WHO'S ASSESSMENT OF HEALTH CARE INDUSTRY PERFORMANCE: RATING THE RANKINGS

ELLIOTT PARKER and JEANNE WENDEL *

Department of Economics, University of Nevada, Reno, NV, USA

SUMMARY

This paper examines the econometric methodology used by World Health Organization (WHO) researchers to develop a template for ranking health system efficiency in 191 countries, as reported in the *World Health Report 2000 (WHR)*. Complementing recent critiques on the quality of the WHR data, we ask whether the methodology would be appropriate for accomplishing the WHR goal, even if the data problems were resolved. We identify three econometric problems and then re-estimate the health production function and the resulting efficiency ratios, using available WHO data. We conclude that the WHR strategy of ranking countries according to econometric efficiency ratios will not produce policy-relevant results until health inputs, outputs, and production relationships are clarified.

KEY WORDS – World Health Organization; WHO; World Health Report 2000; production efficiency; comparative health care system efficiency

* Correspondence to: Jeanne Wendel, Department of Economics /030, University of Nevada, Reno, NV 89557-0207, USA. Tel.: +1 775-784-6695; e-mail: wendel@unr.edu

WHO'S ASSESSMENT OF HEALTH CARE INDUSTRY PERFORMANCE: RATING THE RANKINGS

SUMMARY

This paper examines the econometric methodology used by World Health Organization (WHO) researchers to develop a template for ranking health system efficiency in 191 countries, as reported in the *World Health Report 2000 (WHR)*. Complementing recent critiques on the quality of the WHR data, we ask whether the methodology would be appropriate for accomplishing the WHR goal, even if the data problems were resolved. We identify three econometric problems and then re-estimate the health production function and the resulting efficiency ratios, using available WHO data. We conclude that the WHR strategy of ranking countries according to econometric efficiency ratios will not produce policy-relevant results until health inputs, outputs, and production relationships are clarified.

INTRODUCTION

The World Health Report (WHR) 2000 [1], entitled "Health systems: Improving performance", presents the results of an ambitious body of work undertaken in pursuit of a laudable goal. In an effort to "stimulate action that will eventually improve the overall performance of health systems" [2, p 17], researchers at the World Health Organization define health inputs and outputs, build a five-year data set for 141 of their 191 member countries, estimate a health production function, and use these estimates to calculate rankings of overall efficiency for each country. Anticipating the potential policy implications of such rankings, we ask two questions:

- How reliable and useful are these rankings?
- Does this report provide a template for analyzing international health system efficiency?

In this paper, we focus on the econometric methodology used to create the efficiency rankings. Blendon, et al. [3] and Williams [4] have critiqued the WHR data set. Williams focused on the large proportion of the data set that was estimated rather than observed, while

Blendon, et al., focused on the WHR reliance on expert opinions to proxy public views, without input from public opinion surveys.

We ask a different question: if the data quality issues were resolved, would the WHR econometric approach provide a solid framework for understanding, measuring, and analyzing health system efficiency? We identify three problems in the specification of the WHR production function, and we use their data to re-estimate health care efficiency and assess the practical relevance of the methodological issues. Because our results raise several troubling issues, we conclude that these international efficiency comparisons should not be used to guide management or policy decisions.

The next section of this paper presents an overview of the WHR methodology. The subsequent section identifies two initial problems, aggregation and functional form, and re-estimates the WHR efficiency rankings, using both a production function approach and a cost function approach. The subsequent section focuses on the more fundamental, problem posed by missing variables that are inadequately proxied by dummy variables. Finally, the concluding section summarizes our findings.

THE WHR METHODOLOGY: AN OVERVIEW

The WHR analysis focuses on a composite measure of health care system performance, in which the most fundamental component of this measure is Disability-Adjusted Life Expectancy (DALE). Additional components shown in Table 1 include a measure of the responsiveness of the health care system to demand for non-clinical quality characteristics (RESP), along with three measures of relative equality: the distribution of DALE (DIST), the distribution of RESP

(RDIS), and the fairness of financing (FFIN). These five measures are normalized into the (0,1) interval and combined into a linear composite measure with the weights indicated in the table.

This composite measure of health care system output (Q) is modeled as the outcome of a translog production function of two explanatory variables, health care expenditures and educational attainment. The first explanatory variable is the primary variable of interest, while the second is included to capture the effects of "non-controllable" non-health-system inputs into the production of health. The fixed effects model also includes 190 dummy variables to capture inter-country economic, demographic, political, public health, and cultural differences. A second translog production function, with identical independent variables, is estimated for DALE alone.

The WHR authors estimate the parameters of this production function, and use the results to calculate an overall efficiency ratio for each country. The WHR data set includes five years of data for 141 countries, and one year of data for the remaining 50 countries.

