# Technical Note

August 2001

Supplement to Dean T. Jamison and Martin E. Sandbu, "WHO Ranking of Health System Performance," *Science*, 293 (31 August 2001), 1595-6

This note describes the methods used in our paper "WHO Ranking of Health System Performance" to reconstruct the WHR 2000 performance rankings and to examine the robustness of those rankings to changes in assumptions about the attribution of health outcomes to health system performance versus other factors.

## I. The World Health Report Methodology

This section will specify and scrutinise the conceptual framework on which the World Health Report 2000 (WHR 2000) "performance rankings" are based. That framework is described in a series of papers by WHO staff (Murray and Frenk 2000, Evans *et al.* 2000) and starts by identifying *performance* with *efficiency*. This efficiency measure is a measure of how the actual output of the health system (actual health status) is related to the maximum output that could potentially be achieved with the same health system resources (which are broadly conceived, *e.g.* the total expenditure on health in dollars). A fully efficient health system is one that produces this maximum health output (this health system would have a "performance" score of 1 in WHR 2000). A completely inefficient health system would be one that produces the same health output that we would observe in a country without a health system but with the same resources[1] (a "performance" score of 0). In the graph in figure 1 of our paper, the relative efficiency of a health system is therefore measured by the following formula:

$$(1) \qquad \text{Efficiency(country } i) = [(OD_i + (PHO_i - LF_i)] / [(OD_{max} + (PHO_i - LF_i)] .$$

where OD denotes the observed 'outcome deviation' for the country (the deviation of its health measure from what is statistically predicted given its resources), PHO is the predicted health outcome, LF is the lower frontier (the lowest possible health outcome with that country's resources) and $OD_{max}$ is the outcome deviation of the best-performing country (so $OD_{max}$ + PHO is the upper frontier).

The challenge in this kind of exercise is how to estimate the two light curves in the graph. They represent the maximum health output that can be produced for a given quantity of inputs – we shall refer to this as the *(upper) frontier* – and the minimum health level that can be expected with these resources even without any functioning health system – we shall call this the *lower bound*.

The *country performance index* for female life expectancy in WHR 1999 is simply the outcome deviation (relative to what would be predicted by income) expressed in years, i.e. for India it was + 0.6 years. Is it clear from the above equation that, as with country performance, the algorithm for *health system* performance in WHR 2000 becomes a function of $OD_i$ , albeit a complex and

[1] It is, however, difficult to understand what exactly this means. If a health system is to be understood, as the authors of WHR 2000 suggest, as the set of "health actions" (that is actions aiming to improve health), how could one have expenditure on health without a health system?

non-transparent one that requires multiple assumptions (that could importantly affect rankings) for calculating the upper and lower bounds.

## II. Estimation of upper frontier and lower bound in WHR 2000

WHR 2000 uses Disability-Adjusted Life Expectancy (DALE) for 1997 as the measure of health status in each country.[2] The methodology for estimating the frontier is given in Evans *et al.* (2000). It uses a multiple regression model of the relationship between (the natural log of) actual DALEs and two independent variables: (the natural logs of) countries' total health expenditures and their level of educational attainment. Using a data panel from 1992-1997, Evans *et al.* (2000) estimate the coefficients on the independent variables as well as country-specific "fixed effects".[3] A country's *upper frontier* is then taken to be the DALE that is predicted by plugging that country's health expenditure and educational attainment into the estimated regression, plus the *highest* fixed effect in the sample. Thus the best-performing country (that with the highest fixed effect/intercept/unexplained time-invariant outcome deviation) is defined to be on its frontier, and for all other countries the difference between their intercepts and that of the best-performing country is taken to be their distance from their own respective upper frontiers.

