

Table 1: Comparing PPP Exchange Rates, 1990

Country	GDP PPP	Health Care PPP	Ratio: Health Care to GDP PPP	Drug PPP	Ratio: Drug to GDP PPP
Australia	1.39	1.02	0.733813	0.83	0.597122
Austria	14	8.59	0.613571	11.29	0.806429
Belgium	39.5	21.15	0.535443	25.08	0.634937
Canada	1.3	0.93	0.715385	1.15	0.884615
Denmark	9.39	6.94	0.739084	7.94	0.84558
Finland	6.38	4.5	0.705329	4.07	0.637931
France	6.61	3.62	0.547655	3.02	0.456884
Ireland	0.69	0.48	0.695652	0.58	0.84058
Italy	1421	876.8	0.61703	768	0.540464
Netherland	2.17	1.36	0.626728	2.12	0.976959
New Zealand	1.61	1.04	0.645963	1.16	0.720497
Norway	9.73	6.09	0.625899	5.63	0.578623
Portugal	104	66.4	0.638462	64.3	0.618269
Spain	110	65.65	0.596818	55.3	0.502727
Sweden	9	6.06	0.673333	4.95	0.55
Switzerland	2.2	1.69	0.768182	1.68	0.763636
United Kingdom	0.6	0.34	0.566667	0.43	0.716667
United States	1	1	1	1	1

Source: OECD (2000), as summarized in Comanor, Frech and Miller 2006, p. 8

Figure 1
Ratio of Health PPP to GDP PPP, 1990

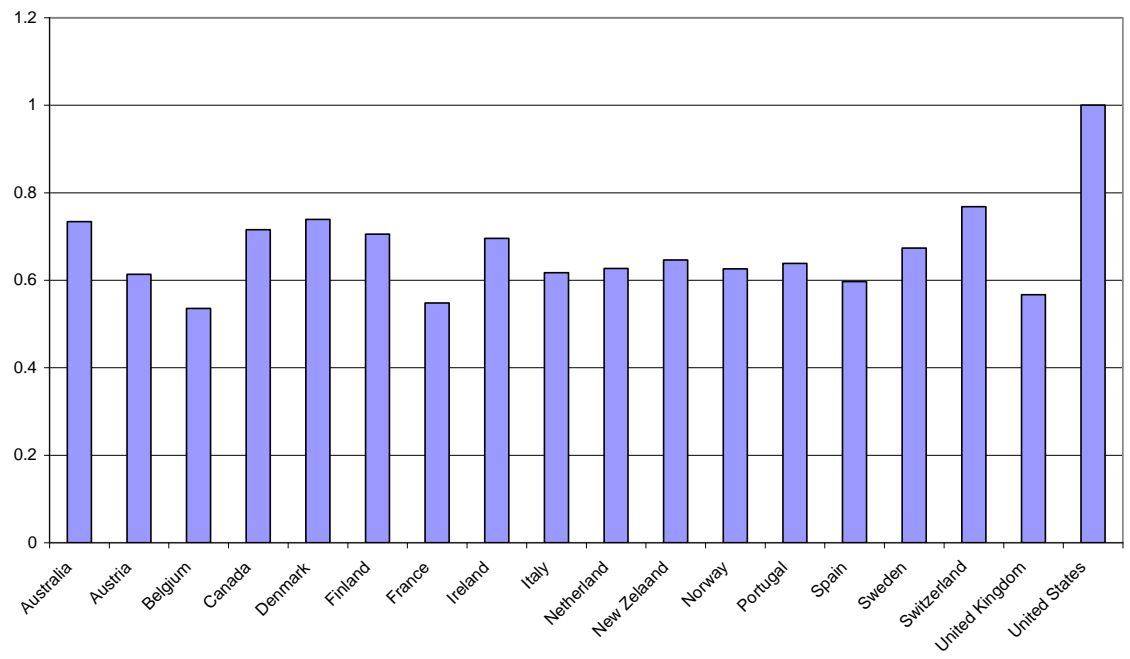
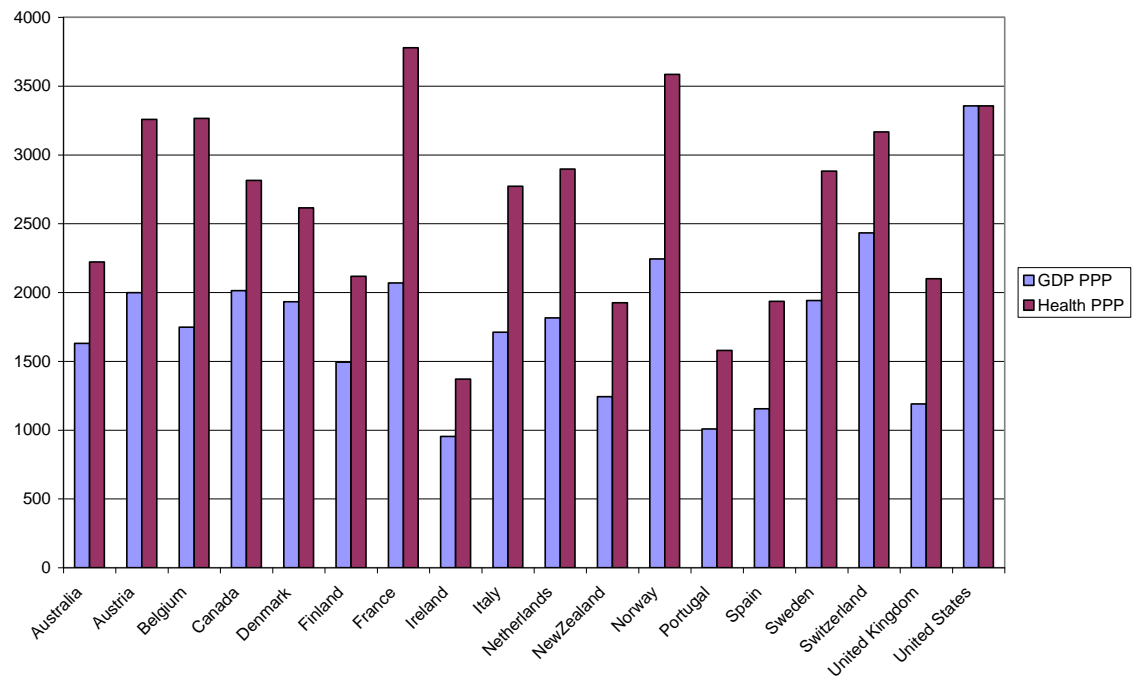


Table 2: Real Health Expenditures at Health versus GDP PPP Exchange Rates, 1990
U.S. dollars

Country	GDP PPP Exchange Rate	Health PPP Exchange Rate
Australia	1630.259	2221.627
Austria	1998.736	3257.544
Belgium	1748.302	3265.15
Canada	2013.446	2814.494
Denmark	1932.548	2614.788
Finland	1493.159	2116.968
France	2069.349	3778.562
Ireland	953.1897	1370.21
Italy	1710.289	2771.807
Netherlands	1815.531	2896.84
New Zealand	1243.447	1924.952
Norway	2243.863	3585.022
Portugal	1007.868	1578.589
Spain	1155.262	1935.702
Sweden	1940.782	2882.35
Switzerland	2432.971	3167.182
United Kingdom	1189.661	2099.403
United States	3356	3356

Sources: OECD (2000), as summarized in Comanor, Frech and Miller 2006, p. 8; OECD data set personal communication, 2008

Figure 2, Real Health Expenditures at Health versus GDP PPP, 1990 U.S. dollars



The fact that health care prices are higher in the U.S. than other countries has been noted by several analysts (Pauly 1993; Anderson, Reinhardt, Hussey and Petrosyan 2003; Frech and Miller 1999, pp. 22-28; Comanor, Frech and Miller 2006, p. 8; Bailey and Garber 1997, p. 188; Garber and Skinner 2008, pp. 43-44). The authors of the Report note that physician incomes are higher in the U.S. (Joumard, André, Nicq and Chatal 2008, p. 31).³⁶ The incomes of other health care workers are also higher in the U.S. For example, a recent study of nurse migration compared nursing wages in the five countries, including the U.S., that have experienced immigration of nurses. Nurses in the U.S. earned the most. The difference ranged from a minimum of 8 percent more than in Australia to a maximum of 43 percent more than in France (Vujicic, Zurn, Diallo, Adams, Dal Poz 2004, p. 10 (of web version)). Also, as is discussed below, pharmaceutical prices are higher in the U.S.

For some purposes, the comparisons using GDP PPPs are appropriate. They measure the financial flows. However, for a health care production function, expenditures need to be put into real terms with the correct prices in order to measure real resources used in health care. Further, most informal, descriptive analyses take place implicitly in a household production function context.

An observer may also be interested in reducing the pay of health care workers, especially physicians, and reducing the prices and profits of the pharmaceutical firms. Indeed, one sometimes sees analyses or arguments focused on these issues.³⁷ But, it is important to keep the issues of health care system productivity separate from the issue of factor prices.

In a study of the OECD's PPP exchange rate program, Ian Castles showed the large difference in health care resource use one gets for Japan versus the U.S., depending on whether one uses the health care PPP exchange rate or the overall GDP PPP exchange rate. Using the GDP PPP exchange rate, 1993 spending in the U.S. is 224.5 percent of spending in Japan. Using the health care PPP exchange rate, U.S. real resource use is only 86.9 percent of Japan's (Castles 1997, pp. 31, 32). Taken at face value, this huge difference occurs because the relative price of health care is much lower in Japan (61 percent lower). Castles does not find it plausible that the price difference is really that large, so he takes this finding as an indication that the health care PPP exchange rate is not very reliable.³⁸ Japan's apparent price for health care is lower than any other of the OECD countries analyzed above.³⁹ The belief that the health PPP exchange rates are not reliable is and was the view of the OECD Statistics Directorate and the authors of the Report (Joumard, André, Nicq and Chatal 2008, p. 52). There are many other ways to demonstrate that the prices of health care differ greatly across countries and that the price of U.S. health care is high, that do not rely on the health care PPP exchange rate, as we shall see below.

³⁶ For more on international differences in physician pay, see Rie Fujisawa and Geatan Lafortune (2008).

³⁷ These policies can reduce spending, especially in the short run. But, a policy to suppress wages or prices by monopsony (buyer market power) creates distortions and causes inefficiencies. See Pauly (1988, 2008, pp. 25-28), Jill Herndon (2002) and Mark Ramseyer (2009). Suppressing pharmaceutical prices reduces incentives for research and development. See Patricia Danzon (1997b), John Calfee (2000) and F.M. Scherer (1993).

³⁸ In contrast to Castles's skepticism about the magnitude of the Japan-U.S. price differences, Ramseyer reports a study that shows even more difference than is implied by Castles's calculations (75 percent below U.S. prices) (Ramseyer 2009, p.3).

³⁹ Another possibility is that Japanese health data are unreliable. Indeed, concern with that has led to the exclusion of Japanese data from earlier studies, e.g. Comanor, Frech and Miller (2006)..

Price Controls and Systematic Measurement

Errors

Other problems with using spending to measure health care inputs arise because many health care systems use price controls to some extent. This leads to two types of measurement problems that systematically understate the economic cost of the health care system. The first is the hidden cost of nonprice rationing, while the second is the hidden cost of informal, black market co-payments.

A fundamental rationing problem arises that when price controls are set below market-clearing prices, as they generally are. In this situation, price controls cause excess demand. That is, more consumers seek more care than the providers are willing to supply at the controlled price. Potential buyers must be rationed out of the market in some way: discouraged or prevented from consuming care. Formal health care rationing is historically rare, so that most actual rationing is informal and typically well hidden.^{40,41}

Often, the rationing is by waiting time and queuing, an especially wasteful practice, but one that is hard to observe and measure.⁴² In spite of the difficulties of observation and measurement, some indicia of the problem are available. For example, long waiting times and short physician visits, are common in tightly price-controlled Japan and Korea. In those countries, patients are often asked to return for multiple visits and appointments are generally unavailable—requiring lengthy waits in physician's offices. For example, in Japan 25 percent of patients wait over two hours to see a physician in the public (generally higher quality) hospital. Waits are only slightly shorter in the private sector. In Japan and Korea, doses of medicine are often split in two in order to generate two dispensing fees and two short office visits (Ramsayer 2008, pp. 5, 6).

Studies in Quebec when Universal Insurance was instituted showed a major shift away from home visits. Like rationing by waiting, dropping home visits raises costs for consumers, but this higher cost is not captured in any accounting systems. Also, as one would expect from the theory of nonprice competition, quality, at least some dimensions, appears to be lower under price controls (Frech 2000, pp. 350-352; Ramsayer 2008). Consumers bear costs of nonprice rationing, partly in the form of waiting time, in price-controlled systems. These costs are subtle and not picked up in health spending accounts, leading to systematically understating the resource costs of the health care systems that rely heavily on nonprice rationing.⁴³ Canada is a prime example. See Danzon (1992, 1993) for an analysis of the hidden economic cost imposed by Canadian price controls. As a result, the health care resources use of non-

⁴⁰ Formal health care rationing is becoming more common and more important, particularly in the approval system for coverage of new pharmaceuticals in governmental health insurance systems. See Dranove (2003).

⁴¹ The waste of informal nonprice rationing is not so hidden in some other markets. For example, U.S. price controls on gasoline in 1973 and 1979 caused long lines at gas stations that made the waiting costs clear. See H.E. Frech III and William Lee (1987) and Robert Deacon and Jon Sonstelie (1985).

⁴² See H.E. Frech III and Paul Ginsburg (1978, esp. pp. 36-41), Frech (2000) and Danzon (1992, 1993). Formal rationing and informal rationing among competing managed care plans both have the potential to reduce inefficiency, relative to informal rationing in centralized health care systems. See Dranove (2003) and Frech (2000, pp. 353-354.).

⁴³ For an interesting study of the U.S. Medicaid system, a state-run system for the poor, see Thoman Koch (2008). Much like the price-controlled national health insurance system, Medicaid provides generous coverage, but low prices to providers. Koch shows that the result is poor access to physician services.

U.S. OECD countries will be systematically underestimated, making the U.S. appear to be less efficient.