The WHR authors modify the traditional concept of technical efficiency, to develop the overall efficiency scores. Technical efficiency, also called multifactor productivity or total factor productivity in the literature, is usually defined as the ratio of weighted outputs to weighted inputs, normalized to the (0,1) interval. In a normal production process, however, zero inputs would lead to zero outputs. Noting that this would not apply to the provision of health care (because life expectancy would not fall to zero in the absence of health care and education), the authors of the WHR define a measure of overall production efficiency relative to the minimum life expectancy that would occur with zero health care expenditures:

$$WTE_i = \frac{Q_i - \breve{Q}_i}{\hat{Q}_i - \breve{Q}_i},$$

where Q_i = average observed DALE for the *i*th country, $\breve{Q}_i$ = minimum DALE expected with zero health care expenditures, and $\widehat{Q}_i$ = maximum potential DALE consistent with actual health care expenditures. Thus, overall efficiency measures the extent to which a country succeeds in realizing its potential to increase health above the level expected to prevail in the absence of any health care, given its level of health care expenditures.

This minimum level of the health composite measure is developed from a combination of estimation and assumption. The minimum level of DALE is estimated from 1908 data on 25 countries. This estimation is based on two key assumptions: that no country enjoyed the benefits of a functioning health system in 1908, and that 1908 DALE can be modeled as a function of literacy. The 1908 relation between DALE and literacy is used to estimate the hypothetical level of DALE that would have been observed in each country in 1994-1997 in the absence of any health system. The zero-input values of the remaining four performance measures are assumed to be equal to zero (for DIST and RESP) or one (for RDIS and FFIN).

Production efficiency measures estimated by WHR, for both the production of *DALE* alone (this W̲HR D̲ALE e̲fficiency estimate is denoted WDE) and the production of the composite output (WCE denotes this W̲HR c̲omposite e̲fficiency estimate), are provided in the WHR annex tables for all 191 WHO countries. The rank correlation coefficient between the two efficiency measures is 0.87, reflecting the fact that the composite measure is highly correlated with DALE.

Results of the World Health Organization's Global Programme on Evidence for Health Policy are reported in the WHR [1], with more detailed information provided in 30 working papers posted on the WHO website.

RE-ESTIMATING THE RANKINGS: ARE THEY ROBUST?

This section focuses on two of the econometric problems, inappropriate aggregation of heterogeneous performance measures prior to estimation and the logically inconsistent functional form. We examine the practical relevance of these issues by first re-specifying and re-estimating the production function, and then estimating a cost function that provides a more direct approach for addressing the two issues. The re-estimated production function yields efficiency measures and rankings that are positively correlated with the WHR measures and rankings. The cost function, however, yields efficiency rankings that are uncorrelated with the WHR rankings.

First Specification Issue: Inappropriate Aggregation of the Composite Output

The first problem embedded in the WHR specification is inappropriate aggregation of heterogeneous performance measures prior to estimation of the production function. The WHR composite health system performance measure constitutes a valid output measure only if the right-hand-side variables exert equal impacts on each of the components of the composite measure. For a production process in which goods are weighted by marginal values and producers equate the marginal products of inputs across products, it is usually valid to specify aggregate production as a function of common inputs. The five health system outputs are un-priced, however, and three of them measure relative equality rather than output quantities. The WHR study offers neither theoretical nor empirical evidence to support the implicit assumption that these measures meet the equal-impact requirement for aggregation prior to estimation.

We employ two strategies to assess the practical relevance of this issue. First, we estimate separate production functions for each of the five outputs to permit differential impacts of the explanatory variables on the performance measure components, and test for equality of the coefficients in the five equations. This strategy facilitates comparison with the WHR results, but provides only a partial solution to the aggregation problem, because subsequent calculation of five efficiency measures employs the WHR weights to compute a weighted composite efficiency measure. We subsequently address this second problem by estimating a cost function that models health expenditures as a function of input prices and the five outputs.

Second Specification Issue: Logically Inconsistent Functional Form

The WHR translog functional form is logically inconsistent with the explicit assumption that life expectancy would remain positive if healthcare expenditures fell to zero. In contrast to this assumption, the function $Q=f(X)$ is transformed by the translog form to lnQ=g(lnX), which implies the condition $f(0)=0$ (or, in logs, $g(-\infty) \rightarrow -\infty$). Thus, the translog production function does not define Q when one of the inputs (X) is zero, even though the authors of WHR argue Q would be positive when health expenditures are zero. This inconsistency forces the WHR authors to estimate the minimum from out-of-sample data, instead of using the regression results to estimate the minimum.

We address this problem by specifying a non-homothetic Generalized Leontief function of the form:

$$q_{k,i} = \alpha_{k0} + \sum_m \alpha_{km} \sqrt{X_{m,i}} + \frac{1}{2} \sum_m \sum_n \beta_{kmn} \sqrt{X_{m,i}} \sqrt{X_{n,i}} + \varepsilon_{k,i} ,$$

where i indexes countries 1-191, k indexes the five health system outputs (q) included in the composite measure (*DALE*, *DIST*, *RESP*, *RDIS*, and *FFIN*) and m and n index the inputs X (which are assumed by WHR to include only health expenditures H and educational attainment E). This functional form fits the diminishing non-homothetic property apparent in the data, both visually and from preliminary Box-Cox estimates, and it permits output to be positive when input quantities are zero, thus avoiding the logical inconsistency of employing the translog production function in this case. Because this functional form can assume non-zero values when *H* equals zero, this also permits direct estimation of the minimum health output associated with zero health system inputs.