Evans *et al.* (2000) gives the estimated frontier function used to construct the performance rankings:

$$\ln(\text{FRONTIER}_i) = 3.8125 + 0.2144 + 0.0088*\ln(\text{HEXP}_i) + 0.0630*\ln(\text{EDUC}_i) + 0.0217*[\ln(\text{EDUC}_i)]^2 \quad (2)$$

where FRONTIER is the upper frontier DALE value, the first constant is the average fixed effect or intercept, the second constant is the highest positive country deviation from the average fixed effect, HEXP is absolute health expenditures in international dollars and EDUC is the average number of years of schooling in the adult population. The values for these variables, except country estimates for FRONTIER, are given in WHR 2000 and reproduced in the attached spreadsheet.

As for the lower frontier, WHR 2000 and the accompanying technical paper Evans *et al.* (2000) give scant and sometimes inconsistent information. It emerges that it was based on the health status of a small number of countries *circa* 1900, since these data points are taken to provide information on the relationship between a country's resources and its health level in the absence of a modern health system. This relationship was then applied to countries' current level of resources and shifted down to ensure that no country fell below their lower bound.

---

[2] For a description of DALEs, see Mathers *et al.* (2000). Note that data on life expectancy and DALE are virtually perfectly correlated (female life expectancy in 1992 has a .95 correlation coefficient with DALEs in 1997). This indicates that the use of DALEs imposes the cost of lower transparency for virtually no gain and, in particular, that this is not a source of the difference between the WHR 2000 and the WHR 1999 rankings that we draw attention to in our paper.

[3] "Fixed effects" can be though of as the over-time average deviation of a country's DALE from the level predicted by the estimated equation. In other words, given the estimated "health production function", each country has its own specific, time-invariant, DALE-intercept which shifts that production function up or down.

None of the WHO publications to date give the estimated function for the lower bound.

## III. The Problems with the WHR 2000 Methodology

While the lower bound is not unproblematic, we have taken it as given for the purposes of this paper that such a lower bound makes sense and that it can be estimated in a satisfactory manner. The rest of this note is therefore focused on the interpretation and estimation of the upper frontier.

Evans *et al.* (2000) use a specification with total health expenditure and educational achievement as the independent variables and disability-adjusted life years as the dependent variable. While it is of course straightforward to carry out the econometric exercise described above, it is not at all clear how to interpret it, and it is the chosen interpretation of the results that validates or invalidates the choice of inputs in the production function. The papers that accompany WHR 2000 offer two alternative and incompatible accounts. The first is that the upper frontier reflects how much the health system could achieve if it used its resources optimally to improve those determinants of health *for which it can be held accountable*. Under this interpretation, the frontier must of course be adjusted (econometric controls must be included in its estimation) for everything for which the health system should *not* be held accountable. This is the interpretation suggested by Murray and Frenk (2000, pp. 722, 727).

The alternative interpretation is that the upper frontier reflects the maximum health level that could be achieved if the health system made better use of all the determinants of health that it *controls* to a certain extent. The difference is subtle but important. The health system may have as much (or rather, as little) effective control over macroeconomic policy or the quality of the judicial system[4] as over education policy, but we may think that it should be held more accountable for the effects on health of the latter than the former.

It is unclear which interpretation has been used to set up the model, and under both interpretations the setup is flawed: the upper frontier is estimated with either too many or too few variables.

The inclusion of health expenditures in the production function is unproblematic. After all we want to examine how well a health system can do given the resources at its disposal. The problem concerns what else we include. Let us start with the second approach, which says that "*controllable* inputs into the production process are … related to outputs to measure efficiency"[5] whereas *uncontrollable* ones are not included.[6] However, the rationale for considering the educational attainment as a "controllable" variable is unclear. Why is educational attainment any more "controllable" than macroeconomic policy, quality of governance, or other spheres of policy? Moreover, to the extent that it is "controllable", it is presumably already accounted for in health expenditures – given that the latter are the resources of the health system, they capture the

---

[4] Murray and Frenk (2000, p. 727) offer "the presence of an effective judicial system" as a factor that might influence the performance of the health system.

[5] Evans *et al.* (2000, p. 9), emphasis added.