Another effect of price controls and nonprice rationing that has gotten far less attention is the system of corruption and informal co-payment that is important in some places. Called “co-payment by envelope” in Japan, or “red envelope” in Taiwan, or “fakellaki” in Greece, the practice is apparently common. These payments are not normally captured by health care accounting systems (Ikegami 1991; p. 104; Frech 2000, pp. 351-352; Chiu, Yu-Chan, Smith, Morlock and Wissow 2007; Liaropoulos, Siskou, Kaitelidou, Theodorou and Katostaras 2008, Ramsayer 2009, p. 5). A recent study of the Greek system found that 36 percent of those treated in public hospitals reported making informal payments. Further, it reported that the Greek National Statistical Service recently recalculated the 2004 health care spending, including a part of the “black economy” payments, raising the estimated share of private spending from 46.1 to 55.4 percent (Liaropoulos, Siskou, Kaitelidou, Theodorou and Katostaras 2008, pp 72-74). Assuming that public health spending was not revised downward, this implies that the recalculation raised the estimated total spending by about 21 percent or more.⁴⁴ Informal bribes for health care are generally criticized on various policy grounds, such as efficiency, transparency and income distribution.⁴⁵ But, the issue of concern here is independent of those issues. Relatively large informal payments in some countries are normally missed by accounting systems, providing another source of downward bias in the apparent health care resources used in those countries. Since price controls are relatively less common and less stringent in the U.S., this biases results towards an appearance of U.S. inefficiency.

b. Physical Measures of Health Care Resources

Another way to measure health care resources is to use an aggregated measure of physical inputs. The Report creates an index of weighted health workers per 1,000 population, based on weighting a nurse as one half of a physician (Joumard, André, Nicq and Chatal 2008, pp. 29, 30) as an alternative to the health spending measure discussed above.⁴⁶ The Report states that the weighting is *ad hoc*, but a weighting of this sort can be based on objective market data, as was done by Mark Pauly (1993). Pauly includes a much broader array of workers (including many unskilled and semiskilled workers) and uses relative wages in the U.S. to form the weights. Thus, it is conceptually superior to the more limited measures. Further, the difference is quantitatively important. Physicians and nurses in total make up only 18.6 percent of the U.S. health care workforce, 3.4 percent for physicians and 15.2 percent for nurses (Bureau of Labor Statistics 2008, p. 6). Pauly’s analysis is for 1988 data. The weight for physicians is 4.83 times the weight for other workers (p. 156). Written the same

⁴⁴ Setting the initial total Greek spending to 1.0 implies that initial private spending is 0.461 and public spending is 0.539. After revisions, if public spending is unchanged, the new ratio of private to total spending can be written in terms of revised private spending

$$\text{Private Spending} / (\text{Private Spending} + 0.539) = 0.554,$$

Implying that revised private spending = 0.668 and revised total spending = 0.668 + 0.539 = 1.207, an increase of 20.7 percent.

⁴⁵ For example, see Camiola Ionescu (2005).

⁴⁶ This measure ignores technicians and other health workers and nonlabor inputs, such as pharmaceuticals, devices, equipment and buildings. Pharmaceuticals are discussed below.

way, the OECD weight for physicians is 2.0 times the weight for nurses.⁴⁷ The OECD numbers are expressed as the number of health workers per 1,000 population. This is conceptually not quite as good as using a proportion of the work force. Another approach is to concentrate on physicians.⁴⁸

One could also simply use the number of physicians per 1,000 population. This is one of measures of health resources (as opposed to spending) used by Anderson, Reinhardt, Hussey and Petrosyan (2003, p. 95), and Anderson, Bianca, Frogner and Reinhardt (2007, p. 1485) in their descriptive analysis. It is also the only approach of Zeynep Or, Jia Wang and Dean Jamison in their health production study (2005). See Table 3 for these measures.

Analyzing all these measures of physical resource use in health care, there are several things to note. First, they paint a very different picture of the real resources used by the U.S. health care system from that one gets with the commonly-used expenditures at the GDP PPP exchange rate. Much like using the health PPP exchange rate, but even more striking, it is clear that the U.S. health care system is not a particularly high user of health care resources. For example, using Pauly's physical input measure, the most comprehensive, the U.S. resource use is 6th of 12, slightly below the mean. Using the more narrow measure based on weighted physicians and nurses only, the U.S. is 4th out of 14. Looking only at physicians, the U.S. is only 9th of 18, and again, slightly below the mean. Clearly, the U.S. uses relatively more nurses and less other types of nonphysician workers than the other OECD countries, so the Report's measure overstates U.S. resource use. Most importantly, the U.S. is not a high user of labor resources in its health care system. Using the GDP PPP exchange rates to calculate real resource use is highly misleading. Using that data in a health production model creates a large bias towards inaccurately portraying the U.S. system as inefficient in producing health with health care resources.

Another point to note is that these physical measures differ quite a lot, even though they are based on health care personnel. The creation of broader indexes with actual weights, in the spirit of Pauly's work, would be welcome.

⁴⁷ For an alternative, Pauly also simply computes the percentage of the population and the workforce who work in health care (1993, p. 156). This amounts to weighting all workers the same. The basic message is similar, but this method is less similar to the Reports than the weighting discussed above.

⁴⁸ One could also examine the number of physician visits *per capita*. On this measure, the U.S. is quite low, at about 3.6, while Germany is 8.5 and France is 7.0. The U.S. ranks 15th out of 18 (van Doorslaer, Masseria and Koolman 2006, p. 181).

Table 3: Expenditures and Physical Resources, cir. 1990

Country	GDP PPP Ex Rate, 1990*	Health PPP Ex Rate 1990*	Weighted Health Workers % of Workforce 1988**	Weighted Physicians and Nurses, per 1,000 Pop 1990***	Physicians Per 1,000 Pop 1990****
Australia	1630.259	2221.627	8.24	31.60	2.3
Austria	1998.736	3257.544			2.2
Belgium	1748.302	3265.15		20.50	3.3
Canada	2013.446	2814.494	6.39	36.90	2.1
Denmark	1932.548	2614.788		14.10	3.4
Finland	1493.159	2116.968	7.65	29.10	3.1
France	2069.349	3778.562	8.31	28.15	3.1
Ireland	953.1897	1370.21	5.94	19.10	1.6
Italy	1710.289	2771.807	5.12	17.70	4.7
Netherlands	1815.531	2896.84		22.20	2.5
New Zealand	1243.447	1924.952	6.8	17.00	1.9
Norway	2243.863	3585.022	10.66		2.6
Portugal	1007.868	1578.589	4.52	11.00	2.1
Spain	1155.262	1935.702			2.3
Sweden	1940.782	2882.35	10.87		2.9
Switzerland	2432.971	3167.182		51.6	3
United Kingdom	1189.661	2099.403	5.52	23.70	1.4
United States	3356	3356	7.03	30.70	2.4

Note: OECD data is interpolated between 1988 and 1993 for France, 1991 data for Switzerland

Sources: *Comanor, Frech and Miller, 2006, p. 8, **Pauly. 1993, p. 156,

OECD Dataset, 2008 *Anderson, Reinhardt, Hussey and Petrosyan, 2003, p. 95

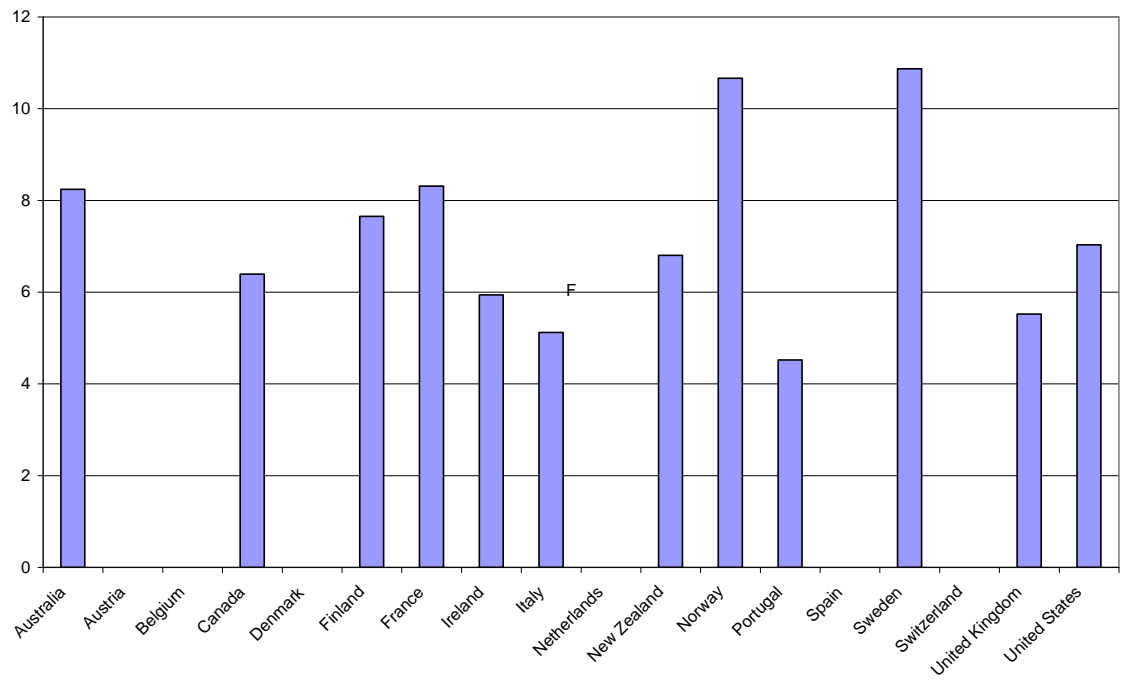
Figure 3, Weighted Health Workers as a Percent of Workforce, 1988

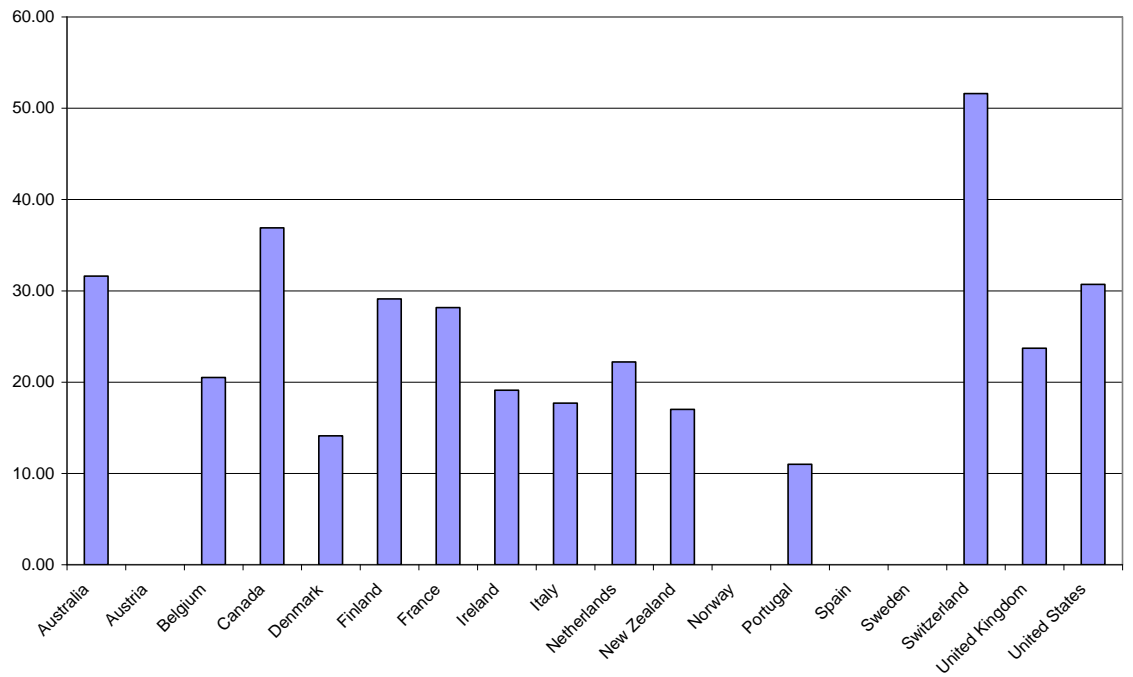
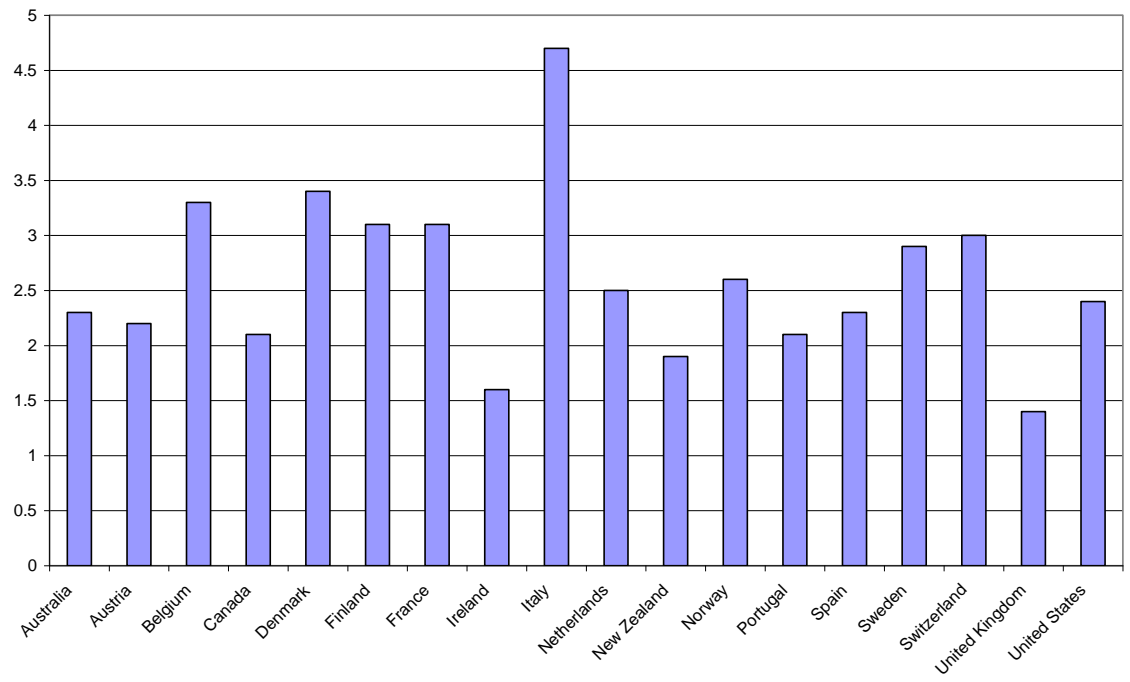
Figure 4, Weighted Physicians and Nurses per 1,000 Population 1990

Figure 5, Physicians per 1,000 Population 1990



c. Pharmaceuticals

These measures discussed so far are focused on labor inputs. Of course, there are nonlabor inputs as well. For the particular case of pharmaceuticals, previous work can be analyzed to get a particularly clear picture of the inaccuracies of the GDP PPP exchange rate.⁴⁹ As discussed above, it is only appropriate to convert spending in other countries to U.S. dollars by the GDP PPP exchange rate if health care prices differ across countries in the same way that prices differ in general. In the case of pharmaceuticals, one might think that this would be true because pharmaceuticals are traded internationally. However, researchers including Tadeusz Szuba (1986) and Patricia Danzon and Allison Percy (1995), have demonstrated that this is far from the truth.⁵⁰

For instance, relatively strict price regulation is practiced in France and Italy. Other OECD countries, such as the United Kingdom and Germany, also regulate pharmaceutical prices, indirectly and typically much less stringently.⁵¹ The U.S. at the other extreme, generally permits free pricing of pharmaceuticals, subject to market forces. Thus, one would expect GDP PPP exchange rates to be inaccurate for cross-national comparisons.