Re-estimation of the Production Function

To assess the practical relevance of these specification issues, we re-estimate the production function using: (1) five separate equations instead of one composite equation, (2) a nonhomothetic Generalized Leontief production function, (3) the one year of data publicly available from WHR[1] (see Table 2), and (4) no dummy variables (which implies that the country characteristics will be captured in the error term ε). We include only the two independent variables (*H* and *E*) included in the WHR production function, in both the full second-order version and a first-order version that facilitates interpretation of the marginal effects.

Estimation results for these two versions of the model are reported in Table 3. We do not present this model as a template for rating health system efficiency; rather, we use the estimation results to assess the econometric issues identified in the previous section.

The re-specified production functions yield reasonable results: the coefficients of the explanatory variables are positive and statistically significant in most cases. While individual coefficients are not all statistically significant, the joint hypothesis that either of the two variables can be dropped is rejected.

Confirming our hypothesis that the composite output was inappropriately aggregated prior to estimation, these results reveal significant differences in the impacts of health expenditure and education on the five separate output components. The joint hypothesis that relative values of both the coefficients for *H* and the coefficients for *E* are equal to the weights given in Table 1 is rejected with an F-statistic equal to 45, which exceeds the critical value of 2.42 for a two-tailed 5 percent test. Thus, the five performance measures should not be aggregated into a composite measure prior to estimation, at least not with the weights given in WHR.

Calculating Production Efficiency Rankings

How do these econometric issues affect the efficiency ratios and country rankings? We calculate overall efficiency for each country for our production function and cost function results, using the WHR efficiency definition. Calculating the production function efficiency ratios requires computation of maximum and minimum values, followed by computation of overall efficiency. Using the first set of reported production function estimations, with both first and second-order effects but excluding additional variables, we calculate the maximum attainable output for each performance measure, given current levels of $X = [H, E]$ and no country-specific differences:

$$\hat{q}_{k,i} = \hat{\alpha}_{k0} + \sum_{m} \hat{\alpha}_{km} \sqrt{X_{m,i}} + \frac{1}{2} \sum_{m} \sum_{n} \hat{\beta}_{kmn} \sqrt{X_{m,i}} \sqrt{X_{n,i}} + \max\left(\hat{\varepsilon}_{k,i}\right).$$

Minimum output for each component is calculated as output given zero health care expenditure, i.e. $X = [0, E]$, which simplifies to:

$$\breve{q}_{k,i} = \hat{\alpha}_{k0} + \hat{\alpha}_{ke} \sqrt{E_i} + \frac{1}{2} \hat{\beta}_{kee} E_i + \min\left(\hat{\varepsilon}_{k,i}\right).$$

These quantities permit computation of an overall efficiency ratio for each performance measure. The numerator is the difference between expected output and the minimum output (expected with zero health system inputs); the denominator is the difference between efficient output and the minimum output:

$$PE_{k,i} = \frac{q_{k,i} - \breve{q}_{k,i}}{\widehat{q}_{k,i} - \breve{q}_{k,i}}.$$

Using the WHR weights, the resulting efficiency measures are used to calculate the composite efficiency index:

$$CPE = 0.25 \times PE_{1,i} + 0.25 \times PE_{2,i} + 0.125 \times PE_{3,i} + 0.125 \times PE_{4,i} + 0.25 \times PE_{5,i}.$$

From this index we derive country rankings; correlations between our calculated efficiency rankings and those given in WHR are reported in Table 4. The correlation coefficient for the rankings produced by the WCE (the WHR measure of composite production efficiency) and the RCE (our revised measure of composite production efficiency) is 0.67. While it is significantly greater than zero, the hypothesis that the correlation between WCE and RCE equals one is rejected[2]. On average, countries move 36 places in rank between the two efficiency measures.

To examine the sources of the differences between the two sets of efficiency ratios, we calculate an alternative measure, RCE2 by (1) using our Generalized Leontief functional form, (2) estimating only one production function for the composite measure *Q*, and (3) calculating the minimum $\breve{Q}$ using the WHR-assumed minimum values for the non-DALE outputs (0 for DIST and RESP, 1 for RDIS and FFIN).[3] The correlation between the RCE and RCE2 rankings is .96, as reported in Table 4. The hypothesis that the correlation between RCE and RCE2 equals one cannot be rejected; hence the difference between the WCE and RCE rankings does not stem from differences in the method used to estimate minimum health. Instead, it reflects changes to either the data set or the functional form. While the correlation between RCE2 and WCE is slightly higher than the correlation between RCE and WCE, the difference is not statistically significant.

Table 4 also includes the correlation coefficients between the efficiency rankings for the DALE-only measures WDE (given in WHR) and our revised DALE-only measure RDE, as derived from the estimation of our first equation, along with those between these rankings and their associated composite measure. The correlation between the WHR results and our results is approximately 0.7, whether output is measured as the WHR composite or DALE-only.

