WHO’s Fooling Who?
The World Health Organization’s Problematic Ranking of Health Care Systems

by Glen Whitman

Executive Summary

The World Health Report 2000, prepared by the World Health Organization, presented performance rankings of 191 nations’ health care systems. These rankings have been widely cited in public debates about health care, particularly by those interested in reforming the U.S. health care system to resemble more closely those of other countries. Michael Moore, for instance, famously stated in his film Sicko that the United States placed only 37th in the WHO report. CNN.com, in verifying Moore’s claim, noted that France and Canada both placed in the top 10.

Those who cite the WHO rankings typically present them as an objective measure of the relative performance of national health care systems. They are not. The WHO rankings depend crucially on a number of underlying assumptions—some of them logically incoherent, some characterized by substantial uncertainty, and some rooted in ideological beliefs and values that not everyone shares.

The analysts behind the WHO rankings express the hope that their framework “will lay the basis for a shift from ideological discourse on health policy to a more empirical one.” Yet the WHO rankings themselves have a strong ideological component. They include factors that are arguably unrelated to actual health performance, some of which could even improve in response to worse health performance. Even setting those concerns aside, the rankings are still highly sensitive to both measurement error and assumptions about the relative importance of the components. And finally, the WHO rankings reflect implicit value judgments and lifestyle preferences that differ among individuals and across countries.

Glen Whitman is an associate professor of economics at California State University at Northridge.
Introduction

The World Health Report 2000, prepared by the World Health Organization, presented performance rankings of 191 nations’ health care systems.¹ Those rankings have been widely cited in public debates about health care, particularly by those interested in reforming the U.S. health care system to resemble more closely those of other countries. Michael Moore, for instance, famously stated in his film SiCKO that the United States placed only 37th in the WHO report. CNN.com, in verifying Moore’s claim, noted that France and Canada both placed in the top 10.²

Those who cite the WHO rankings typically present them as an objective measure of the relative performance of national health care systems. They are not. The WHO rankings depend crucially on a number of underlying assumptions—some of them logically incoherent, some characterized by substantial uncertainty, and some rooted in ideological beliefs and values that not everyone shares. Changes in those underlying assumptions can radically alter the rankings.

More Than One WHO Ranking

The first thing to realize about the WHO health care ranking system is that there is more than one. One ranking claims to measure “overall attainment” (OA) while another claims to measure “overall performance” (OP). These two indices are constructed from the same underlying data, but the OP index is adjusted to reflect a country’s performance relative to how well it theoretically could have performed (more about that adjustment later). When using the WHO rankings, one should specify which ranking is being used: OA or OP.

Many popular reports, however, do not specify the ranking used and some appear to have drawn from both. CNN.com, for example, reported that both Canada and France rank in the top 10, while the United States ranks 37th. There is no ranking for which both claims are true. Using OP, the United States does rank 37th. But while France is number 1 on OP, Canada is 30. Using OA, the United States ranks 15th, while France and Canada rank 6th and 7th, respectively. In neither ranking is the United States at 37 while both France and Canada are in the top 10.

Which ranking is preferable? WHO presents the OP ranking as its bottom line on health system performance, on the grounds that OP represents the efficiency of each country’s health system. But for reasons to be discussed below, the OP ranking is even more misleading than the OA ranking. This paper focuses mainly on the OA ranking; however, the main objections apply to both OP and OA.

Factors for Measuring the Quality of Health Care

The WHO health care rankings result from an index of health-related statistics. As with any index, it is important to consider how it was constructed, as the construction affects the results. WHO’s index is based on five factors, weighted as follows:³

1. Health Level: 25 percent
2. Health Distribution: 25 percent
3. Responsiveness: 12.5 percent
4. Responsiveness Distribution: 12.5 percent
5. Financial Fairness: 25 percent

The first and third factors have reasonably good justifications for inclusion in the index: Health Level. This factor can most justifiably be included because it is measured by a country’s disability-adjusted life expectancy (DALE). Of course, life expectancy can be affected by a wide variety of factors other than the health care system, such as poverty, geography, homicide rate, typical diet, tobacco use, and so on. Still, DALE is at least a direct measure of the health of a country’s residents, so its inclusion makes sense.
Responsiveness. This factor measures a variety of health care system features, including speed of service, protection of privacy, choice of doctors, and quality of amenities (e.g., clean hospital bed linens). Although those features may not directly contribute to longer life expectancy, people do consider them aspects of the quality of health care services, so there is a strong case for including them.