Table 4 presents measures of *per capita* pharmaceutical expenditures converted to U.S. dollars using pharmaceutical PPP exchange rates and GDP PPP exchange rates

⁴⁹ This discussion is partly based on Frech and Miller, (1999, pp. 23-28).

⁵⁰ There is also the issue of price discrimination across countries, usually favoring poor countries, by manufacturers of patented pharmaceuticals (Schut and Van Berek 1986; Danzon 1997a; Danzon and Towse 2003). However, within the OECD, lower prices do not track lower income countries, suggesting that price discrimination is not the primary cause of price differences.

⁵¹ See Garattini et al. (1994) for a comparison of the pharmaceutical markets and price regulation in Italy, France, Germany, and the United Kingdom.

Table 4: Comparing Measures of Real Pharmaceutical Consumption Using Pharmaceutical Purchasing Power Parity and GDP Purchasing Power Parity Exchange Rates, 1990

Rank	Country	Pharm PPP	GDP PPP	Difference
1	France	560.927	256.123	304.804
2	Italy	448.060	242.160	205.899
3	Germany	374.138	311.782	62.356
4	Luxembourg	348.086	224.572	123.514
5	Belgium	304.466	193.561	110.904
6	Spain	286.618	144.749	141.870
7	Canada	262.609	231.773	30.836
8	Iceland	248.776	212.707	36.068
9	United States	236.000	236.000	0.000
10	Sweden	225.859	119.751	106.107
11	Norway	216.341	125.167	91.174
12	Greece	216.032	95.128	120.904
13	Australia	197.590	118.241	79.350
14	New Zealand	194.828	140.460	54.360
15	Austria	191.940	154.345	37.595
16	Finland	190.418	121.397	69.020
17	Switzerland	188.095	143.767	44.328
18	United Kingdom	183.721	131.229	52.492
19	Portugal	152.193	94.368	57.824
20	Netherlands	130.189	127.483	2.706
21	Ireland	120.482	101.449	19.033
22	Denmark	112.846	95.390	17.456
23	Turkey	61.818	34.703	27.116

Source: Frech and Miller 1999, p. 24

for 1990. Just as it does for health care in general, the GDP PPP exchange rate underestimates real pharmaceutical expenditures outside of the U.S. For example, the pharmaceutical PPP exchange rate gives an estimate of \$561 for France, while the estimate is only \$256 using the GDP PPP exchange rate. The biggest differences are found for those countries with the strictest price regulations, France and Italy.

More recent analysis by Danzon and Michael Furukawa (2008 p. 228) for a larger set of countries and more recent data (from 2005), confirms the result: U.S. prices are relatively higher. At the consumer level, prices in the rich countries of the OECD ranged from a low of 66 percent of U.S. prices (Australia) to a high of 95 percent (Germany). France's prices were 78 percent of U.S. prices. On average, non-U.S. prices were about 77 percent of U.S. prices. Thus, U.S. real resources or quantities are overstated. Interestingly, the difference was substantially greater for manufacturers' prices than for retail prices. Distribution costs are apparently lower in the U.S.

Danzon and Furakawa (2008 p. 223, 224) also examined physical usage, with their detailed proprietary data. Physical usage showed the same pattern. Although U.S. spending evaluated at GDP PPP exchange rates was the highest of the comparison countries, the U.S. health care system uses fewer actual doses *per capita* than most other countries. An alternative physical measurement is grams of active ingredient *per capita*. Using this measure, the U.S. places higher, but still not near the top of the distribution. France is the highest: 71 percent higher than the U.S. Note that this is consistent with the real pharmaceutical spending as evaluated with the pharmaceutical PPP exchange rate above in Table 4. France was the highest in that data in 1990 also, by an even higher percentage.

Danzon and Percy (1995) provide highly accurate Fisher price indexes to convert pharmaceutical expenditures to U.S. dollars.⁵² These price indexes are carefully calculated based on proprietary data that is only available for a few countries. These measures should be regarded as the "gold standard" for this time period. They are only available for France, Italy, Germany and the United Kingdom and the U.S. Further, the detailed prices were converted from other currencies into dollars using market exchange rates, rather than PPP exchange rates. Though not ideal, this seems to have had little effect on the rankings of consumption of various countries.

One can convert pharmaceutical consumption in France, Italy, Germany and the United Kingdom to U.S. dollars using the using the Danzon and Percy measures and also the pharmaceutical PPP and the GDP PPP exchange rates for 1980, 1985 and 1990. Correlations among the these three measures showed that pharmaceutical consumption using the Danzon and Percy measure was substantially more closely correlated with the consumptions using the pharmaceutical PPP exchange rate than with consumption using the GDP PPP exchange rate. See Table 6.

Tadeusz Szuba (1986) also assembled price ratios for the single year 1983 using proprietary data, though with a slightly different approach. He includes the same countries as Danzon and Percy, plus Switzerland. Szuba's price coefficients are converted to U.S. dollars using market exchange rates (like Danzon and Percy). See Table 5 for a comparison of all the methods He found that in 1983 Italy had the

⁵²The Fisher price index is the geometric mean of the Laspeyres and Paasche price indexes. Like both the Laspeyres and Paasche indexes, it is intransitive. For example, the product of the indexes between the U.S. and Canada and between Canada and Denmark is not equal to the index between the U.S. and Denmark. Unlike the other two indexes, the Fisher price index yields results which are invariant to which country is used as a base.

lowest pharmaceutical prices and that the U.S. had the highest pharmaceutical prices among the six countries he studied (those countries listed in Table 5). We apply his price coefficients to 1985 expenditure estimates which have been converted to U.S. dollars using the market exchange rates.

In Table 5 one can see differences among the measures, but a common pattern emerges. France seems to be the highest pharmaceutical user while the U.S., Switzerland and the United Kingdom are lower users.

Table 6 presents correlations among the different measures of pharmaceutical consumption for 1985. The consumption using the pharmaceutical PPP measure is highly correlated with consumption using both the gold standard Danzon and Percy measure and the Szuba measure. Importantly, consumption using the GDP PPP exchange rates is not highly correlated with the consumption using the Danzon and Percy or Szuba measures.

Table 5: Comparing Measures of Real Pharmaceutical Expenditures for 1985 in Six Countries Using Various Conversions to 1985 U.S. Dollars					
Country	Market Exchange Rates	GDP PPP	Pharm PPP	Danzon & Percy's Fisher Price Indexes	Szuba's Price Coefficients
France	129.48	176.23	401.38	387.16	556.2
Italy	94.31	148.01	269.99	258.97	457.06
Germany	174.15	231.67	257.29	290.83	256.79
Switzerland	102.44	115.33	NA	NA	147.81
U. S.	151.00	151.00	151.00	151.00	151.00
U. K.	66.67	94.72	207.17	103.44	126.01
Source: Frech and Miller 1999 p. 26					

Table 6: Correlations Among Pharmaceutical Consumption Measures, 1985

	Market X Rate	GDP PPP X Rate	Pharmaceutical PPP X Rate	Danzon & Percy Index	Szuba Coefficients
Market X Rate	1.000	0.988***	0.527	0.686	0.225
GDP PPP X Rate		1.000	0.574	0.728	0.313
Pharmaceutical PPP X Rate			1.000	0.979***	0.924***
Danzon & Percy Index				1.000	0.851
Szuba Coefficients					1.000

Source: Frech and Miller 1999, p. 26 and calculations of the author
 **Significant at the 1 % level

At best, though, the pharmaceutical PPP measure of real pharmaceutical consumption is an approximation. Subtle differences are missed. The most important problem is differing patterns of consumption. These differences appear to be substantial. Livio Garratini and his colleagues (1994) attempted to analyze the prices of the 100 best selling pharmaceuticals in 1992 in Italy, Germany, France, and the United Kingdom. Only eight products were common in the top 100 in each country. They found similar inconsistencies when they examined consumption by therapeutic classes. Danzon and Percy (1995) and Szuba (1986) also found large differences across countries.

V. Results of the Panel Data Regressions

The main econometric results of the Report are reproduced below. These Tables 7 and 8 here, reproduced from Tables 3 and 4 in the Report present the econometric results of the main analyses. As one can see, the basic regression is run for different output measures and, in some cases, separately for males and females. Table 7 (Table 3 in the Report) measures health resources as weighted physicians and nurses, while Table 8 (Table 4 in the Report) measures health resources in expenditures, evaluated at GDP PPP exchange rates. Since the specification is log-log, all the coefficients can be interpreted as elasticities. I am going to focus on LE and infant mortality. PYLL is less intuitive and more difficult to interpret. In particular, the elasticities (coefficients) have a very different interpretation in terms of life years gained. In this case, the PYLL is defined as the potential years of life lost per 100,000 people aged zero to 70, the assumed potential life span. The average is 3,158. That is only 0.032 years per person per year. When accumulated over a 70 year lifetime, this is still only 2.2 years.. Deaths before age 70 are fairly rare. Driving PYLL to zero, obviously impossible, would increase the expected number of years lived by only 2.2 years. A health care service that prevented some early deaths could cause a large percentage change in PYLL and a small percentage change in LE. Thus, we observe much larger elasticities of almost any input with respect to PYLL, compared to elasticities of life expectancies.

Sampling Error, Statistical Significance and Interpolation

The levels of statistical significance for the coefficients are indicated in the tables. They are generally high, usually higher than the 1.0 percent level on a two-tailed test. Significance at the 1.0 percent level means that a coefficient this far from zero in either a positive or a negative direction is unlikely to occur (probability less than 1.0 percent), based on sampling error, assuming that the specification is correct. The high observed level of significance indicates that sampling error is not generally a large problem. The possibility of specification error, systematic measurement error in one or more variables and unsystematic measurement error in the independent variables remains. Further, it may be that some of the data is interpolated over time because of missing observations. This would artificially increase statistical significance levels. In the discussion that follows, I will not comment on the observed level of statistical significance, except for the relatively few cases where it is noticeably lower than 1.0 percent

Table 7: Reproduced from Table 3 of Report

Table 3. **Health status determinants, with health care resources measured by practitioners**^{1,2}
Econometric results for the main scenario

Dependent variables ³ Explanatory Variables ³	Life expectancy at birth			Life expectancy at 65		Premature mortality (adjusted)			Infant Mortality
	Female	Male	Total	Female	Male	Female	Male	Total	
Constant	3.940***	3.650***	3.800***	2.090***	1.570***	11.600**	12.000***	12.000***	10.600***
Practitioners	0.013***	0.017***	0.015***	0.032**	0.043***	-0.089**	-0.062	-0.072*	-0.440***
Smoking	-0.007***	-0.018***	-0.014***	-0.028***	-0.070***	0.060**	0.190***	0.150***	0.072
Alcohol	-0.011***	-0.018***	-0.015***	-0.024*	-0.010	0.290***	0.040	0.130**	0.370***
Diet	0.003	0.004	0.004	0.002	0.008	0.088**	0.030	0.055*	0.120*
Pollution	-0.003	-0.012***	-0.006**	-0.032***	-0.058***	0.150***	0.170***	0.160***	0.190***
Education	0.040***	0.045***	0.042***	0.056***	0.046***	-0.250***	-0.300***	-0.260***	-0.500***
GDP	0.035***	0.066***	0.051***	0.099***	0.170***	-0.480***	-0.480***	-0.510***	-0.870***
Number of Observations	254	254	254	254	254	236	237	236	254
Number of Countries	22	22	22	22	22	21	21	21	22

Notes:

1. Generalised least square regressions, with country-fixed effects, error terms following a country-specific AR(1) and correction for heteroskedasticity.

*** indicates significance at 1%; ** indicates at 5% and * indicates significance at 10%.

2. Practitioners are calculated as the number of practising physicians and half the numbers of practising nurses.

3. Details on individual variables are provided in Annex 1.

Source: OECD calculations.

Reproduced from (Journard, André, Nicq and Chatal 2008, p. 22, Table 3)

Table 8: Reproduced from Table 4 of Report

Table 4. **Health status determinants, with health care resources measured by spending¹**
Econometric results for the main scenario

Dependent variables ²	Life expectancy at birth			Life expectancy at 65		Premature mortality (adjusted)			Infant mortality
	Female	Male	Total	Female	Male	Female	Male	Total	
Constant	4.009***	3.641***	3.825***	2.178***	1.638***	11.172***	12.871***	12.244***	8.516***
Spending	0.035***	0.045***	0.041***	0.051***	0.061***	-0.272***	-0.300***	-0.282***	-0.572***
Smoking	-0.000	-0.006**	-0.004	-0.019***	-0.057***	0.063***	0.088***	0.077***	0.077*
Alcohol	-0.011***	-0.014***	-0.011***	-0.017	-0.004	0.234***	0.082*	0.115***	0.327***
Diet	0.003	0.004	0.004	0.013*	0.028***	0.044*	0.001	0.014	0.044
Pollution	-0.009***	-0.018***	-0.012***	-0.037***	-0.068***	0.169***	0.153***	0.162***	0.320***
Education	0.029***	0.031***	0.030***	0.064***	0.045***	-0.107**	-0.227***	-0.182***	-0.378***
GDP	0.006	0.035***	0.019***	0.044***	0.107***	-0.285***	-0.292***	-0.292***	-0.379***
Number of Observations	325	325	325	325	325	307	307	307	325
Number of Countries	23	23	23	23	23	22	22	22	23

Notes:

1. Generalised least square regressions, with country-fixed effects, error terms following a country-specific AR(1) and correction for heteroskedasticity.

*** indicates significance at 1%; ** indicates at 5% and * indicates significance at 10%.

2. Details on individual variables are provided in Annex 1.

Source: OECD calculations.

Reproduced from (Journard, André, Nicq and Chatal 2008, p. 23, Table 4)

A. Lifestyle Variables

In the OECD study, alcohol consumption comes out as generally harmful to the production of health, especially in infant mortality. The elasticity of infant mortality with respect to alcohol consumption is very large indeed, estimated at 0.370 and 0.327 in Tables 7 and 8. This implies that a doubling of alcohol consumption would increase infant mortality by between 33 and 37 percent. This is broadly consistent with the medical literature, which suggests that drinking during pregnancy has harmful health effects on the infant (Windham, Fenster, Hopkins and Swan 1995; Walker, Tempkin and Wallace 2009). Further, alcohol consumption during pregnancy is probably correlated with consumption of other recreational drugs, such as cocaine, which also are known to have harmful effects on fetal and infant health, so that the alcohol variable is probably picking up some of this effect as well. This result underscores the idea that health in general, and especially infant mortality, is heavily influenced by lifestyle variables. Since lifestyle is measured very incompletely at best, this leads to omitted variable bias in what limited lifestyle variables are used.

The effects on LE at 65 for are about a tenth the size of the effects on infant mortality, but they still show important negative effects. This negative result is somewhat surprising, given the evidence that moderate alcohol consumption leads to improved cardiovascular health and health outcomes.⁵³ Of course, heavy drinking is harmful to health. The *per capita* alcohol consumption measure is a broad average and hence probably a poor measure of heavy drinking. It is important to remember that excessive alcohol consumption, like most lifestyle variables, affects health with a lag (except for infant mortality). That lag is not allowed for in the Report's specification.