Estimating a Cost Function and Calculating Cost Efficiency Rankings

The cost function provides a more direct approach for addressing these econometric issues. It provides a more compact single-equation representation of the relationship between the primary input *H* and the five health system outputs, it avoids the necessity of calculating a composite efficiency score, it permits positive health output in the presence of zero health inputs, and it provides efficiency scores that do not rely on calculation of minimum outputs. The cost function approach is useful here because it is appropriate for multiple outputs, can be estimated in the presence of fixed inputs, and will permit direct estimation of cost efficiency as the ratio of the minimum potential cost to the observed cost of producing observed output given observed input prices.

Because direct estimation of cost efficiency eliminates the need to calculate minima, there is no need to use the non-homothetic Generalized Leontief form; the translog form originally used in WHR will suffice. This function is of the general form:

$$\ln C_i = \gamma_0 + \sum_m \gamma_m Z_{m,i} + \frac{1}{2}\sum_m \sum_n \theta_{mn} Z_{m,i} Z_{n,i} + v_i \ ,$$

where Z (subscripted by m and n) is a vector of outputs, fixed inputs, and variable input prices, usually expressed in logarithms.

Cost, measured as health care expenditures H, is specified as a function of the five performance measures (DALE, DIST, RESP, RDIS, and FFIN). We assume one fixed input, (education E), and two variable inputs (capital and labor). We assume that the real cost of capital varies over time but does not differ significantly across countries, and cannot be distinguished from the constant because we are using cross-section data. To proxy the relative wage of labor inputs, we include Gross Domestic Product per Capita (G), which is adjusted by the authors of WHR to reflect purchasing power parity. Recognizing the potentially dramatic impact of including G in the cost function, we also estimate this basic cost function without G to facilitate comparison with the WHR results.

Because a cost function is linearly homogeneous in input prices, this specification implies that the elasticity of H with respect to G must lie in the [0,1] interval. Further, the specification of E as a fixed input implies that the elasticity of H with respect to E is negative, and the elasticity of H with respect to each of the five outputs should be positive. Because theory provides no a priori assumptions about returns to scale for the five performance measures, we test whether these should enter linearly or logarithmically; the results indicate we should accept the former and reject the latter. Thus, we specify the elements of Z as [lnE, lnG, *DALE*, *DIST*, *RESP*, *RDIS*, and *FFIN*].

With seven elements in Z, we have 36 parameters to estimate and report. Estimating the full equation using OLS, many of the 28 second-order (θ) terms are insignificant, but we reject the hypothesis that these θ terms are all jointly equal to zero. Retesting this joint hypothesis for the θ parameters with initial t-statistics less than one, indicates that we cannot reject this hypothesis at any reasonable level of confidence. The basic model therefore includes all terms with initial t-statistics greater than one, for which estimation results are presented in Table 5.

Calculating elasticities from the estimated coefficients, we find that the elasticity of *H* with respect to *G* is positive and significant (i.e., the joint hypothesis that its first and second-order coefficients all equal zero can be rejected). However, the elasticity of *H* with respect to *E* has a positive rather than a negative mean (though its range includes negative observations) and it is significantly different from zero. The cost elasticities for both DALE and DIST have a positive mean (with some negative values in their ranges), and they are significant at 5 percent and 10 percent respectively. The cost elasticity for RDIS has a negative mean that is significantly different from zero, while the cost elasticities for both RESP and FFIN are not significantly different from zero. These results reconfirm the hypothesis that the five performance measures should not be aggregated into a composite measure prior to estimation of the production function.

To calculate cost efficiency rankings, we first calculate the minimum attainable health care expenditure given observed values of *Z* as:

$$\ln C_i = \hat{\gamma}_0 + \sum_m \hat{\gamma}_m Z_{m,i} + \frac{1}{2}\sum_m \sum_n \hat{\theta}_{mn} Z_{m,i} Z_{n,i} + \min\left(\hat{v}_i\right),$$

and the cost efficiency index then simply equals:

$$CE_i = \exp\left(\ln \breve{H}_i - \ln H_i\right).$$

The correlation coefficient between the resulting cost efficiency rankings and the original WHR production efficiency rankings equals 0.04, and not significantly different from zero. The cost efficiency rankings are also uncorrelated with the rankings from our production function estimations.

To check whether this lack of correlation is an artifact of our specification, we estimate seven additional versions of the cost function. We consider the effects of excluding *G*, excluding all second-order coefficients, and/or excluding three performance variables (DIST, RDIS, and FFIN) that measure distributional equity rather than output quantity and quality, and we consider all possible combinations of these exclusions. The correlation coefficients between these various cost function efficiency rankings and the WHR rankings range from –0.10 to 0.05. The average absolute change in rank for all estimated versions of the cost function ranges from 64 to 70 places. The correlation is not significantly different from zero in any of these cases. The insignificant correlations between the cost function efficiency scores and the WHR efficiency scores, along with the substantial average changes in rank, are particularly troubling because the cost function provides a more direct approach, compared with the production function, for modeling the relation between health expenditures and multiple health system outputs.