[6] According to Evans *et al.* (2000), these uncontrollable factors are better analysed at a second stage as possible *causes* of differences in efficiency.

extent to which the health system can influence educational attainment by devoting its resources to lobbying the decision-makers in education policy (for example). Therefore it seems that educational attainment ought not be included as a health system *resource*, but rather as one of many types of "interventions" to which the health system may choose to allocate its resources. In any case, in terms of *controllability* there does not seem to be any reason to include educational attainment as the only addition to health expenditures in the estimation of the frontier.

In fact, *controllability* is not a plausible criterion here. It is far preferable to conceive of the upper frontier as that potential level of health that the health system *can be held accountable* for achieving or failing to achieve. Thus we should include controls for the factors that the health system cannot be faulted or credited for. Murray and Frenk acknowledge as much when they say: "The potential for the health system to influence … determinants … such as educational attainment, general social inequalities and biodiversity, is much less [than problems such as tobacco consumption and diet] and *the assessment of health system performance should reflect this fact*." (Murray and Frenk 2000, emphasis added)

In line with the preceding quote, the architects of the WHR performance rankings adjust the health systems' upper frontiers for the effects on health of educational achievement, thereby making health systems unaccountable for those effects. Thus the WHR efficiency measure does not classify a country with lower health status as less efficient if that worse health status is due to a relative lack of educational attainment. Conversely, the measure demands more of countries with highly educated populations in order to give them the same efficiency score.

This seems eminently reasonable. But the same reasons that make it natural to include educational attainment in the regression equation also make it unreasonable to exclude the non-health system determinants of health for which the health system ought to be held even less accountable. These could be general government policy (including economic policy), the quality of governance (rule of law), infrastructure, and, notably, geography. If a health system should not be deemed relatively inefficient because of relatively low educational attainment in the population, then it certainly ought not be considered relatively inefficient because the country happens to be in the tropics!

Thus a measure of health system efficiency, to make sense, must measure actual achievement relative to the maximum that can be attained, where this maximum is adjusted for *all* the determinants of health that the health system should not be held responsible for. WHR 2000 does not do this – it only adjusts for differences in educational attainment. Evans *et al.* (2000) explains that the authors tried to include GDP, but that this did not make much difference to the ultimate rankings. Other determinants were not included because "[i]dentifying relevant variables that are available for all countries" was "difficult" (p. 12). The omission of these variables would be inconsequential if it did not affect the rankings; however, such a robustness test is not documented in WHR 2000 and the related papers. We therefore tried to test the robustness of the rankings to more extensive adjustments of the frontier.

## IV. Replication of the WHR rankings

In order to test the robustness of the rankings to modifications of the frontier, we needed to know the frontier and lower bound used in WHR 2000. Since some of the data and results used in the WHR 2000 analysis remained unavailable at the time our paper went to press, we had to replicate that analysis to the best of our ability. The following paragraphs describe our methods.

From WHR 2000 we get the DALEs and the absolute level of health expenditures in international dollars for the 191 member countries of WHO. For the education data we used the average number of years of education in the adult population (male and female) as per 1990 from the Barro-Lee data set. Data constraints forced us to limit the analysis to 97 out of the 191 WHO member countries. When we refer to a country's ranking in WHR 2000 in what follows, we therefore mean the ranking of its health system efficiency (as given in WHR 2000) *within* our subset of 97 countries, expressed as percentiles. The list of countries is given at the end of this note – the countries represent the entire range of WHO member countries in terms of their WHR 2000 performance score, so the results to be presented below are not merely artefacts of the smaller sample.[7]

Our first challenge was to construct an upper frontier and a lower bound as similar as possible to the (unreported) ones of WHR 2000. We used the WHR 2000 frontier production function (from equation 2) to derive a frontier DALE value for each of the 97 countries. Since we do not have access to exactly the same data as the authors of WHR 2000 used, these are of course not exactly the same values as the ones underlying the WHR 2000 performance table. Specifically, just plugging in the data generated upper frontier values such that a few countries were just above their frontier. To adjust for this, we shifted the frontier values up by 1.33 DALEs, which was enough to place the best-performing country (Portugal in our sample) exactly on its frontier. The exact values, as well as all the other data we have used for this exercise, are given in the attached spreadsheet. While the values cannot be the same as the (unpublished) WHR 2000 values, we are close enough that if the WHR 2000 methods really are robust, the performance rankings derived with our data should be stable as well.