Smoking also has a generally negative effect on the health measures, with quite large effects on infant mortality. The elasticities are 0.072 and 0.077 in the models with different measures of real health care resources used. This, and the other effects are about one third the size of the effects of alcohol consumption. With the exception of infant mortality, smoking also affects health with a lag and that lag is not captured in this work.

The only other lifestyle variable included here is a measure of diet. It seems clear that the major dietary issue in the OECD countries relates to obesity--overeating relative to physical activity. Consumption of calorie-dense animal fat would seem to be the best dietary correlate of obesity and, indeed, it has been used in past studies (Frech and Miller 1999, p. 35; Hertz, Herbert and Landon 1994, p. 16). Total calories or fat plus sugar calories might also make sense as proxies for obesity. The animal fat variable does have some effect in the past studies, though it is a poor proxy for obesity. The Report, however, goes in a different direction. It uses the consumption of fruit and vegetables (Joumard, André, Nicq and Chatal 2008, p. 27).

The authors note that sugar and/or fat consumption might make sense, but that they are likely to have nonmonotonic effects—positive to health at low levels and then turning negative.⁵⁴ Indeed, that is what Frech and Miller found for animal fat in OECD

⁵³ See, e.g., Michael, Colditz, Coakley and Kawachi (1991, p. 15); Gaziano, *et al*, (1993).

⁵⁴ The switching point is called the epidemiological transition. See Gage and O'Connor (1994).

data (1999, p. 42, 43), using a quadratic (in the logs) functional form to allow for this sign reversal.⁵⁵ The authors of the Report state that using similar variables led to unstable or inconsistent results (Joumard, André, Nicq and Chatal 2008, p. 27). It is unclear whether they estimated a flexible functional form that allowed for a nonmonotonic relationship. In any case, the fruit and vegetable diet variable has a generally very small effect and the effects are not tightly estimated. Most of the estimates are not statistically significant at even the 10 percent level. Probably, some weak correlations with powerful omitted variables—such as obesity—is the reason for getting any results at all.

Obesity Considered and Not Used

The Report briefly considers obesity:

Obesity is sometimes considered as a determinant of the population health status because it can be considered as a proxy for a broad range of nutritional and physical activity patterns. In practice, obese people tend to die at a younger age (Joumard, André, Nicq and Chatal 2008, p. 51).

This is correct. Furthermore, obese people consume many more health care resources. Roland Sturm (2002) has shown that, in U.S. data, obesity leads to 36 percent higher total health care consumption and impressive 77 percent higher consumption of pharmaceuticals. Other work by Eric Finkelstein, Ian Flebelkorn and Guijin Wang (2003, pp. w3-219, w3-224) confirms this with a larger dataset and calculates that 5.3 percent of U.S. health care spending is caused by obesity and 9.1 percent are caused by obesity and overweight taken together. The OECD Report goes on to say:

Data on obesity are, however, not easily comparable: 28 countries collect data for obesity but on a very irregular basis. Furthermore, in some countries data refer to self-reported status, while in others they are derived from actual heights and weights (Joumard, André, Nicq and Chatal 2008, p. 51).

This is a problem, but the magnitude of the error from different measurements is probably far smaller than the error caused by omitting obesity. The variation in obesity across OECD countries is so large that it probably overwhelms the measurement error. For example, in the mid 1990s, female obesity was 25.1 percent in the U.S. and only 6.8 percent in France (Miler and Frech 2004, p. 68). Further, the self-report problem has been studied in the literature, leading to adjustments for the differences between self-reported versus physical measures of BMI (Shields, Gorber and Tremblay 2008a, 2008b; Baum and Rheum 2007, p. 7; Michaud, van Soest and Andreyeva 2007, pp. 5, 6, 21-23). In the OECD data, most countries use self-reported data (25 out of 30 in both 1995 and

⁵⁵ The log-log form imposes monotonicity—ruling out U-shaped or inverted U-shaped relationships where the effect changes sign. Adding a term where the log is squared relaxes this monotonicity constraint. It also allows testing on whether that new squared term is statistically significant.

2005) (OECD 2007b, p. 10). Another simple approach would be to add a dummy variable for the countries that use actual measurements.

In the same section, the Report also raises a conceptual issue:

More fundamentally, one could question whether obesity should be considered as a determinant of the population health status (*i.e.* a right-hand side term of the health status production equation) or instead as a measure of the health status itself (left-hand side term). It is clear, in practice, that obesity is influenced by education, income, lifestyle factors and, though probably less, by health-care resources (Joumard, André, Nicq and Chatal 2008, p. 51).

The authors argue that obesity may be an output of the health care system. This reflects a definition of health care that is so wide as to be meaningless. At a basic physical level, obesity results from consuming too many calories, relative to the level of physical activity. It varies greatly across countries (including within the OECD and within Europe) and over time. Further, there is a steep education/obesity gradient within countries (Baum and Ruhm 2007; Cutler, Glaeser and Shapiro 2003; Michaud, van Soest and Anfeyeva 2007). The later two papers, Cutler, Glaeser and Shapiro and Michaud, van Soest and Anfeyeva are focused on the causes of obesity. Both studies find that time spent in sedentary activities and especially differences in time spent eating and cooking are very important. Obesity appears to be fundamentally a result of lifestyle choices and the economic and physical environment, only slightly if at all, influenced by the medical system. It is not plausible that better physicians or more physicians or more complete health insurance affect this by very much. Social norms and traditions, length of work days, industry mix and even urban design probably have a much larger effect. Further, obesity has a very powerful effect on shortening LE and in leading to higher spending on health care. Both effects make the health care system in more obese countries look inefficient. Excluding obesity leads to large upward bias in the apparent inefficiency of the U.S. health care system. See (Comanor, Frech and Miller 2006) and the discussion above.

B. Socioeconomic Variables

There are a great number of variables one might think of as socioeconomic variables that help produce health. In the Report, however, the category contains only air pollution, education and income (GDP).

The Report measures exposure to pollution by *per capita* nitrogen oxide (NO_x) emissions. This is probably a poor measure for many reasons. First, NO_x is not generally considered to be the most important air pollutant. Particulates, and possibly sulfur dioxide (SO₂), are considered more important from a health viewpoint (Liu 1989, p. 188; Chay and Greenstone 2003, pp. 1121-1126).^{56,57}

⁵⁶ There are also visual aspects of air pollution, which are not necessarily closely related to the health effects.

Second, even if NOx were a good choice, the variable does not measure exposure, it measures emissions. A variable for emissions is not the relevant input for the production of health. The correct measure would be the NOx content of ambient air. Assuming a dose/response curve that is nearly linear in the relevant range, a reasonable approximation in aggregate data would be the weighted (by local population) average of NOx exposure.⁵⁸ NOx emissions *per capita* is probably a very poor measure. Also, it is fairly highly correlated with the income measure, *per capita* GDP, 0.5539 in the OECD dataset. This correlation complicates the interpretation of the coefficients on income and on NOx emissions. Within the limits of easily available data, the measure could be improved by using NOx emissions per square mile.⁵⁹ Using emissions *per capita* makes a country like Canada or Iceland incorrectly look bad (high emissions, but low exposure because of locations of pollution verses population) (Joumard, André, Nicq and Chatal 2008, p. 26). Probably, NOx emissions are fairly strongly correlated with other pollutants and also with the industrial structure.

In the reported regressions, the NOx emission variable has a surprisingly strong negative impact on health in most regressions. In the infant mortality equations it is particularly important, with an elasticity of 0.190 when real health care resources are measured by weighted physicians and nurses and 0.320 when real health care resources are measured by spending. Given that NOx emission is such a poor measure of actual exposure to pollution, these results are difficult to interpret. In spite of these definitional and measurement problems, the NOx emission variable may be picking up some pollution exposure effects, along with some effects of industrial composition.

The next socioeconomic variable is education. Going back to the very beginning of research on the household production of health, education has been viewed as an important determinant of health (Grossman 1972). Eliot Jamison, Dean Jamison and Eric Hanushek have found a large effect of both quantity and quality of education on infant mortality across a large and diverse set of countries and 40 years (2007, pp. 782-784). Within the rich countries, health differences across educational groups are large (Banks, Marmot, Oldfield and Smith 2006; Rachel Tolbert Kimbro, Sharon Bzostek, Noreen Goldman, and Germán Rodríguez 2008). However, it is not so clear that the relationship is causal, rather than simply picking up the effects of omitted variables, such as higher ability people both obtaining more education and being healthier or reverse causation (better health allows one to obtain more education).

An import recent study by Damon Clark and Heather Royer (2008) uses a powerful natural experiment in the U.K., where the compulsory education level was raised all at once, to isolate the truly causal relationship. This study finds virtually no effect. And the result is statistically strong—it rules out major effects. While it may not be the last word, this study raised doubts about the belief that education has much impact on health in the rich and highly-educated countries.

⁵⁷ For a recent study showing a large effect of particulate pollution on life expectancy, see C. Arden Pope III, Majid Ezzati and Douglas Dockery (2009)

⁵⁸ Such a weighted average exposure variable is not available at the country level (Greenstone 2008).

⁵⁹ This measure would be exactly correct if the population and the emission sources were evenly spread out over the country and wind was not an issue. The emissions *per capita* variable, on the other hand, is highly biased (overstating exposure in sparsely populated countries), even under these idealized circumstances.

The mechanism by which education is thought to influence health is not so clear, but, as the Report states, better educated people are more likely make better health choices, such as complying with treatments and medical advice.⁶⁰ This cognitive interpretation is favored by Jamison, Jamison and Hanushek (2007, pp. 783-784) and also by David Cutler, Angus Deaton and Adriana Lleras-Muney (2006, pp. 113-115). The Report goes on to say that education may affect choice of lifestyles, such as “smoking less, exercising more, etc.” (Joumard, André, Nicq and Chatal 2008, p. 28). But, then it maintains that these lifestyle effects should not be the mechanism reflected in the coefficients from the regression, because lifestyle variables are controlled for separately. Given the weaknesses of the lifestyle variables, this argument seems dubious. As one example of the connection between education and lifestyle, Charles Baum and Christopher Ruhm have shown that education of a person’s mother is a powerful determinant of obesity (2007). As discussed above, there is a great deal of unmeasured lifestyle variation in the data. Some of the unmeasured lifestyle effects are partly picked up in the education variable and also partly picked up in the income variable and in the country-specific dummy variables.

In any case, consistently measuring education across different countries with different educational systems is challenging indeed. Here, the education measure is defined as the proportion of people, aged 25 to 64, who have attained an upper secondary education. This corresponds to finishing high school in the U.S. This is a imperfect measure, partly because the knowledge and mastery necessary to complete “upper secondary education” varies substantially across countries. For example, the Report notes that this proportion is low in Australia, yet the mean number of years of schooling completed is fairly high (Joumard, André, Nicq and Chatal 2008, pp. 51-52). The data are not presented, but this is probably true of some other countries, like Germany, where completing high school is more demanding and normally takes longer than in the U.S.

In the regressions, the effects of education are fairly large. Elasticities in LE regressions range from 0.029 to about 0.064. The effect on infant mortality is very large, -0.500 when real health care resources are measured by spending and -0.378 when they are measured by weighted physicians and nurses. Either of these is a very large effect, implying that increasing the proportion of people educated to this level by 10 percent (about 7 percentage points)⁶¹ would decrease infant mortality by about 4 or 5 percent. As is discussed above, the controls on lifestyle in the analysis are quite incomplete; most of the apparent effect of education in the regression probably comes from picking up correlated and unmeasured cultural and lifestyle factors. It is an open question to what extent more educational attainment causes consumers to adopt healthier lifestyles or is simply correlated with them.

The most important socioeconomic variable, of course, is income. In these regressions, GDP *per capita* is the measure. It is a nearly universal result that populations with higher wealth or income are healthier. The effect is more obvious and more often well-estimated in samples including both poor and rich countries where there is more variation in income, but it continues throughout the income range (Prichett and

⁶⁰ For an interesting demonstration of the relationship between education and compliance, specifically for HIV infection and diabetes, see Dana Goldman and James Smith (2002).

⁶¹ The OECD mean percentage aged 25 to 64 year olds who have attained upper secondary educations is about 65 percent.

Summers 1996). More income can purchase better living conditions and better and less risky working conditions. Higher income can also pay for more health care and more education, both of which are imperfectly measured across countries.

In most of the regressions, GDP has a strong effect, larger when real health care consumption is measured by weighted physicians and nurses than when it is measured by spending. For LE, the elasticities vary quite a bit, ranging from 0.006 (and insignificant at even the 10 percent level) to 0.170. Generally speaking, the effects are quite a bit larger for males. The effects on infant mortality are -0.379 when real resources are measured by spending and -0.870 when they are measured by weighted physicians and nurses. These are very large estimates. The effect of income on infant mortality has been very carefully measured by Lant Prichett and Larry Summers in a much larger set of countries where one might expect a much larger effect than within the OECD countries.. They find an elasticity between -0.2 and -0.4 (1996, p. 866).

There is a substantial problem with this GDP *per capita* variable in this OECD study (and in any cross-country study). GDP *per capita* is highly correlated with real health care resources: 0.7717 with weighted physicians and nurses and 0.9504 with health care spending. That has two implications for interpretation of the effect of income on health. First, the ability of any statistical procedure to apportion the effects between income and health care resources is limited. Second, it is important to measure real health care resources well in order to control for it in order to avoid omitted variable bias in the coefficient on income.

C. Health Care Resources

1. Physical Measures

The OECD researchers report experimenting with several physical measures (Joumard, André, Nicq and Chatal 2008, p. 34). A variable for hospital beds *per capita* was entered into the equation, but was generally statistically insignificant or had the wrong sign.⁶² The number of scanners (e.g. MRI machines) was always insignificant. In the end, the Report used the variable described above, weighted physicians and nurses, with the weight on the nurses being one half of a physician. In the regressions, this variable has a notable positive effect on health. In the LE regressions, the elasticities range from 0.013 for females at birth to 0.043 for males at age 65. Since mean LE at birth is about 75 years for females, an elasticity of 0.013 implies that doubling health care resources would increase LE by almost one year.

In highly related work, also using a panel of OECD countries, Zeynep Or also found health care resources to be highly productive (2000a, 2000b). The earlier work focused on the effects of health spending on PYLL and will be discussed below. The later work used very similar techniques to the current Report, but measured health resources with the physician/population ratio only. Or found estimated elasticities of

⁶² The weak results for hospital bed make sense, given the international differences in hospital organization. For example, small, primitive physician-owned hospitals are common in Japan. As an indicator of how different they are from a high tech American hospital is provided by comparing length of stay. The length of hospital stay in Japan averaged 36.5 days, while it was only 6.5 days in the U.S., in 2004 (Ramayer 2009, p4.)

about 0.10 for LE at 65, roughly twice as large as the effects found in the Report. More recently, Dynep Or, Jia Wang and Dean Jamison analyzed the impact of the physician/population ratio on LE in a panel of OECD countries (2005). In their analysis that is most comparable to the Report, Or, Wang and Jamison find elasticities with respect to physicians varying between 0.037 and 0.077 for the LE analyses (p. 558). This is a substantially bigger effect than the Report finds. Note that even Or's lowest elasticity indicates that health care is very productive. A doubling of resources would raise female LE by over two years.