Detailed examination of the cost function results (not reported here, but available upon request) also highlights the relevance of the excluded socio-economic variables. The cost elasticity of *G* is always positive and significant, but the cost elasticity of *E* is only negative (as hypothesized) when *G* is excluded. When *G* is excluded, the cost elasticities of DALE, DIST, and RESP are positive and significant. When *G* is included, however, only the elasticity for RESP is consistently positive and significant.

THE MISSING VARIABLES PROBLEM

The cost function results highlight the significance of the WHR decision to exclude from the production function the economic, demographic, political, public health, and cultural conditions that are widely expected to exert substantial impacts on the production of health. The WHR researchers argue that the production function should not include such variables for three reasons [1]. First, the 190 country dummy variables included in the fixed effects model account for these impacts. Second, these additional variables represent "controllable" variables, which should not be included in a production function. Third, a Hausman test indicates that inclusion of GDP per capita would not increase the explanatory power of the model.

This modeling strategy does not, however, yield results that can be meaningfully applied to the stated goals of the WHR Study. An OLS regression of the WHR composite efficiency ratios on three variables (GDP per capita (*G*), fertility rates (*F*), and reported HIV/AIDs infection rates (*I*)) yields an R-square of 0.66 (see Table 6), with each explanatory variable significant at the 5 percent level.

This type of efficiency measurement system, which indicts countries with low incomes, high birthrates and high reported HIV infection rates, suggests that country leaders can utilize health system resources more efficiently by reducing birthrates, preventing disease and stimulating economic growth. This implication is not very helpful, and not consistent with the stated WHR goal of stimulating "actions that will eventually improve the overall performance of health systems" [2, p 17]. Measuring progress toward this goal requires a measurement system that focuses on how well each country's health system utilizes its resources, given the level of economic development, fertility rates, and HIV infection rates.

The WHR decision to exclude "non-controllable" variables (other than education) is not consistent with the WHR definition of the health system boundary. WHR 2000 researchers define the boundary of the health system by focusing on the concept of a health action [7, p 4]. The primary intent of a health action is to improve or maintain health. WHR 2000 researchers explain this concept via an example: measures taken to increase education of young girls are not health system inputs, even though increased female education may improve health outcomes because health improvements are not the primary intent. According to this definition, expenditures designed to increase development, for example, are not health system inputs; hence the measurement of health system efficiency should be independent of a country's level of development.

Do the Dummy Variables Substitute for the Relevant Socioeconomic Variables?

The WHR approach of estimating a fixed effects model, instead of including relevant socio-economic variables in the production function, is problematic for two types of reasons. First, the data set is not sufficient for estimating the fixed effects model and, second, this approach does not mathematically separate country characteristics from country inefficiency.

The WHR data set is not sufficient to estimate a fixed effects model because (a) the five-year observation window is too short to observe meaningful intra-country variation in variables such as life expectancy, educational attainment, and GDP per capita, (b) only one year of data is available for approximately one-fourth of the 191 countries [2, p 7], and (c) a substantial portion of the five-year data set is estimated rather than observed.

The estimation of much of the data set is particularly problematic in this situation, because some of the potentially relevant (but excluded) variables are employed as independent variables in the estimating equations. RESP and RDIS, for example, are estimated from information developed in a one-time survey of knowledgeable individuals in 35 countries (See Valentine [10] and Williams [4]). This information is used to estimate relationships between the components of responsiveness and independent explanatory variables such as geographic access, percent of population below the poverty level, GDP per capita, and education. These relationships are used to estimate five years of data for RESP and RDIS for all 191 countries. Thus, some dependent variables are constructed as linear combinations of potentially relevant (but intentionally excluded) explanatory variables. Williams [4] discusses the data estimation procedures in more detail and draws the conclusion that the data is insufficient to support the WHR research program. We point here to the consequent modeling problem: if the multi-year data set does not include independent observations for each year, country dummy variables will not account for country characteristics in a meaningful way.

While fixed effects models are appropriate for estimating production functions in some situations, this approach is problematic in this situation because (a) the missing variables are expected to exert substantial impacts on the production of health and (b) the stated goal is to develop efficiency measures that provide incentives for efficient resource use within the health care system. Inter-country differences in health-care requirements are not distinguished in a fixed effects model, however, from inter-country differences in inefficiency. The dummy variables simply measure the average residual for each country; they do not prevent relevant country characteristics factors from affecting the estimation of efficiency. Inclusion of both a dummy variable and an error term in the fixed effects model separates country characteristics into two terms, but it is not clear that it will do this by isolating country-specific inefficiency in the error term. Including socio-economic variables in the production function gives explicit consideration to country characteristics that impact the production of health.

Should Socioeconomic Variables be Excluded Because they are Controllable?