The other three factors, however, are problematic:

Financial Fairness. A health system’s financial fairness (FF) is measured by determining a household’s contribution to health expenditure as a percentage of household income (beyond subsistence), then looking at the dispersion of this percentage over all households. The wider the dispersion in the percentage of household income spent on health care, the worse a nation will perform on the FF factor and the overall index (other things being equal).

In the aggregate, poor people spend a larger percentage of income on health care than do the rich. Insofar as health care is regarded as a necessity, people can be expected to spend a decreasing fraction of their income on health care as their income increases. The same would be true of food, except that the rich tend to buy higher-quality food.

The FF factor is not an objective measure of health attainment, but rather reflects a value judgment that rich people should pay more for health care, even if they consume the same amount. This is a value judgment not applied to most other goods, even those regarded as necessities such as food and housing. Most people understand and accept that the poor will tend to spend a larger percentage of their income on these items.

More importantly, the FF factor, which accounts for one-fourth of each nation’s OA score, necessarily makes countries that rely on market incentives look inferior. The FF measure rewards nations that finance health care according to ability to pay, rather than according to actual consumption or willingness to pay. In most countries, a household’s tax burden is proportional to income, or progressive (i.e., taxes consume an increasing share of income as income rises). Thus, a nation’s FF score rises when the government shoulders more of the health spending burden, because more of the nation’s medical expenditures are financed according to ability to pay. In the extreme, if the government pays for all health care, then the distribution of the health-spending burden is exactly the same as the distribution of the tax burden. To use the existing WHO rankings to justify more government involvement in health care—such as via a single-payer health care system—is therefore to engage in circular reasoning because the rankings are designed in a manner that favors greater government involvement. If the WHO rankings are to be used to determine whether more government involvement in health care promotes better health outcomes, the FF factor should be excluded.

The ostensible reason for including FF in the health care performance index is to consider the possibility of people landing in dire financial straits because of their health needs. It is debatable whether the potential for destitution deserves inclusion in a strict measure of health performance per se. But even if it does, the FF factor does not actually measure exposure to risk of impoverishment. FF is calculated by (1) finding each household’s contribution to health expenditure as a percentage of household income (beyond subsistence), (2) cubing the difference between that percentage and the corresponding percentage for the average household, and (3) taking the sum of all such cubed differences. Consequently, the FF factor penalizes a country for each household that spends a larger-than-average percentage of its income on health care. But it also penalizes a country for each household that spends a smaller-than-average percentage of its income on health care.

Put more simply, the FF penalizes a country because some households are especially likely to become impoverished from health costs—but it also penalizes a country because some households are especially unlikely to become impoverished from health costs. In short, the FF factor can cause a country’s

To use the existing WHO rankings to justify more government involvement in health care is to engage in circular reasoning because the rankings are designed in a manner that favors greater government involvement.
rank to suffer because of desirable outcomes.

**Health Distribution and Responsiveness Distribution.** These two factors measure inequality in the other factors. Health Distribution measures inequality in health level within a country, while Responsiveness Distribution measures inequality in health responsiveness within a country.

Strictly speaking, neither of these factors measures health care performance, because inequality is distinct from quality of care. It is entirely possible to have a health care system characterized by both extensive inequality and good care for everyone. Suppose, for instance, that Country A has health responsiveness that is “excellent” for most citizens but merely “good” for some disadvantaged groups, while Country B has responsiveness that is uniformly “poor” for everyone. Country B would score higher than Country A in terms of responsiveness distribution, despite Country A having better responsiveness than Country B for even its worst-off citizens. The same point applies to the distribution of health level.

To put it another way, suppose that a country currently provides everyone the same quality of health care. And then suppose the quality of health care improves for half of the population, while remaining the same (not getting any worse) for the other half. This should be regarded as an unambiguous improvement: some people become better off, and no one is worse off. But in the WHO index, the effect is ambiguous. An improvement in average life expectancy would have a positive effect, while the increase in inequality would have a negative effect. In principle, the net effect could go either way.

There is good reason to account for the quality of care received by a country’s worst-off or poorest citizens. Yet the Health Distribution and Responsiveness Distribution factors do not do that.

in essence, give greater weight to the well-being of the worst off. Alternatively, a separate health performance index could be constructed for poor households or members of disadvantaged minorities. These approaches would surely have problems of their own, but they would at least be focused on the absolute level of health care quality, which should be the paramount concern.