In percentage terms, LE at age 65 is substantially more sensitive to health resources than at earlier ages, roughly twice as sensitive. Most of the difference is simply a mechanical implication of the fact that most death occur after age 65. LE at birth is about 70 to 75 years in most of this data, while LE at 65 is about 15 to 20 years. Thus, an elasticity at the later years of twice as high implies, somewhat paradoxically, a smaller number of life years gained, not a larger number of life years gained. Frech and Miller (1999) and Miller and Frech (2004) find similar results when comparing the productivity of pharmaceuticals at 40 to that at 60. It would have been interesting to see the effect of health care resources on LE at 40.

The effects of health care resources on PYLL are larger in terms of elasticities, ranging from -0.062 for males to -0.089 for females and -0.072 to the total sample. However, they are not very precisely estimated, being statistically insignificant for males and statistically significant at only the 5 percent level for females. In Or's previous work with similar panel data, she found less consistency, but even larger effects, as high as -0.38 for women and -0.28 for men (2000b). The results are not exactly comparable, because the PYLL in the Report has apparently been defined to exclude deaths from land transport accidents, accidental falls, suicides and assaults (Joumard, André, Nicq and Chatal 2008, pp. 8, 19, 47).

As is discussed above, the effects of any input on LE and PYLL are not comparable, even though they both are expressed in years. These large elasticities for effects on PYLL do not translate into large effects on life years. For example, if we take the Report's largest estimate of -0.072 as if it were correct, this implies that a doubling of resources would cause a 7.2 percent decline in PYLL. But, the average is only 3,158 per 100,000 people 0-70 years old, or 0.032 years per person per year. Reducing that by 7.2 percent, we get an increase of 0.0023 years per year. Even accumulating these effects over 70 years, this is only 0.15 years. The effect of health care on PYLL is small because health care does not have so much effect at the earlier ages. Even though the effect is statistically significant, it is not so significant from a scientific or policy viewpoint. In the Report, as we have seen, even poorly measured lifestyle and socioeconomic factors seem to be much more powerful for PYLL than for LE. It is not clear why this should be, especially with the exclusions of some lifestyle-related causes of death.

Looking at infant mortality, weighted physicians and nurses have a notable effect, with an elasticity of -0.440. This is a large effect, though not as large at GDP *per capita*, education or alcohol consumption. Or, Wang and Jamison also estimated large effects of physicians alone, -0.548 in the most comparable formulation (2005, p. 558). Considering the importance of omitted lifestyle variables and the measurement problems with infant mortality, discussed above, it is difficult to know what weight to give to the results for infant mortality.

2. Health Care Expenditures

The OECD researchers also used a measure of resources in monetary terms. As is discussed above, they converted the spending in any one country and year to U.S. dollars at constant prices, using the GDP PPP. The results show a generally larger effect on health outcome than did weighted physicians and nurses. Also, the elasticities are more constant across differing LE measures. The elasticities for LE range from 0.035 to 0.061. This is a significant benefit. These results are roughly comparable to, though slightly larger than, the results for the productivity of pharmaceuticals in cross sectional analyses (Frech and Miller 1999, p. 42; Miller and Frech 2004, p. 39; Shaw, Horrace and Vogel 2005, p. 775). These works do not contain reliable estimates for non-pharmaceutical health care, probably because it is so correlated with income.⁶³ Nixon and Ulmann obtained a lower estimate, roughly half of the Report's (2006, p. 15), using the same GDP PPP exchange rate as the Report.

Using health expenditures, the estimates rise very little with age (comparing LE at birth to LE at age 65). For females, the elasticity is 0.035, rising to only 0.051 at 65. For males, the corresponding elasticities are 0.045 and 0.061. In comparable work, also using a panel of OECD countries (for a slightly earlier period, 1970-2000) for LE at 60, Peter Zweifel, Lukas Steinmann and Patrick Eugster found elasticities that are somewhat lower.⁶⁴ Comparing these Report's results for LE at 65 to those for LE at birth implies that health care has a substantially smaller effect on life years for older people. This is surprising.

For PYLL, the effect of health care resources are much larger than it is for LE. The elasticities vary in a tight range -0.272 to -0.300. Also, all of these results were highly statistically significant, in contrast to the PYLL estimates using weighted physician and nurses. Some of Or's previous work also used expenditures. She found less consistency across males and females and also smaller estimates, -0.18 for women and a very small, -0.04 for men (2000a).

Turning to infant mortality, again, the estimated effects are large, at -0.572. This is a large effect and larger than any single other input. John Nixon and Philippe Ulmann (2006), find similarly large effects of health care spending on infant mortality (2006, p. 15). As with the other infant mortality estimates, it is hard to know what to make of the results, given the specification and data problems discussed above.

3. Physical Resources v. Expenditures

Comparing the results with the two measures, one can say that the real expenditures measure gives generally larger effects on health measures. However, both measures are flawed, as is explained above. Also, there are major problems of omitted

⁶³ These papers used the health care PPP exchange rate, while the Report used the GDP PPP exchange rate.

⁶⁴ Zweifel, Steinmann and Eugster (2005) use a quadratic function, not a log-log function. Therefore, elasticities are not constant and must be calculated at some specified values. When calculated at the mean, these values are 0.035 for females and 0.045 for males (2005, pp. 135-137) Their equation has fewer controls on lifestyle and socioeconomic variables and does not include country-specific fixed effects. One might therefore have expected a larger effect for health care spending, rather than the somewhat smaller effect they obtain.

variables that are correlated with both measures. So, the estimates confound the influences of the omitted variables and health care resources. Which flawed measure is preferable for estimating the effect of health care resources on health is not clear.

Overall, this study, taken with the rest of the literature, supports the idea that more health care is productive. This general result contrasts with the view that much health care in the rich countries is wasteful; the “flat of the curve” medicine.⁶⁵

D. Other Statistical Issues

1. Treatment of Randomness (Error Modeling)

The modeling uses feasible generalized least squares (FGLS or GLS) to account for heteroskedasticity and for serial correlation of the errors. Further, as explained above, it includes country fixed effects. This is a reasonable approach to this data, although there are other possibilities.⁶⁶

A major problem is that the OECD approach implicitly counts all the variation at the country level as inefficiency in the health care system. This is a result of the interpretation given to the coefficient on the country-specific dummy variables and to the residual variation. In reality this coefficient also picks effects of three other types. First, as explained above, it picks up variation in excluded variables (e.g. lifestyle variables), including many for which there is no data available. Second, it picks up the effects of systematically mismeasured variables, such as using GDP PPP exchange rates, rather than real health care PPP exchange rates. Third, it picks up random variation. This is discussed in more detail below.

2. The Treatment of Time and Health Care Productivity

As is discussed above, the Report’s modeling does not include a variable for time. Thus, is it likely that some of the variables pick up the influence of time-related improvements in technology. Since resources devoted to health care have been increasing over time, the problem of confounding the influence of health care resources and the passage of time is particularly likely to be a problem with the measured productivity of health care. The result will be to overstate the productivity of health care. Victor Fuchs believes that this is a major problem, especially in time-series analyses such as the Report’s (2004). The effect should be smaller in cross-sectional analyses. In fact, it would vanish in cross-sectional analysis if technological diffusion were equal across countries.

VI. The DEA Approach

The report seems to favor the operations research technique of data envelopment analysis, DEA, in principle, but chooses panel data regressions for data reasons (Joumard,

⁶⁵ See, e.g., Fuchs (2004).

⁶⁶ See Greene (2004) for a discussion of the stochastic production function approach and Or, Wang and Jamison (2005) for an alternative allowing for the effect of health care to vary by country, rather than only the level of health to vary.

André, Nicq and Chatal 2008, p. 5). But, there are strong reasons to suggest that DEA is generally inferior, less stable and less reliable. It relies on simply assuming that the apparently most efficient observations (highest observed output, lowest observed input) are on the efficient frontier (curve) and all others are inefficient. Efficiency by country is measured as the distance from the frontier to the actual data point. This is method implicitly assumes that all unmeasured variation in health can be attributed to differences in health care system efficiency. This is the same implicit assumption that underlies the regression-based measures of health care system inefficiency. Recasting the analysis in a DEA framework does not make this assumption any more reasonable. The DEA approach is sensitive to measurement error, especially for observations at the extremes of the variables.

Further, the technique requires the use of a small number of inputs. Reportedly, results were not reasonable when several inputs were used (Joumard, André, Nicq and Chatal 2008, pp. 35, 36). This limitation exacerbates the problem of omitted variable bias. In the actual estimation, there were only three independent variables, health care resources, diet and a proxy for what the authors call economic, social and cultural status (ESCS) (Joumard, André, Nicq and Chatal 2008, pp. 36, 52). This last variable is taken from the OECD Programme for International Student Assessment (PISA). It is used here to stand in for both income and education to reduce the number of variables. The index is based on occupational status, parental education, family wealth, an index of home educational resources and an index related to culture in the home (OECD *Glossary*, undated).

VII. The Productive Efficiency of Different Health Care Systems

A. Estimates

The Report states that health care efficiency gains might increase LE by 3 years, on average (Joumard, André, Nicq and Chatal 2008, p. 6). This is a huge effect. The statement is derived from the estimates by summing up country fixed effects (the estimated coefficient on the dummy for the country) plus the residual error to estimate country-specific efficiency (Joumard, André, Nicq and Chatal 2008, p. 33).

The residual error, often just called the residual, is the difference between the actual value of LE and the value predicted by the model. It summarizes all the variation in the health measure that is not accounted for by the model (here, the model includes the country-specific fixed effects). As stated above, this amounts to attributing all unexplained variation plus all county effects to inefficiency. As the Report puts it:

The implicit assumption here is that all unexplained country-specific effects and residuals reflect inefficiency, and not measurement error, omitted variables and other factors (Joumard, André, Nicq and Chatal 2008, p. 33).

Indeed, this is the assumption. And the analysis presented above casts a great deal of doubt on this assumption. Raising some doubts itself, the Report observes in a footnote that:

Health and safety regulations, work and housing conditions and poverty could also play a role but the lack of data constrains the inclusion of these variables in the analysis (Joumard, Isabelle, Christophe André, Nicq and Chatal 2008, p. 33, fn. 30).

In the following text, though, the Report argues for the strong assumption that whatever is not measured by the model is correctly interpreted as health care system inefficiency, saying:

Supporting this assumption are the very low correlations, if any, between the unexplained differences in health status indicators and recent values of key variables which could not be included in the panel regressions—in particular income dispersion (as measured by Gini coefficients), obesity and population density (Joumard, André, Nicq and Chatal 2008, p. 33).

This is not a convincing test of whether the country inefficiency measures for any country, or for the U.S. in particular, are affected by the omission of these variables. The bias from omitted confounding variables results from several different correlations among the data: the correlations of the omitted variables with the included variable (here, the U.S. dummy and the residual) and the strength of the actual impact of the omitted variables. Further, the short list of imperfectly measured variables tried in this analysis hardly exhausts the categories of relevant lifestyle, cultural and economic variables. Related work with a much larger WHO panel has been criticized for exactly the same assumption: attributing all unmeasured country-level heterogeneity to inefficiency (Greene 2004, pp. 959, 960, 977).

Comanor, Frech and Miller (2006) examine the effect of excluding or including just one of these relationships—involving obesity—in a model with controls for several other variables. They find that the residual for the U.S. in a cross-section equation that is similar to the Report's, is highly dependent on whether obesity is included or not. For example, the residual for male LE at 60 drops from -0.62 years to -0.18 years when obesity is included; for females at 60, it drops from -1.46 to -1.01 years. Also, when obesity is included, none of the U.S. residuals are statistically significant at even the 10 percent level (Comanor, Frech and Miller 2006, p. 13). The controls are still far from perfect even when obesity is included. Obesity itself is not perfectly measured and other, unmeasured or imperfectly measured lifestyle, cultural, environmental and economic influences remain. Therefore, we would not interpret the residuals from our equations as good measures of health care system efficiency, not even in our models that include obesity.

In a classic study, showing the importance of cultural and lifestyle effects, Victor Fuchs compared age-specific mortality in Nevada to that in neighboring Utah. These states are similar in every observable way, from dry climate to health care systems. Yet, the excess mortality in Nevada was stunning. For adults 40-49, it was 54 percent for males and 69 percent for females. The cause is clearly the difference in lifestyle and culture. The Mormons in Utah live healthy, stable lives, with low use of alcohol, tobacco

and low divorce rates. Nevada is the opposite (Fuchs 1974, pp. 52, 53). It would be a large mistake to attribute the difference in health between the two states to health care system efficiency.

Adding the residual to the country-specific effect is somewhat odd. By construction, the country fixed effect picks up all the cross-sectional variation—all the fixed, unexplained differences associated with each country. The residual for that country can only differ from zero in any particular year because of time-varying unexplained effects. They must average zero for each country by construction. As a result, the estimated inefficiency varies over time. Only the estimates for 2003 are shown, but the authors of the Report state that the estimates do not vary much over time (Joumard 2008). That statement implies that most of the variation in inefficiency across countries is due to the coefficients of the dummy variables, not the residuals.

The inefficiency estimates from the econometric analysis are presented centered around zero. That is, centered around the efficiency of the average country, not the most efficient one. Therefore, some countries' health care systems efficiency score would be positive if they appeared to be more efficient than average and some would be negative if they appeared to be less efficient than average. Another, perhaps more common, way to present them is to set the most productive country as a benchmark and measure all deviations from them as apparent inefficiency (Greene 2004, p. 961; Frech and Mobley 2000, p. 372; Schmidt and Sickles 1984). Scaled as they are in the Report, it is more convenient to refer to the measures as efficiency measures, bearing in mind that they can take on a value that is either positive or negative. In any case, the estimates are very large indeed. For example, in using expenditures and looking at LE at birth, the U.S. country-specific efficiency measure is -4.0 years.⁶⁷ Other apparently low performers are Hungary at -3.1 and Denmark at -1.5 years. The U.S. comes out as the worst of all and quite a bit worse than the next rich country (Denmark). At the other extreme, for Iceland the score is 2.6 and for Australia, it is 2.5 years (Joumard, André, Nicq and Chatal 2008, pp. 33, 34). This means that, if the U.S. had as an efficient a health care system as Iceland's, U.S. life expectancies would be greater than they are by 6.6 years. If it was as efficient as Australia's, LE would be 6.5 years greater (Joumard, André, Nicq and Chatal 2008, p. 25).⁶⁸ These estimates seem implausibly large.