The WHR Report argues that socio-economic variables should not be included in the production function because they are not “controllable”. Instead, the WHR production function includes education as a controllable variable. This variable, according to WHR, measures education as a fixed input into the process of producing health, and serves as a proxy for non-health systems inputs [8, p. 70]. This logic suffers from three problems. First, the concept of "controllable" and "uncontrollable" inputs is problematic in the context of an international health care production function because it begs the questions, "controllable by whom?" and "controllable to what extent?" Second, education is likely to proxy development. Since development is frustratingly non-controllable, the WHR logic would exclude it from the production function. Third, excluding non-controllable inputs from the production function yields efficiency measures that are strongly influenced by variables that are outside the sphere of influence of each country's health system, as defined by WHR [8, p 4]. These measures are not, therefore, useful for assessing the efficiency with which each health system utilizes its resources.

Does the WHR Hausman Test Justify Exclusion of Per-Capita GDP?

After arguing that per-capita GDP (*G*) is not a conceptually appropriate explanatory variable, the WHR study authors conducted a statistical test to assess whether *G* should be included in the production function. They regressed *G* on the independent variables, [8, p. 12] and then included the residual from this regression in the production function. They found that the residual coefficient is statistically insignificant; hence neither *G* nor the residual are included in the WHR model.

The Effects of Including Socioeconomic Variables in the Production Function

We assess two issues. First, we repeat the Hausman test with a production function that does not include the dummy variables. As reported below, the estimated coefficient on *G* is significant in this case. Second, we examine the impact of including three socio-economic variables on the production function and efficiency estimation results.

To assess the practical relevance of the decision to exclude socioeconomic variables, we re-estimate the first-order version of the five production functions given earlier, adding *G*, *F*, and *I* to each equation. These results, which are reported in Table 7, confirm that including per-capita GDP (*G)*, fertility (*F)*, and reported HIV/AIDS infection rates (*I*) exerts significant impacts on the estimation results. When these variables are included in the production function, the coefficient on health care expenditures becomes insignificant in three of the equations (DALE, DIST and FFIN) and significantly negative in another (RDIS). The coefficient for *H* is only positive in the RESP equation, where it is significant at 10 percent.

Because the effect of *H* is no longer significantly positive, calculation of the efficiency scores is problematic. How can we measure the efficiency of health care expenditures if they do not improve health performance? We calculate it nonetheless, to assess the magnitude of the impact, and find that inclusion of the missing variables *G*, *F*, and *I* in the production function alters the efficiency ranking dramatically. The correlation coefficient ($\rho_{WCE/RCE3}$) between this new revised composite efficiency ranking (RCE3) and the original WHR rankings (WCE) drops to 0.33. While this correlation is significantly greater than zero, adding three socioeconomic variables to the production function reduced the correlation between our re-estimated efficiency scores and the WHR scores by approximately 50 percent. This correlation coefficient reduction is statistically significant: the confidence intervals for the two correlation coefficients ($\rho_{WCE/RCE3}$ and $\rho_{WCE/RCE}$) do not overlap. The average change in rank between the original WHR rankings and these new rankings is 51 places out of 191 countries.

CONCLUSION

We conclude that the efficiency ranking effort, while laudable, is premature. The results are not robust with respect to two fundamental decisions: to focus on a production function, rather than a cost function, and to exclude all socioeconomic explanatory variables except education.

The strong statistical relationship among the WHR efficiency measures, HIV/AIDs infection rates, fertility rates, and G undermines the WHR vision that the efficiency scores can be used to measure and induce increased productivity of health care expenditures. An international aid allocation strategy designed to reward countries with efficient health care systems, as measured by the WHR methodology, would channel aid to countries with low reported HIV/AIDs infection rates, low fertility rates, and high GDP per capita; clearly, such a system would not provide incentives for more efficient use of health system resources.

The WHR's strategy of including only controllable variables in the production function is widely used, and is appropriate for studies designed to analyze exogenous factors that affect production efficiency. This two-step approach (first estimate efficiency without consideration of the exogenous factors, then analyze the impacts of the exogenous factors on the efficiency measures) does not provide the information needed to meet the WHR goals, however. Meeting these goals will require efficiency measures that answer the question: how well does each country's health care system utilize its resources, given its socioeconomic characteristics.

The results of our paper point to the fundamental problem that plagues every effort to analyze health production: while many products, such as steel, are largely produced within reasonably well-defined industries, health is produced both inside and outside the health care industry. The relationship between non-health system inputs and health outputs has been well documented, but it is not well understood. This knowledge gap may lead researchers to omit such variables from efficiency analysis, but failure to include relevant inputs in the health production function is likely to distort efficiency results. The complexity of the issues at hand are illustrated by the fact that GDP per capita may be positively correlated with health system output measures for three distinct reasons. First, high GDP per capita facilitates provision of public health measures (such as clean water, improved nutrition, sanitation and vaccinations) and personal consumption patterns (such as adequate diet) that are inputs into the public and household production of health. Second, GDP per capita proxies labor costs. Failure to include a measure of inter-country wage differences will bias the health expenditure coefficient if countries with high wage rates experience relatively high health care expenditures that primarily reflect high input prices, rather than high input quantities. Third, patients in high-GDP countries may have above-average willingness to pay for types of health care that do not extend life expectancy. As income rises, health care delivery may broaden its focus beyond prolonging life to include more quality of life measures that alleviate pain, increase joint mobility, or address reproductive problems. Such health care expenditures would appear, in the WHR template, as inefficient resource utilization. Thus, producing efficiency measures that will accomplish the WHR goal will require better clarification of health system inputs, the relationship between inputs and outputs, and actual health system outputs.