Recall that WHR 2000 defines health system efficiency by:

$$(3) \qquad EFF_i = [DALE_i - MIN_i]/ [FRONTIER_i - MIN_i],$$

where EFF is efficiency, DALE denotes the actual observed DALE values and MIN denotes the lower bound values. Since we derived values for FRONTIER as described in the previous paragraph, we were able to extrapolate lower bound values for each country. By construction, then, our set of frontier values and lower bound values together with WHR 2000 DALE observations exactly replicates the WHR2000 efficiency values and thus the rankings.

## V. Adjusting for long-term fixed effects

Given the range of important cross-country differences that are not attributable to the health system but still determine disparities in health, we expected the rankings to be sensitive to how the frontier was adjusted for these factors. The problem, as Evans *et al.* (2000) point out, is that very little evidence about these differences is available on a cross-country basis. Indeed, this is one reason why the WHR rankings should be taken with a large pinch of salt.

There is one set of factors that are both easily measured and relevant for health, however, namely geography. The geographical circumstances of a country are clearly important determinants of health outcome; moreover, they are even less the responsibility of the health system than is educational attainment. If the WHR 2000 method adjusts the frontier for educational attainment,

---

[7] As shown in the attached spreadsheet, the subsample we used ranges evenly over the entire WHR 2000 efficiency ranking, so our robustness test is not invalidated by an unrepresentative sample of countries.

then, it should also adjust it for geographical factors. One cannot blame the health system for a country's being located in the tropics.

We now used our derived upper frontier and lower bound (recall that they were constructed to be as faithful to the WHR 2000 methods as possible without better access to the underlying data, and specifically, that they exactly replicate the WHR 2000 efficiency scores) to test the robustness of the method. Ideally, we would have wanted to carry out an exact replication of the WHR 2000 regression but with geographical variables included. Since the full data panel used by the authors of WHR 2000 is unavailable, we instead ran the following simple regression:

$$(4) \qquad \ln(DALE_i) = \alpha + \beta_1 * \ln(HEXP_i) + \beta_2 * \ln(EDUC_i) + \beta_3 * [\ln(EDUC_i)]^2 + \gamma_1 * TROPICS_i + \gamma_2 * COASTAL_i + \varepsilon_i$$

where DALE and HEXP values are from WHR 2000, EDUC is from Barro-Lee as above, and TROPICS and COASTAL denote the fraction of a country's land area situated between the geographical tropics and within 100km of the coast or an ocean-navigable waterway, respectively.[8]

We could now calculate the partial effect on health of the geographical variables. We took the average values of TROPICS and COASTAL to be the 'neutral' geographical state, so that a country with average geographical location was taken to have a geography effect of zero. Thus the effect of geography on health in a country is the effect of deviations from mean in the geographical variables:

$$(5) \qquad GEOFE_i = \gamma_1 * (TROPICS_i - AVGTROP) + \gamma_2 * (COASTAL_i - AVGCOAST)$$

where GEOFE is the geographical effect, and AVGTROP and AVGCOAST are the sample averages of the variables TROPICS and COASTAL. Thus the sample average of GEOFE is zero.