These estimates are larger than those in the literature. For example, in a comparable work for a single cross section, Comanor, Frech and Miller (2006) estimated the relative shortfall, for LE at birth, of the U.S. at -1.56 years for males and -0.53 for

⁶⁷ The inefficiency estimates are random variables, but the Report does not present the standard deviations or confidence intervals for the econometric estimates. However, I have obtained the standard deviations (of the country effects alone) from the authors (OECD, 2008). The standard deviations are less than 1.0 year for most countries. For the U.S., the standard deviation is 0.6 years when health care resources are measured by physical units and 0.7 years when measured by health care spending. The 95 percent confidence interval for the U.S. would be plus or minus 1.4 years. Sampling error does not seem to be overwhelming, at least not for the U.S. estimate that is featured in the Report. But note that sampling error will be understated if there was interpolation of any data.

⁶⁸ There are mathematical errors of approximation involved in converting log estimates to levels. This procedure is called exponentiation. The errors occur because the log function is not linear. There is a fix for this, called "smearing estimates," which was done in Comanor, Frech and Miler (2006, p. 15) in similar OECD data. The errors were found to be small.

females when obesity was controlled for (2006, p. 13).⁶⁹ This is far smaller effect (less in absolute value) than the -4.0 years from the Report. Even when they do not control for obesity, their estimated shortfall is -2.19 years for males and -1.56 for females (Comanor, Frech and Miller 2006, p. 13).

The results are not robust to using different health outcome and health care resources measures in the OECD Report itself. The OECD Report states that country efficiency rankings are “roughly similar” using LE at 65 versus at birth (Joumard, André, Nicq and Chatal 2008, p. 33, fn. 31).⁷⁰ But that’s apparently not so for the ranking of the U.S. As mentioned above, the U.S. inefficiency is -4.0 years for LE at birth (Joumard, André, Nicq and Chatal 2008, pp. 33, 34). But, for female LE at 65 (the only one presented) the U.S. estimated inefficiency using health spending is about -0.5 years.⁷¹ The U.S. rank is 17th of 23 and the U.S. now does better than the U.K. and Ireland (Joumard, André, Nicq and Chatal 2008, p. 69).

Further, the OECD results are not robust to using different measures of of health care resources. Looking at LE at birth, as mentioned, the estimated inefficiency for the U.S. in the featured specification is about -4.0 years (Joumard, André, Nicq and Chatal 2008, pp. 33, 34) and the U.S. underperforms the most. When health care resources are measured in physical terms, the U.S. inefficiency estimates drops to -2.5 years and it ranks above Hungary (Joumard, André, Nicq and Chatal 2008, p. 69). The difference between these two probably partly reflects the bias in using GDP PPP exchange rates for convert health spending to dollars, which is discussed above. Looking at female LE at 65, for the physical health resources measure, the U.S. does even better, with an inefficiency of about zero and a ranking at the median, 12th out of 23 (p. 69). See Table 9 for a display of the different apparent inefficiency estimates from the various regression versions in the Report and in Comanor, Frech and Miller. As one can see, the measure of relative inefficiency varies greatly, depending on how the inputs are measured, on whether obesity is controlled for and also on which measure of LE is used.

⁶⁹ This is based on the residual for the U.S. One would get an identical answer by inserting a dummy variable for the U.S. But, a full set of country dummies cannot be used in a cross section because that would lead to negative degrees of freedom—mathematically impossible to estimate.

⁷⁰ In this footnote (Joumard, André, Nicq and Chatal (2008), p. 33, fn. 31), the Spearman rank correlation between two different sets of rankings is noted. But, this is a poor measure of correlation that wastes information. The underlying data here (efficiency scores) are inherently and are interpreted as cardinal numbers, not mere rankings. Therefore, the more commonly used Pearson correlation (often just called the correlation) would be more meaningful.

⁷¹ Country-by-country inefficiency estimates at age 65 are only presented for females.

Table 9: Apparent U.S. Efficiency Differences: LE, Regression Models				
Source	Measure of Health Care Resources	Obesity Controlled For?	Measure of Health	Apparent Inefficiency (Relative to Mean)
JCMC, p. 25, Table 6	Spending at GDP PPP	No	LE at Birth	-4.0 years
JCMA, p. 34, Figure 9	Physicians and Nurses	No	LE at Birth	-2.5 years
JCMC, p. 56, Figure A3.2	Spending at GDP PPP	No	Female LE at 65	-0.5 years
JCMC, p. 56, Figure A3.2	Physicians and Nurses	No	Female LE at 65	0.0 years
CFM	Spending at Health Care PPP	No	Female LE at Birth	-1.56 years
CFM	Spending at Health Care PPP	Yes	Female LE at Birth	-0.53 years
CFM	Spending at Health Care PPP	No	Male LE at Birth	-2.19 years
CFM	Spending at Health Care PPP	Yes	Female LE at Birth	-1.56 years
CFM	Spending at Health Care PPP	No	Female LE at 60	-1.46 years
CFM	Spending at Health Care PPP	Yes	Female LE at 60	-1.00 years
CFM	Spending at Health Care PPP	No	Male LE at 60	-0.62 years
CFM	Spending at Health Care PPP	Yes	Male LE at 60	-0.18 years
Sources, Joumard, André, Nicq and Chatal (2008), pp. 25, 34, 56; Comanor, Frech and Miller (2006), p. 13.				

One can look to other estimates of country-specific apparent efficiency in the literature. These results are quite different from the ones stressed in the Report. Or, Wang and Jamison's estimates are not directly quantitatively comparable to the Report's. In effect, they use a dummy variable for each country to control for unexplained fixed effects, like the Report. But, they do not interpret the coefficient on this dummy variable as measuring the efficiency of the health care system. Rather, they allow for the effect of health care resources (here, the physician/population ratio) to vary across countries. They interpret differences in the coefficient on this variable as the efficiency difference across countries. Their efficiency measures are, therefore, differences in slopes across countries, while the OECD Report's efficiency differences are differences in the constant term across countries. Estimating efficiency by differences in slopes is conceptually superior to the Report's interpretation. It is less confounded by other influences on health. Still, the Or, Wang and Jamison approach is vulnerable to a weaker version of same the criticism. I.e. the slope of the production function can also differ across countries because of confounding influences (Garber and Skinner 2008, p. 30).

Because the basis of the Or, Wang and Jamison estimates is so different, their quantitative estimates in terms of years are not comparable to the Report's. On the other hand, the country productivity rankings can be compared. Table 10 shows the large variation in ranking for the U.S. in different models of LE that use different measure of the health care inputs, different statistical approaches and different measures of LE.

They are very different from those in the Report. In particular, the U.S. comes out generally much higher. In Or, Wang and Jamison's model for LE at birth for females, the U.S. ranks 12th out of 21 in health care efficiency, ranking higher than the U.K. Norway, and Sweden. In their rankings for males, the U.S. is above the mean and the median, ranking 5th out of 21. The rankings are not consistent across measures of health. In infant mortality, the U.S. is 9th of 21. Looking at LE at 65, the U.S. is 17th of 21 for females and 9th of 21 for males. In avoiding premature mortality from cardiovascular disease, the U.S. health care system is superior, ranked 7th of 21 for females and first, the top performer, for males (Or, Wang and Jamison 2005, pp 543-546).

Nixon and Ulmann (2006) estimate a model that is similar to the OECD one, on a panel of 9 European countries. Looking at male LE at birth, they find Greece, the U.K. and Germany as the most efficient. For female LE, the most efficient are Luxemburg, Greece and the U.K (Nixon 2008).⁷² Only the relatively good showing of Greece is common across this study and the OECD's featured result

The Report also estimates different country-specific effects for other health measures: PYLL and infant mortality. The PYLL results are not presented, but some results for a (heart-disease-specific) version of this are available from Or, Wang and Jamison (2005, p. 545). In these estimates, the U.S. comes out as very efficient. For females, the U.S. is the most efficient in the OECD. These results can be reconciled with the Report's easily. Or, Wang and Jamison measure health care resources by physicians only, but most importantly, as mentioned above, they do not attribute country-fixed effects to efficiency differences. Further, the Or, Wang and Jamison estimates for the heart disease are consistent with estimates from a detailed micro study that found U.S.

⁷² The rankings mentioned in the text are based on corrections to what was published, from a private communication from John Nixon, (Aug. 1, 2008).

health care productivity to be quite comparable, and often more productive than, health care productivity in Germany or the United Kingdom (Bailey and Garber, 1997).

In the Report's estimates for infant mortality, the U.S. comes out poorly, either the lowest or the second lowest (second to Turkey) (Joumard, André, Nicq and Chatal 2008, p. 69). This is the health measure that seems to be most sensitive to omitted variables, especially those reflecting cultural and lifestyle influences. But, the results for infant mortality are quite different in Or, Wang and Jamison, where the U.S. comes out in the middle of the pack (2005, p. 546). See Table 11. Here again, the differences can be easily reconciled. The most important difference is that Or, Wang and Jamison do not attribute estimated country differences to health care productivity, while the OECD Report does.

Table 10: Rankings by Apparent Efficiency, for LE							
JCMC LE at Birth Health Spending at GDP PPP	JCMC Female LE at Birth Physicians and Nurses	JCMC Female LE at 65 Health Spending at GDP PPP	JCMC Female LE at 65 Physicians and Nurses	OWJ Female LE at Birth Physicians	OWJ Male LE at Birth Physicians	OWJ Female LE at Birth Physicians	OWJ Male LE at Birth Physicians
Iceland	Greece	Australia	Australia	Japan	Canada	Japan	Japan
Australia	Australia	France	France	Canada	Japan	Austria	Austria
N. Zealand	Iceland	N. Zealand	Canada	Australia	Australia	France	France
Korea	France	Canada	N. Zealand	Austria	Austria	Australia	N. Zealand
Greece	N. Zealand	Iceland	Iceland	Portugal	U.S.	Canada	U.K.
Canada	Korea	Korea	Korea	France	N. Zealand	N. Zealand	Portugal
Finland	Switzerland	Switzerland	Switzerland	N. Zealand	U.K.	Belgium	France
Poland	Canada	Belgium	Greece	Germany	Portugal	Switzerland	Finland
Sweden	Sweden	Finland	Austria	Belgium	Finland	Portugal	U.S.
France	Netherlands	Sweden	Sweden	Greece	Germany	France	Belgium
Belgium	Germany	Austria	Finland	Switzerland	France	Germany	Germany
Ireland	Austria	Greece	U.S.	U.S.	Switzerland	Sweden	Greece
U.K.	Turkey	Poland	Germany	U.K.	Belgium	Greece	Switzerland
Czech Rep.	Iceland	Netherlands	Poland	Finland	Greece	Spain	Canada
Netherlands	Finland	Norway	Netherlands	Ireland	Sweden	Italy	Spain
Switzerland	U.K.	Germany	Norway	Spain	Italy	Norway	Sweden
Austria	Poland	Ireland	U.K.	Sweden	Netherlands	U.S.	Italy
Germany	Czech R.	U.S.	Ireland	Italy	Ireland	Netherlands	Ireland
Turkey	Denmark	U.K.	Denmark	Netherlands	Spain	Ireland	Netherlands
Norway	Norway	Denmark	Czech R.	Norway	Norway	U.K.	Denmark
Denmark	U.S.	Czech R.	Hungary	Denmark	Denmark	Denmark	Norway
Hungary	Hungary	Hungary	Turkey				
U.S.		Turkey					

Sources: Joumard, André, Nicq and Chatal (2008), pp. 25, 34, 56; Or, Wang and Jamison (2005), pp. 543, 544.

Table 11: Rankings by Apparent Efficiency: PYLL Heart Disease and Infant Mortality				
OWJ Female PYLL by Heart Disease Physicians	OWJ Male PYLL by Heart Disease Physicians	JCMC Infant Mortality Health Spending at GDP PPP	JCMC Infant Mortality Physicians and Nurses	OWJ Infant Mortality Physicians
Australia	U.S.	Czech R.	Korea	Canada
Japan	Australia	Ireland	Czech R.	Portugal
New Zealand	Canada	Finland	Greece	Austria
Finland	Finland	Korea	Iceland	Germany
Canada	Netherlands	Greece	Finland	Greece
Switzerland	New Zealand	Poland	Poland	U.K.
U.S.	Denmark	New Zealand	France	Australia
Sweden	Switzerland	Australia	New Zealand	France
Netherlands	Belgium	Hungary	Australia	U.S.
Denmark	U.K.	Sweden	Hungary	New Zealand
Belgium	Sweden	Belgium	Denmark	Japan
France	Japan	France	Sweden	Switzerland
Portugal	France	Denmark	Germany	Denmark
U.K.	Portugal	Ireland	Austria	Italy
Italy	Italy	Canada	U.K.	Spain
Spain	Germany	U.K.	Canada	Finland
Greece	Norway	Austria	Netherlands	Belgium
Germany	Austria	Germany	Ireland	Sweden
Austria	Spain	Netherlands	Norway	Norway
Ireland	Greece	Norway	Switzerland	Ireland
Norway	Ireland	Switzerland	Turkey	Netherlands
		U.S.	U.S.	
		Turkey		

Sources: Joumard, André, Nicq and Chatal (2008), p. 69; Or, Wang and Jamison (2005), pp. 545, 546.

B. DEA versus Panel Econometric Estimates of Apparent Country-Specific Efficiency

The Report states that “Panel data results and DEA efficiency scores are broadly consistent” (Joumard, André, Nicq and Chatal 2008, pp. 34, 35). They further state that, based on scatter plots (Joumard, André, Nicq and Chatal 2008, p. 38) that:

Figure 12 shows that the two techniques give a broadly consistent picture of the gains, as measured by the number of years of life that could be saved, if efficiency were to be raised to the level implied by the estimated efficiency frontier. (Joumard, André, Nicq and Chatal 2008, p. 21, fn. 32).

This seems like a very favorable reading. The comparison is limited to a single health variable, LE at birth. As is discussed above, this is probably not the best measure for health care efficiency. Further, it matters to the estimates. Estimated efficiency by country differs greatly according to which health measure is used. Further, “broad consistency” is in the eye of the beholder. My visual interpretation of the scatter plots comparing the two methods is that they do not line up very well country by country, though the averages are similar.⁷³ Note that the time periods are different (1981-2003 versus 2004 only) and that the list of countries is also different. Lastly, the methods have similar problems of confounding the effect of omitted variables, especially lifestyle and cultural variables, with country-specific efficiency.

C. Apparent Efficiency, Policy and Institutions

The Report attempts, unsuccessfully, to relate its estimated efficiency measures to various aspects of the institutional organization of the health care system. Variables examined include the percentage of the population covered by insurance, the relative importance of out-of-pocket payments or the general model (e.g. private v. public insurance) (Joumard, André, Nicq and Chatal 2008 p. 23). This is in line with the results in the literature (Glied 2008; Or, Wang and Jamison 2006, pp. 553-554). It is always difficult to make inferences from a negative result (a lack of a statistically significant relationship). But, the negative result on these issues is consistent with the idea that country-specific health outcomes are determined largely by external factors that are not closely related to the health care system, nor to health policy and institutions.