---

[1] The complete data set, with five years of data for 141 of 191 countries, was requested from Dr. Murray and others in the WHR research project, but it was not provided to us. Dr. Evans graciously provided a portion of Evans, et al. [5] which had been cleared by WHO.

[2] We focus our discussion on the correlations between sets of country rankings because the of policy relevance of these rankings. We test whether these ranking correlations are significantly greater than zero. However, all other hypothesis tests are conducted for correlations among the sets of efficiency ratios.

[3] The logic of these assumptions is murky because the last three performance measures are conceptually meaningless in the absence of a functioning health care system (see Murray, et al. [7], p 8).

REFERENCES

1. World Health Organization. *The World Health Report 2000: Health Systems: Improving Performance.* Geneva, World Health Organization, 2000.

2. Tandon A, Murray C, Lauer J, Evans D. Measuring Overall Health System Performance for 191 Countries, Geneva, World Health Organization, Global Programme on Evidence for Health Policy, Discussion Paper Series, No. 30, 2000.

3. Blendon RJ, Kim M, Benson JM, The public versus the World Health Organization on health system performance. *Health Affairs*, May/June 2001.

4. Williams A, Science or marketing at WHO? A Commentary on 'World Health 2000'. *Health Economics* 2001; **10**: 93-100.

5. Evans DB, Bendib LM, Tandon A, Lauer J, Ebener S, Hutubessy RCW, Asada Y, and Murray CJL. Estimates of income per capita, literacy, educational attainment, absolute poverty, and income Gini coefficients for the World Health Report 2000. Geneva: World Health Organization, Global Programme on Evidence for Health Policy, Discussion Paper No. 7, 2000.

6. World Health Organization. Global situation of the HIV/AIDS pandemic, end 1999. Geneva: World Health Organization, 2000. http://www.who.int/ emc/ diseases/ hiv/ Documents/ Report1999.pdf .

7. Murray CJL and Frenk J. A WHO Framework for Health System Performance Assessment. Geneva: World Health Organization, Global Programme on Evidence for Health Policy, Discussion Paper No. 6, 2000.

8. Evans DB, Tandon A, Murray CJL, and Lauer JA. The Comparative Efficiency of National Health Systems in Producing Health: An Analysis of 191 Countries. Geneva: World Health Organization, Global Programme on Evidence for Health Policy, Discussion Paper No. 29, 2000.

9. Valentine NB, de Silva A, Murray CJL. Estimating Responsiveness Level and Distribution for 191 Countries: Methods and Results, Geneva: World Health Organization, GPE Discussion Paper Series No. 22, 2000.

Table 1. Components of the WHR 2000 composite output for health system performance

Measure	Composite Weight
DALE – Disability-Adjusted Life Expectancy	25.0%
DIST – Distribution of DALE	25.0%
RESP – Responsiveness	12.5%
RDIS – Distribution of Responsiveness	12.5%
FFIN – Fairness in Financing	25.0%

Table 2. Data set

Variable	Year	Source
Performance Measures:		
DALE and DIST	1999	Annex Table 5 of WHR 2000
RESP and RDIS	1999	Annex Table 6 of WHR 2000
FFIN	1997	Annex Table 7 of WHR 2000
WHR Efficiency Measures:		
WCE – Composite production efficiency		Annex Table 10 of WHR 2000
WDE – Production efficiency for DALE		Annex Table 10 of WHR 2000
Exogenous Variables:		
G – GDP per capita, in thousands of 1997 purchasing-power-parity US Dollars	1997	Annex 1 of Evans, et al. [5]
E – Educational attainment	1997	Annex 1 of Evans, et al. [5]
H – Health care expenditures per capita, calculated as the product of G and the health expenditure share, including total public and private health care expenditures	1997	Annex Table 8 of WHR 2000
A – Share of population aged 65 or older	1999	Annex Table 2 of WHR 2000
F – Fertility rate	1999	Annex Table 2 of WHR 2000
I – Percentage of population with reported HIV/AIDS infection	1999	Table 2 of WHO 2000 [6]