To adjust the frontier, we once again had to approximate, given the fact of unavailable data. Ideally, we would have adjusted the fixed effects of all the countries' production functions by the effect of geography and computed the new upper frontier function now using the highest positive *net* fixed effect as the maximum achievable level of health. This was not possible since the individual country fixed effects have not been published (only that of the best-performing country). We therefore had to make do with the second best, which was simply to add the geography effect to the (logarithmic) frontier values whose derivation we described above. This amounts to assuming that if part of a country's high achievement on health is a geographical 'windfall', so to speak, then we should expect correspondingly more from its health system, that is, the frontier is shifted up. Similarly, if a country's low health status is partly due to its unfortunate geographical location, we do not want to blame its health system for that, and so we shift the frontier down by the part caused by inclement geography.

For example, Austria has a frontier value (in DALEs) of $75.2 = e^{4.32}$ in our original construction. The geographical effect is .07 (in logarithmic terms), so its frontier is shifted up to $80.4 = e^{4.39}$. Since its good health is partly caused by favorable geography, in other words, this raises our expectations of what the health system should be able to achieve. For another example, consider

[8] These variables were provided by John Gallup and Jeffrey Sachs. More detailed information about them can be found on http://www.cid.harvard.edu/ciddata/ciddata.html. They are also reproduced in the attached spreadsheet.

Cameroon: Its original frontier value (in DALEs) is $65.5 = e^{4.18}$, but the effect of geography on health is -.11, so the adjusted frontier value is $58.8 = e^{4.07}$.

We did this for all countries, and then recalculated the efficiency scores.[9] This procedure improves the efficiency scores of countries with relatively adverse geography and worsens the score of those whose geography is relatively beneficial for their health outcomes. To continue the previous example, Austria's efficiency score falls from .914 to .812, whereas as that of Cameroon increases from .399 to .482.

Finally, we re-ranked the countries in our sample according to our new efficiency scores.

The estimation of the geography effects described in this section was carried out with the STATA statistical software. The STATA log-file is included at the end of this note.

## VI. Robustness of rankings with the WHR 2000 methods

The above derivation puts us in a position to test the stability of the WRH 2000 rankings. These are given with 'uncertainty intervals', and it seems that a sensible test for their robustness is to check whether the ranks derived with an adjustment for geography mostly fall within those intervals. One problem with this was how to transform the intervals from WHR 2000 when going to our smaller sample. We converted the WHR 2000 ranks and rank uncertainty intervals into percentile terms. Since our sample included countries from the whole range of WHR 2000 ranks, without any lumping (see the chart in the accompanying spreadsheet), this is a valid comparison. We then also converted the new, adjusted ranks into percentile terms, and checked if the countries' new percentile scores fell within their old percentile scores.

The result is shown in the graph in the main paper and in the accompanying spreadsheet. As can be seen from those documents, an overwhelming number of countries – 79 out of 96 – fall outside their uncertainty intervals. The average country moves 14 percentiles between rankings and 26 countries move 20 percentiles or more. Bolivia, for example, performed poorly according to *WHR2000,* ranking at the 26th percentile. After adjusting for geography, however, Bolivia's percentile rank jumps to 52, i.e. by a distance of over *six* times the width of its uncertainty interval (which places it between the 24th and the 28th percentiles). Thus we conclude that the method used to generate the ranking of 'health system performance' is not robust to the exact specification of which factors the health system is or is not held responsible for.

[9] The only country we could not do this for, was Portugal, the highest-ranking country. This is because we had adjusted the frontier function to make sure no countries were above their frontier. We did this by shifting it up until Portugal was exactly on its frontier. This, of course, means that we could not derive a meaningful value for Portugal's lower bound: Mathematically, it is exactly the same as its upper frontier and its actual DALE value. This makes the efficiency score definitionally zero when we adjust the frontier by the effect of geography, but this is not a meaningful change. The rest of the results discussed therefore exclude Portugal and compare the ranks of the remaining 96 countries with their ranks within the same subset of countries in WHR 2000.

## VII. Cited works


Evans, David B., Ajay Tandon, Christopher J. L. Murray and Jeremy A. Lauer. 2000. "The Comparative Efficiency of National Health Systems in Producing Health: An Analysis of 191 Countries". Global Programme on Evidence Discussion Paper No. 29. Geneva: World Health Organization.