It is interesting to note that, in spite of its reputation for a more private and market-oriented health care system, U.S. relies substantially *less* on out-of-pocket payments than the OECD average, 13.3 percent v. 19.3 percent, ranking 23th of 28 (Joumard, André, Nicq and Chatal 2008, p. 40). Also, proportion of the population covered by some private insurance is greater in France, Switzerland and the Netherlands than it is in the U.S. (Glied 2008, p. 5). As discussed above, the efficiency measures being explained differ substantially.

D. Sensitivity Analysis in the Report

⁷³ The Report presents only a scatter plot, not the correlation estimates.

The authors of the Report have done some sensitivity analyses to test the robustness of their results. This is an excellent idea, but the Report does not push it far enough. The Report presents different versions of the specification, alternatively excluding subsets of the explanatory variables. First, the authors exclude GDP, then GDP and education and finally they reinstate GDP and education and remove the lifestyle variables. In summarizing the results of these truncated models, they say that:

A close look at the results from these alternative scenarios suggests that most estimated coefficients are broadly stable in level and significance (Tables A.3.1. to A.3.8). There are some deviations, however. In the “lifestyle” scenario, estimated coefficients are generally higher but their significance is sometimes reduced. The estimated spending elasticity is also somewhat higher in models without GDP, reflecting the correlation between these two variables. When health care resources are measured by the number of practitioners, estimations are less stable across alternative scenarios (Joumard, André, Nicq and Chatal 2008, p. 62).

The results for the country-specific inefficiency measures are not discussed. It would have been very informative to see them. The most important remaining issue is the productivity of health care. This is not so stable across specifications. When health care resources are measured by spending, the elasticity for female LE at birth increases from 0.035 to 0.064 when GDP and education are removed. For males, the elasticity increases from 0.045 to 0.083. These are increases of 83 and 84 percent. When health care resources are measured by physical units, the exclusion of GDP and education makes more difference. The elasticity for females increases from 0.013 to 0.083. For males, the elasticity increases from 0.017 to 0.105. These are increases of 439 and 418 percent. Further, the values from the truncated model seem implausibly high. Since the estimates are not nearly so sensitive to excluding GDP, most of this striking change in the estimated productivity of health care probably comes from excluding education. Most likely, the education variable is picking up lifestyle and cultural variation, so excluding it exacerbates the omitted variable bias problem.

The thrust of the sensitivity analysis goes the wrong direction. The main problem with the analysis is not that too many variables, redundantly covering the same forces, have been included. The main problem is that important variables, especially lifestyle and cultural variables have been excluded. The model is already too truncated. This problem could be partially explored by augmenting the model with more relevant lifestyle variables, at the cost of fewer observations, but it cannot be explored by dropping variables from a model that is already incomplete.

VIII. Suggested Improvements

Possible improvements to the study flow from the analysis above. Many of the improvements involve data that are not available for some countries and some years. This might limit the size and the variation in the sample. There is often a tradeoff between using the conceptually superior variables in a smaller sample and using conceptually inferior variables in a larger sample. Here, the harm from not using the conceptually superior variables seems large. One could keep the larger sample with the poorer variables for comparison purposes. There is no definitive model. Analysts will differ on the best choice among imperfect alternatives.

More specifically, to reduce the problem of confounding omitted variables, the specification could be augmented. Variables for obesity, and for the prevalence of accidents and violence would be especially useful. Both of these are available at least for some countries and some years. If they were available, variables for disease prevalence might be useful.

The second issue is measurement error and choice of the best measure or measures for any particular concept. Measurement issues interact with omitted variable issues because some health measures are more influenced by factors external to the health care system than others. The worst health variable for confounding non-health care factors seems to be infant mortality. However, many of those external factors are reflected in birthweight. Therefore, defining the health measure as birthweight-specific infant mortality would appear to be an improvement. LE at birth and PYLL are seriously contaminated with infant mortality. Restricting attention to LE at later years would help. One could also consider a version of the PYLL that excluded most of the causes of death that affected infants. Since morbidity is so important, it would be helpful a measure of quality-adjusted or disability-adjusted LE.

Turning to measurement issues on the input side, it would improve matters to replace the GDP PPP exchange rate with the health PPP exchange rate. Since the health PPP exchange rate has its own flaws, it would be good to continue to use a physical measure in parallel. The physical measure could be improved by including all or most types of health care labor and by using an objective market-based weighting, such as using OECD average relative wages. For pollution, replacing NO_x with a more relevant chemical measure than NO_x, if possible, would be an improvement. Whatever chemical measure was used, converting emissions to exposure, as discussed above, would be a natural improvement.

In terms of technique, lags of all the input variables five to 10 years would fit the underlying economic processes better. Further, explicit control for time, either with a full set of time fixed effects or with a time trend variable, makes sense to control for time-related influences, such as technological progress.

But, the most important change is not technical or statistical. The most important change is a matter of assumption and interpretation. It makes sense to interpret country-specific effects broadly as reflecting a mixture of country-specific long-lasting differences in culture, lifestyle, industrial structure and so

on, as well as health care system efficiency. In this context, good statistical reporting would suggest reporting these measures and their associated standard deviations for all or most of the different versions of the model. Further, interpretation would be clearer if the estimates of the country-specific fixed effects were separately reported from the residuals.

IX. Conclusion

The OECD Report is an important and useful effort. It advances our understanding of the usefulness of health care in the household production of health in the OECD countries. The Report concludes that health care is highly productive in improving health outcomes and that productive efficiency varies greatly across countries. Going much further, it provides country-specific estimates of that efficiency. Unfortunately, there are major problems in the analysis that render those conclusions, especially the country-specific conclusions, unreliable.

Many factors that influence the production of health are either omitted or poorly measured. This confounds the true productivity of the input with other factors. The resulting coefficients include omitted variable bias. Because of inherent data limitations, this problem can only be minimized and not eliminated completely.

Even among included variables, there is the problem of measurement error. Systematic measurement error in either inputs or outputs causes bias. Even unsystematic measurement error of an input causes problems (a bias towards zero). Further, measuring one input badly implies that one has not actually held it constant. The consequences are similar to the situation where that variable is omitted altogether from the analysis. This creates an issue of omitted variable bias in the coefficients of the other included variables.

Estimation of health care productivity and the efficiency of different countries' health care systems requires adequate statistical controls for other determinants of health, to avoid confounding health care resource use or the efficiency of a country's system with other factors that operate in the society or the economy that are largely outside the health care system. Examples include healthy lifestyles, favorable cultures, high income, low pollution, good genes, favorable industrial and urban structure, and good education. As we have seen, this study has problems with both omitted variables and systematic and unsystematic measurement errors. The net effect is to bias upward the estimated apparent inefficiency of the U.S. health care system and probably to bias upward the estimated productivity of health care.

Even if the study had been done perfectly, it would have been overreaching to interpret country-specific variation in health outcomes as a measure of health care system productivity. The country-specific estimates, in reality, reflect all slow-moving differences in country-level influences, whatever their source. As William Greene stated in a similar context, "there is considerable heterogeneity that has masqueraded as inefficiency (2004, p. 959)."

References

- Adang, Eddy M. and George F. Borm, “Is there an association between economic performance and public satisfaction in health care?” *European Journal of Health Economics* 8 (2007): 279–285
- Anderson, Gerard F., Uwe E. Reinhardt, Peter S. Hussey and Varduhi Petrosyan, “It’s the Prices, Stupid: Why the United States is so Different from Other Countries,” *Health Affairs* 22 (3) (May-June 2003): 89-105.
- Anderson, Gerard F., Bianca K. Frogner and Uwe E. Reinhardt, “Health Spending in OECD Countries In 2004: An Update,” *Health Affairs* 26 (5) (Sept,-Oct. 2007): 1481-1489.
- Banks, James, Michael Marmot, Zoe Oldfield and James P. Smith, “The SES Health Gradient on Both Sides of the Atlantic,” NBER Working Paper No. 12674 (November 2006).
- Baily, Martin Neil and Alan M. Garber, “Health Care Productivity,” *Brookings Papers on Economic Activity: Microeconomics* (1997): 143-215.
- Baum, Charles L., II, and Christopher J. Ruhm, “Age, Socioeconomic Status and Obesity Growth,” NBER Working Paper No. 13289 (August 2007).
- Becker, Gary S., *Economic Theory*, New York: Knopf (1971).
- _____, Tomas J. Philipson and Rodrigo R. Soares, “The Quantity and Quality of Life and the Evolution of World Inequality,” *American Economic Review* 95 (1) (Mar. 2005): 277-291.
- Board of Governors of the Federal Reserve System, 2008
http://www.federalreserve.gov/releases/h10/hist/dat00_eu.htm, (Aug. 1, 2008).
- Braveman, P. and S. Gruskin, “Defining equity in health,” *Journal of Epidemiology and Community Health* 57 (2003): 254-258.
- Bureau of Labor Statistics, U.S. Department of Labor, *Career Guide to Industries, 2008-09 Edition*, Health Care on the Internet at <http://bls.gov/oco/cg/cgso35-htm> (Aug. 28, 2008).
- Busse, Reinhard, Jonas Schreyogg, and Peter C. Smith, “Variability in Healthcare Treatment Costs Amongst Nine EU Countries—Results from the Healthbasket Project,” *Health Economics* 17 (Supplement 1) (Jan. 2008): S1-S8.
- Calfee, John E., *Prices, Markets, and the Pharmaceutical Revolution*, Washington, D.C.: The AEI Press (2000)

Castles, Ian, "Review of the OECD-Eurostat PPP Program," OECD (1997) <http://www.oecd.org/dataoecd/33/15/2405262.pdf>, accessed July 14, 2008. (Also sometimes cited as *Review of the OECD-Eurostat PPP Programme*, I. Castles, unpublished OECD document, STD/PPP(97)5, (September 1997.))

_____, and David Henderson, "International Comparisons of GDP: Issues of Theory and Practice," *World Economics* 6 (1) (Jan-Mar. 2005): 1-32.

Center for Disease Control, U.S., WISQARS Injury Mortality, WISQARS Injury Mortality Reports, 1999 – 2005.

Chay Kenneth Y and Michael Greenstone, "The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession," *Quarterly Journal of Economics* 118 (3) (August 2003): 1121–1167.

Chiu, Yu-Chan, Katherine Clegg Smith, Laura Morlock and Lawrence Wissow, "Gifts, bribes and solicitations: Print media and the social construction of informal payments to doctors in Taiwan," *Social Science and Medicine* 64 (2007): 521-530.

Clark, Damon and Heather Royer, "The Effect of Education on Adult Mortality and Health: Evidence from the United Kingdom," Case-Western Reserve University, unpublished paper," (May 7, 2008).

Comanor, William S., H.E. Frech III and Richard D. Miller, Jr. "Is the United States an outlier in health care and health outcomes? A preliminary analysis," *International Journal of Health Care Finance and Economics* 6 (1) (Mar. 2006): 3-23.

Cutler, David M. *Your Money or Your Life: Strong Medicine for America's Health Care System*, Oxford University Press (2004).

_____, and Mark McClellan, "Is Technological Change In Medicine Worth It?" *Health Affairs*, 20 (5) (Sept./Oct. 2001): 11-29.

Cutler, David M., Edward L. Glaeser and Jess M. Shapiro, "Why Have Americans Become More Obese?" *Journal of Economic Perspectives* 17 (3) (Summer 2003): 93-118.

Cutler, David M. and Edward Glaeser, "What Explains Differences in Smoking, Drinking and Other Health-Related Behaviors?" *American Economic Review* 95 (2) (May 2005): 238-242.

Cutler, David M., Angus Deaton and Adriana Lleras-Muney, "The Determinants of Mortality," *Journal of Economic Perspectives* 20 (3) (Summer 2006): 97-120.

Danzon, Patricia M., "Price Discrimination for Pharmaceuticals: Welfare Effects in the U.S. and the E.U." *International Journal of the Economics of Business* 4 (3) (Nov. 1997a): 301-322.

_____, *Pharmaceutical Price Regulation: National Policies versus Global Interests*. Washington, D.C.: AEI Press (1977b).

_____, and Allison Percy, "The Effects of Price Regulation on Productivity in Pharmaceuticals." unpublished paper, The Wharton School, University of Pennsylvania, (July 1995).

Danzon, Patricia M., and Michael Furukawa, "International Prices and Availability of Pharmaceuticals in 2005," *Health Affairs* 27 (1) (Jan.-Feb., 2008): 221-233.

Danzon, Patricia M., and Adrian Towse, "Differential Pricing for Pharmaceuticals: Reconciling Access, R&D and Patents," *International Journal of Health Care Finance & Economics* 3 (2) (Sept. 2003):183-206.

David, Richard and James Collins, Jr., "Disparities in Infant Mortality: What's Genetics Got to Do With It?" *American Journal of Public Health* 97 (7) (July 2007): 1192-1197.

Deacon, Robert T. and Jon Sonstelie, "Rationing by Waiting and the Value of Time: Results from a Natural Experiment," *Journal of Political Economy* 93 (4) (Aug. 1985): 627-647.

Dranove, David, *What's Your Life Worth? Health Care Rationing... Who Live? Who Dies? And Who Decides?* Upper Saddle River, N.J.: FT Prentice-Hall, (2003).

Economist, "The price of being well," 388 (8595) (Aug. 20, 2008): 59-60.

Ehrlich, Isaac, "Uncertain Lifetime, Life Protection, and the Value of Life Saving," *Journal of Health Economics* 19 (3) (May 2000): 341-367.

Finkelstein, Eric A., Ian C. Fiebelkorn and Guijing Wang. "National Medical Spending Attributable To Overweight And Obesity: How Much, And Who's Paying?" *Health Affairs* (May 14, 2003): w3-219—w3-226 (web exclusive) 10.1377/hlthaff.w3.219.

Folland, Sherman, Allan C. Goodman and Miron Stano, *The Economics of Health and Health Care*, Upper Saddle River, NJ: Prentice-Hall (2001).

Frech, H.E. III, "Physician Fees and Price Controls" in *American Health Care: Government, Market Processes and the Public Interest*, ed. by Roger Feldman, New Brunswick: Transaction Publishers (2000): 346-364.

_____ and Paul B. Ginsburg, *Public Insurance in Private Medical Markets: Some Problems of National Health Insurance*, Washington, D.C.: American Enterprise Institute for Public Policy Research, (1978).

Frech, H.E. III, and William C. Lee, "The Welfare Costs of Rationing-by-Queuing across Markets: Theory and Estimates from the U.S. Gasoline Crises," *Quarterly Journal of Economics* 102 (1) (Feb. 1987): 97-108.

Frech, H.E. III and Richard D. Miller, Jr., *The Productivity of Health Care and Pharmaceuticals: An International Comparison*, Washington, D.C.: AEI Press, (1999).

____ and _____, "The Effects of Pharmaceutical Consumption and Obesity on the Quality of Life in the Organization of Economic Cooperation and Development (OECD) Countries," *Pharmaceoeconomics* 22 (Suppt. 2) (2004): 25-36.