Table 3. Production function re-estimations

Independent Variables	Dependent Variables *DALE*	*DIST*	*RESP*	*RDIS*	*FFIN*
first and second order:					
Constant	0.230 (0.094)**	0.370 (0.087)**	0.199 (0.038)**	0.507 (0.063)**	0.945 (0.074)**
$\sqrt{H}$	1.124 (0.186)**	1.133 (0.172)**	0.098 (0.075)	0.567 (0.124)**	0.123 (0.145)
$\sqrt{E}$	0.055 (0.105)	-0.092 (0.097)	0.274 (0.042)**	0.149 (0.070)**	-0.065 (0.082)
$H/2$	-0.018 (0.130)	-0.260 (0.120)**	0.018 (0.052)	-0.146 (0.086)*	-0.065 (0.101)
$\sqrt{H \times E}$	-0.331 (0.087)**	-0.254 (0.080)**	0.039 (0.035)	-0.121 (0.058)**	-0.002 (0.068)
$E/2$	0.089 (0.056)	0.139 (0.052)**	-0.105 (0.023)**	-0.021 (0.037)	0.028 (0.044)
R^2	*0.744*	*0.799*	*0.906*	*0.660*	*0.136*
first-order only:					
Constant	0.299 (0.031)**	0.285 (0.029)**	0.403 (0.012)**	0.666 (0.021)**	0.889 (0.022)**
$\sqrt{H}$	0.151 (0.023)**	0.190 (0.022)**	0.211 (0.009)**	0.086 (0.016)**	0.067 (0.016)**
$\sqrt{E}$	0.160 (0.015)**	0.157 (0.014)**	0.055 (0.006)**	0.080 (0.010)**	-0.001 (0.011)
R^2	*0.684*	*0.739*	*0.890*	*0.567*	*0.130*

Note: Standard errors in parentheses

*, ** Difference from zero statistically significant at 10% and 5% levels (two-tailed), respectively

Table 4. Correlation coefficients among production function efficiency rankings

Measure	WHR 2000 composite efficiency index (WCE)	Re-estimated composite efficiency index (RCE)	WHR 2000 DALE-only efficiency index (WDE)	Re-estimated DALE-only efficiency index (RDE)	Re-estimated composite efficiency index with G,F,I (RCE3)
WCE	1.00				
RCE	0.67	1.00			
RCE2	0.72	0.96			
WDE	0.88		1.00		
RDE		0.78	0.68	1.00	
RCE3	0.33				1.00

Table 5. Cost function estimation

Estimated Parameter	Coefficient	Standard Error	Estimated Parameter	Coefficient	Standard Error
γ_0 (CONSTANT)	-0.406	(3.124)	θ_{G1} (lnG×DALE)	-0.876	(0.371)**
γ_E (lnE)	2.272	(0.627)**	θ_{E1} (lnE×DALE)	-0.736	(0.727)
γ_1 (DALE)	-11.538	(4.173)**	θ_{E4} (lnE×RDIS)	-2.207	(0.762)**
γ_2 (DIST)	7.746	(5.504)	θ_{11} ($DALE^2$)	8.003	(4.750)*
γ_3 (RESP)	-9.806	(7.171)	θ_{12} (DALE×DIST)	3.669	(2.089)*
γ_4 (RDIS)	2.607	(1.093)**	θ_{15} (DALE×FFIN)	6.289	(4.543)
γ_5 (FFIN)	-3.672	(3.202)	θ_{25} (DIST×FFIN)	-10.372	(6.178)*
γ_G (lnG)	0.918	(0.250)**	θ_{35} (RESP×FFIN)	11.092	(7.834)
θ_{GE} (lnG×lnE)	0.385	(0.128)**	R^2	*0.951*	

Note: Standard errors in parentheses
*, ** Difference from zero statistically significant at 10% and 5% levels (two-tailed), respectively

Table 6. Determinants of WHR 2000 composite efficiency

Dependent Variable: WCE
OLS regression

Independent Variable	Coefficient	Standard Error
Constant	0.723	(0.031)**
G	0.015	(0.002)**
F	-0.050	(0.007)**
I	-0.284	(0.072)**
R^2	*0.657*	

Note: Standard errors in parentheses
** Difference from zero statistically significant at 5% level (two-tailed)

Table 7. Production function with additional variables

Independent Variables	Dependent Variables *DALE*	*DIST*	*RESP*	*RDIS*	*FFIN*
WHR variables + additional variables, first-order only:					
Constant	0.826 (0.063)**	0.669 (0.070)**	0.333 (0.029)**	0.806 (0.052)**	0.718 (0.059)**
$\sqrt{H}$	0.002 (0.050)	-0.019 (0.055)	0.043 (0.023)*	-0.127 (0.041)**	0.037 (0.047)
$\sqrt{E}$	0.057 (0.015)**	0.085 (0.016)**	0.062 (0.007)**	0.050 (0.012)**	0.025 (0.014)*
$\sqrt{G}$	0.044 (0.015)**	0.059 (0.016)**	0.054 (0.007)**	0.064 (0.012)**	0.015 (0.014)
$\sqrt{F}$	-0.148 (0.019)**	-0.123 (0.021)**	0.013 (0.009)	-0.053 (0.015)**	0.054 (0.017)**
$\sqrt{I}$	-0.205 (0.028)**	-0.072 (0.031)**	0.005 (0.013)	-0.017 (0.023)	-0.013 (0.026)
R^2	*0.830*	*0.808*	*0.918*	*0.656*	*0.176*

Note: Standard errors in parentheses
*, ** Difference from zero statistically significant at 10% and 5% levels (two-tailed), respectively