Mathers, C., R. Sadana, J. Salomon, C. J. L. Murray and A. D. Lopez. 2000. "Estimates of DALE for 191 countries: methods and results". Global Programme on Evidence Discussion Paper No. 16. Geneva: World Health Organization.

Murray, Christopher. J. L. and Julio Frenk. 2000. "A framework for assessing the performance of health systems". *Bulletin of the World Health Organization*, 78 (6), 717-731.

World Health Organization. 1999. *The World Health Report 1999: Making a Difference*. Geneva: World Health Organization.

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


## VIII. Countries included in the sensitivity analysis

The following is a list of the countries included in the sensitivity analysis. We have also made available an Excel spreadsheet with all our data and estimates.

Algeria
Argentina
Australia
Austria
Bangladesh
Belgium
Benin
Bolivia
Botswana
Brazil
Bulgaria
Cameroon
Canada
Central African Republic
Chile
China
Colombia
Congo
Costa Rica
Denmark
Dominican Republic
Ecuador

Egypt
El Salvador
Finland
France
Gambia
Germany
Ghana
Greece
Guatemala
Guinea-Bissau
Haiti
Honduras
Hungary
India
Indonesia
Iran, Islamic Republic of
Iraq
Ireland
Israel
Italy
Jamaica
Japan
Jordan
Kenya
Kuwait
Lesotho
Liberia
Malawi
Malaysia
Mali
Mauritius
Mexico
Mozambique
Myanmar
Nepal
Netherlands
New Zealand
Nicaragua
Niger
Norway
Pakistan
Panama
Papua New Guinea
Paraguay
Peru
Philippines
Poland
Portugal

Republic of Korea
Romania
Rwanda
Senegal
Sierra Leone
Singapore
South Africa
Spain
Sri Lanka
Sudan
Sweden
Switzerland
Syrian Arab Republic
Thailand
Togo
Trinidad and Tobago
Tunisia
Turkey
Uganda
United Kingdom
United Republic of Tanzania
United States of America
Uruguay
Venezuela, Bolivarian Republic of
Yemen
Zambia
Zimbabwe

## IX. Log-file

```
. use geography

. reg lndale lnhexp lneduc lneduc2 tropics coastal

  Source |       SS       df       MS              Number of obs =      97
---------+------------------------------           F(  5,    91) =   50.81
   Model |  5.32999825     5  1.06599965           Prob > F      =  0.0000
Residual |   1.9090435    91    .0209785           R-squared     =  0.7363
---------+------------------------------           Adj R-squared =  0.7218
   Total |  7.23904175    96  .075406685           Root MSE      =  .14484

------------------------------------------------------------------------------
  lndale |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnhexp |   .0712409   .0219214      3.250   0.002       .0276967    .1147851
  lneduc |   .2705847   .0746808      3.623   0.000       .1222404     .418929
 lneduc2 |  -.0535683   .0325991     -1.643   0.104      -.1183224    .0111858
 tropics |  -.0817062   .0424749     -1.924   0.058      -.1660774     .002665
 coastal |   .1898304   .0450915      4.210   0.000       .1002618    .2793991
   _cons |   3.304257   .1063967     31.056   0.000       3.092913    3.515601
```

```
--------------------------------------------------------------------------------

. quietly sum tropics

. sca de avgtrop=r(mean)

. quietly sum coastal

. sca de avgcoast=r(mean)

. display avgland
.49038144

. gen dtrop=tropics-avgtrop

. gen dcoast=coastal-avgcoast

. gen geofe=_b[tropics]*dtrop+_b[coastal]*dcoast

. sum geofe

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
   geofe |      97   -8.27e-10   .0841384  -.1314902   .1400464

. save, replace
file geography.dta saved

. outsheet using geography.txt, replace

. log close
```