Frech, H.E. III and Lee R. Mobley, "Efficiency, Growth, and Concentration: An Empirical Analysis of Hospital Markets," *Economic Inquiry* 38 (3) (July 2000): 369-384.

Fuchs, Victor R., *Who Shall Live? Health, Economics and Social Choice*, New York: Basic Books (1974).

____, "Time Preference and Health: An Exploratory Study," in *Economic Aspects of Health*, edited by Victor R. Fuchs, National Bureau of Economic Research, Univ. of Chicago Press (1980).

____. "More Variation in Use of Care, More Flat-Of-The-Curve Medicine," *Health Affairs*, Web Exclusive (Oct. 7, 2004): VAR-104 to VAR-107.

John G. Frohna, John G., Paula M. Lantz and Harold Pollack, "Maternal Substance Abuse and Infant Health: Policy Options across the Life Course," *The Milbank Quarterly* 77 (4) (1999): 531-570.

Fujisawa, Rie and Gaetan Lafortune, "The Renumeration of General Practitioners and Specialists in 14 OECD Countries: What are the Factors Influencing Variations Across Countries?" OECD Health Working Papers, No. 41 (Dec. 18, 2008), DELSA/HEA/WD/HWP(2008)5.

Gage, Timothy B., and Kathleen O'Connor, "Nutrition and the Variation in Level and Age Patterns of Mortality," *Human Biology* February 66 (1) (Feb.,1994): 77-103.

Garattini, Livio, Franco Salvioni, Diego Scopelliti, and Silvio Garattini, "A Comparative Analysis of the Pharmaceutical Market in Four European Countries," *PharmacoEconomics* 6 (5) (1994): 417-423.

Garber, Alan M. and Jonathan Skinner, "Is American Health Care Uniquely Inefficient?" *Journal of Economic Perspectives* 22 (4) (Fall 2008):27-50.

Gaziano, J. Michale, Julie E. Buring, Jan L. Breslow, Samuel Z. Goldhaber, Bernard Rosner, Martin VanDenburgh, Walter Willett, and Charles H. Hennekens, "Moderate Alcohol Intake, Increased Levels of High-Density Lipoprotein and Its Subfractions, and Decreased Risk of Myocardial Infarction," *New England Journal of Medicine* 329 (25) (Dec. 16, 1993): 1829-1834.

Gibson, Eric, Jennifer Culhane, Terri Saunders, Donald Webb and Joy Greenspan, "Effect of Nonviable Infants on the Infant Mortality Rate in Philadelphia, 1992," *American Journal of Public Health*, 90 (8) (August 2000): 1303-1306.

Glied, Sherry A., "Health Care Financing, Efficiency and Equity," Working Paper No. 13881, National Bureau of Economic Research, (Mar. 2008).

Goldman, Dana P and James P. Smith, "Can Patient Self-Management Help Explain the SES Health Gradient?" *Proceedings of the National Academy of Science* 99 (16) (August 6, 2002): 10929-10934.

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

Greenstone, Michael, private communication (July 23, 2008).

Grossman, Michael, "On the Concept of Health Capital and the Demand for Health," *Journal of Political Economy* 80 (2) (April 1972): 223-255.

Hamilton, James D., *Time Series Analysis*, Princeton University Press, (1994).

Herndon, Jill Boylston, "Health insurer monopsony power: the all-of-none model," *Journal of Health Economics* 21 (2) (Mar. 2002): 197-206.

Hertz, Erica A., James R. Hebert and Joan Landon, "Social and Environmental Factors and Life Expectancy, Infant Mortality, and Maternal Mortality Rates: Results of a Cross-National Study," *Social. Science and Medicine* 39 (1) (1994): 105-114

Huber, Manfred and Eva Orosz, "Health Expenditure Trends in OECD Countries, 1990-2001," *Health Care Financing Review* 25 (1) (Fall, 2003): 1-22.

Ikegami, Naoki, "Japanese Health Care: Low Cost Through Regulated Fees," *Health Affairs* 10 (3) (Fall 1991): 87-109.

Ionescu, Carmiola, "Romanians pays \$1 million extra a day for health," *Lancet* 366 (Oct. 29, 2005): 1515.

Jamison, Eliot A., Dean T. Jamison and Eric A. Hanushek, "The effects of education quality on income growth and mortality decline," *Economics of Education Review* 26 (6) (Dec. 2007): 772–789.

Joumard, Isabelle (OECD), personal communication, (July 11, 2008).

_____, Christophe André, Chantal Nicq and Olivier Chatal, "Health Status Determinants: Lifestyle, Environment, Health Care Resources and Efficiency," OECE Economics Department Working Papers No. 627, (Aug. 4, 2008), ECK/WKP (2008)35.

Kimbo, Rachel Tolbert, Sharon Bzostek, Noreen Goldman, and Germán Rodríguez, "Race, Ethnicity, And The Education Gradient In Health," *Health Affairs* 27 (2) (March-April 2008): 361-372

Koch, Thomas G., "Public Insurance, Crowd-out and Health Care Utilization," unpublished paper, Economics Department, University of California, Santa Barbara (October 2008).

Liu, Ben-chieh, "The Costs of Air Quality Deterioration and Benefits of Air Pollution Control: Estimates of Mortality Costs for Two Pollutants in 40 U.S. Metropolitan Areas," *American Journal of Economics and Sociology* 38 (2) (Apr. 1979): 187-195.

Liu, Korbin and Marilyn Moon, "International Infant Mortality Numbers: A Look Behind the Numbers," *Health Care Financing Review* 13 (4) (Summer 1992): 105-122.

Lairopoulos, Lycourgoes, Olga Siskow, Daphne Kaitelidou, Mamas Theodorou, and Theofanis Katostaras, "Informal payments in public hospitals in Greece," *Health Policy* 87 (2008): 72-81.

Miller, Richard D., Jr. and H.E. Frech III, *Health Care Matters: Pharmaceuticals, Obesity and the Quality of Life*, Washington, D.C.: AEI Press (2004).

Michael, Ynonne, L., Graham A. Colditz, Eugenie Coakley and Ichiro Kawachi, "Health Behaviors, Social Networks, and Healthy Aging: Cross-Sectional Evidence from the Nurses' Health Study," *Quality of Life Research* 8 (8) (Dec. 1999): 711-722.

Michaud, Pierre-Carl, Arthur van Soest and Tatiana Andreyeva, "Cross-Country Variation in Obesity Patterns among Older Americans and Europeans," Rand Corp. Working Paper No. WR-495, (May 2007).

Murphy, Kevin M. and Robert H. Topel, eds., *Measuring the Gains from Medical Research: An Economic Approach*, Chicago: University of Chicago Press (2003).

Nixon, John, personal communication, Aug. 1, 2008.

_____, and Philippe Ulmann, “The relationship between health care expenditure and health outcomes: Evidence and caveats for a causal link,” *European Journal of Health Economics* 7 (2006): 7–18.

OECD, *OECD Health Data 2000: A Comparative Analysis of 29 Countries*. Paris: OECD (2000).

_____, *OECD Health Data 2007*, (2007a)

_____, *OECD in Figures 2007*, OECD Observer, Supplement 1 (2007b).

_____, OECD Dataset for Joumard, Isabelle, Christophe André, Chantal Nicq and Olivier Chatal, “Health Status Determinants: Lifestyle, Environment, Health Care Resources and Efficiency,” OECE Economics Department Working Papers No. 627, (Aug. 4, 2008), ECK/WKP (2008)35, personal communication from Isabelle Joumard, (July 11, 2008).

_____, *Glossary of Statistical Terms*, (undated)

<http://stats.oecd.org/glossary/detail.asp?ID=5401>, accessed July 28, 2008.

O’Neil, David M. and June E. O’Neil, “Health Status, Health Care and Inequality: Canada Vs. the U.S.,” *Forum for Health Economics & Policy* 10 (1) (Article 3) (2008).

Ohsfeldt, Robert L. and John E. Schneider, *The Business of Health: The Role of Competition, Markets and Competition*, Washington, D.C.: AEI Press (2006)

Or, Zeynep, “Determinants of health outcomes in industrialized countries: A pooled, cross-country, time-series analysis,” *OECD Economic Studies* 2000/1 (2000a) (30): 53–77.

_____, “Exploring the effects of health care on mortality across OECD countries.” OECD labor market and social policy occasional papers, no. 46. Paris: OECD, (2000b).

_____, Jia Wang and Dean Jamison, “International differences in the impact of doctors on health: a multilevel analysis of OECD countries,” *Journal of Health Economics* 24 (2005): 531-560.

Pauly, Mark. V., “Market Power, Monopsony, and Health Insurance Markets,” *Journal of Health Economics* 7 (2) (June 1988): 111-128.

_____, “U.S. Health Care Costs: The Untold True Story,” *Health Affairs* 12 (3) (Fall 1993): 152-159.

_____, *Markets Without Magic: How Competition Might Save Medicare*, Washington, D.C.: AEI Press, 2008.

Phelps, Charles E., *Health Economics*, 3rd Ed., Boston: Addison-Wesley (2003).

- Pindyck, Robert S. and Daniel L. Rubinfeld, *Econometric Models and Economic Forecasts*, 4th ed., Boston: Irwin-McGraw-Hill, 1998.
- Pope, C. Arden, III, Majid Ezzati and Douglas W. Dockery, "Fine-Particulate Air Pollution and Life Expectancy in the United States," *NewEngland Journal of Medicine*, 360 (4) (January 22, 2009): 376-386.
- Pritchett, Lant and Lawrence H. Summers, "Wealthier is Healthier," *The Journal of Human Resources* 31 (4) (Autumn, 1996): 841-868.
- Ramseyer, J. Mark, "The Mortality Effects of Cost Containment Under Universal Health Insurance: The Japanese Experience," Discussion Paper No. 619, Harvard Law School (Sept. 2008).
- _____, "The Effect of Cost Suppression under Universal Health Insurance on the Allocation of Talent and the Development of Expertise: Cosmetic Surgery in Japan," *Journal of Law and Economics* (forthcoming, 2009).
- Reinhardt, Uwe E., Peter S. Hussey and Gerard F. Anderson, "Cross-National Comparisons of Health Systems Using OECD Data, 1999," *Health Affairs*. 21 (3) (May/June 2002): 169-181.
- Robb, C. A., S.J. Huston and M. S. Finke, "The mitigating influence of time preference on the relation between smoking and BMI scores," *International Journal of Obesity* 32 (2) (Aug. 2008). 1-8/
- Roehrig, Charles, George Miller and Craig Lake, , "Estimates of NHEA Personal Health Expenditures by Medical Condition," Altarum Institute, Presented to the Society of Government Economists (June 2, 2008).
- Scherer, F. M., "Pricing, Profits, and Technological Progress in the Pharmaceutical Industry," *Journal of Economic Perspectives* 7 (3) (Summer, 1993): 97-115.
- Schmidt, Peter and Robin C. Sickles, "Production Frontiers and Panel Data," *Journal of Business & Economic Statistics* 2 (4) (Oct., 1984), pp. 367-374
- Schreyogg, Jonas M. Oliver Tiemann, Tom Stargardt and Reinhard Busse, "Cross-Country Comparisons of Costs: The Use of Episode-Specific Transitive Purchasing Power Parities with Standardized Cost Categories," *Health Economics* 17 (Supplement 1) (Jan. 2008): S95-S103.
- Schut, Frederick T. and Peter A. G. Van Bergeijk, "International Price Discrimination: The Pharmaceutical Industry," *World Development* 14 (9) (1986): 1141-1150.

Shaw, James W., William C. Horrace and Ronald J. Vogel, "The Determinants of Life Expectancy: An Analysis of OECD Health Data," *Southern Journal of Economics* 71 (4) (2005): 768-783.

Shields, Margot, Sarah Connor Gorber and Mark S. Tremblay, "Estimates of obesity based on self-report versus direct measures," Statistics Canada, Catalogue No. 82-003-XPE, *Health Reports* 19 (2) (June 2008a): 61-76.

_____, "Effects of measurement on obesity and morbidity," Statistics Canada, Catalogue No. 82-003-XPE, *Health Reports* 19 (2) (June 2008b): 77-84.

Stevenson, David A. and John C. Carey, "Contribution of Malformations and Genetic Disorders to Mortality in a Children's Hospital," *American Journal of Medical Genetics* 126A (4) (May 1, 2004): 393-397.

Szuba, Tadeusz J., "International Comparison of Drug Consumption: Impact of Prices," *Social Science and Medicine* 22 (10) (1986): 1019-1025.

Thorpe, Kenneth E., David H. Howard, and Katya Galactionova, "Differences in Disease Prevalence As A Source Of The U.S.—European Health Care Spending Gap," *Health Affairs* 26 (6) (2007): w678-w686 (published online Oct. 2, 2007; 10.3377/hlthaff.26.6.w678).

Turner, Catherina A. "A Well Pay Incentive Program," *Public Productivity Review* 6 (1/2), (Mar. - Jun., 1982): 127-130.

Walker, Mary Beth, Endal Tekin and Sally Wallace, "Teen Smoking and Birth Outcomes," *Southern Economic Journal* 75 (3) (Jan. 2009): 892-907.

Windham, Gayle C., Laura Fenster, Barbara Hopkins and Shanna H. Swan, "The Association of Maternal and Paternal Alcohol Consumption with Birthweight and Gestational Age," *Epidemiology* 6 (6) (Nov. 1995): 591-597.

Winkler, Donald R., "The Effects of Sick-Leave Policy on Teacher Absenteeism," *Industrial and Labor Relations Review* 33 (2) (Jan., 1980): 232-240.

Van Doorslaer, Eddy, Christina Masseria and Xander Koolman, "Inequalities in access to medical care by income in developed countries," *Canadian Medical Association Journal* 174 (2) (Jan. 17, 2006): 177-183.

Vujcic, Marko, Pascal Zum, Khassoum Daillo, Orvill Adams and Mario R. Dal Poz, "The role of wages in the migration of health care professionals from developing nations," *Human Resources for Health* 2 (3) (April 28, 2004): 1-14 (web version).

World Health Organization, *Closing the Gap in a Generation: Health Equity through Action on the Social Determinants of Health*, Final Report of the Commission on the Social Determinants of Health, Geneva: WHO, 2008.

Zhang, Lei and Inas Rashad, "Obesity and Time Preference: The Health Consequences of Discounting the Future," *Journal of Biosocial Science* 40 (1) (2008): 97-113.

Zweifel, Peter and Fredrich Breyer, *Health Economics* Oxford University Press (1997).

Zweifel, Peter and Matteo Ferrari, "Is There a Sisyphus Syndrome in Health Care?" in *Health Economics Worldwide*, ed. by Peter Zweifel and H.E. Frech III, Amsterdam: Kluwer (1992) pp. 311-330.

Zweifel, Peter, Lukas Steinmann and Patrick Eguster, "The Sisyphus Syndrome in Health Revisited," *International Journal of Health Care Finance and Economics* 5 (2) (June 2005): 127-145.