**Uncertainty and Sensitivity Intervals**

The WHO rankings are based on statistics constructed in part from random samples. As a result, each rank has a margin of error. Media reports on the rankings routinely neglect to mention the margins of error, but the study behind the WHO rankings admirably includes an 80-percent uncertainty interval for each country. These intervals reveal a high degree of uncertainty associated with the ranking method.

Using the OA ranking, the U.S. rank could range anywhere from 7 to 24. By comparison, France could range from 3 to 11 and Canada from 4 to 14. The considerable overlap among these intervals, as shown in Figure 1, means one cannot say with great confidence that the United States does not do better in the OA ranking than France, Canada, and most other countries.

These intervals result only from errors associated with random sampling. They do not take into account differences that could result from different weightings of the five component factors discussed earlier. Given that discussion, the proper weight for three of these factors is arguably zero. The authors of the study did not calculate rankings on the basis of that weighting, but they did consider other possible factor weights to arrive at a sensitivity interval for each country’s rank.

It turns out that the U.S. rank is unusually sensitive to the choice of factor weights, as shown in Figure 2. The U.S. rank could range anywhere from 8 to 22, while Canada could range from 7 to 8 and France from 6 to 7.
These intervals depend on the range of weights considered and would therefore be larger if more factor weights were considered.

Furthermore, the rank resulting from any given factor weighting will itself have a margin of error resulting from random sampling. That means the two different sorts of intervals (uncertainty and sensitivity) ought to be considered jointly, resulting in even wider ranking intervals. The ranks as reported in the media, without corresponding intervals, grossly overstate the precision of the WHO study.

**Achievement versus Performance Ranking**

As noted earlier, the WHO report includes rankings based on two indices, OA and OP. The OP index, under which the U.S. rank is notably worse, is the WHO’s preferred measure. It is worth considering the process that is used to convert the OA index into the OP index.9

The purpose of the OA-to-OP conversion is to measure the efficiency of health care systems—that is, their ability to get desirable health outcomes relative to the level of expenditures or resources used. That is a sensible goal. The results of the OP ranking, however, are easily misinterpreted, or misrepresented, as simply measuring health outcomes irrespective of inputs. For instance, according to the WHO press release that accompanied the original report, “The U.S. health system spends a higher portion of its gross domestic product than any other country but ranks 37 out of 191 countries according to its performance, the report finds.”10 The implication is that the United States performs badly in the OP ranking despite its high expenditures—an implica-
tion that has also been drawn by various media outlets and commentators. A more accurate statement would be that the United States performs badly in the ranking because of its high expenditures, at least in part.

When Costa Rica ranks higher than the United States in the OP ranking (36 versus 37), that does not mean Costa Ricans get better health care than Americans. Americans most likely get better health care—just not as much better as could be expected given how much more America spends. If the question is health outcomes alone, without reference to how much has been spent, the more appropriate measure is the OA ranking, where the United States is 15 and Costa Rica is 45. (Even then, this paper’s earlier criticisms of the OA ranking still apply.)

The conversion of OA into OP depends on two constructed variables: first, the maximum level of performance a country could potentially achieve; and second, the minimum level of performance the country could achieve without a modern health care system. The maximum is estimated on the basis of a country’s per capita health expenditure and its level of literacy. The minimum is based on literacy alone. Literacy is used as a proxy for all aspects of a country that might affect health other than the health care system.

Many other variables could have been used to estimate a country’s minimum and maximum possible performance, such as average income, crime rate, geography, nutrition, and so on. None of these were included. But Dean Jamison and Martin Sandbu, in a 2001 Science article, reconstructed the OP ranking while including just one additional variable: geography. For 79 out of 96 countries for which Jamison and Sandbu were
able to recalculate ranks, the resulting rank fell outside—often far outside—the WHO report’s 80-percent uncertainty intervals for those ranks. In other words, inclusion of just one additional variable could drastically affect the resulting ranks. Inclusion of other variables could result in even greater deviations from the reported ranks. For this reason, the OP ranking is even more misleading than the OA ranking, which simply reports health outcomes without a spurious “efficiency” adjustment.

**Underlying Paternalistic Assumptions**

The WHO rankings, by purporting to measure the efficacy of health care systems, implicitly take all differences in health outcomes not explained by spending or literacy and attribute them entirely to health care system performance. Nothing else, from tobacco use to nutrition to sheer luck, is taken into account.

To some extent, the exclusion of other variables is simply the result of inadequacies in the data. It is difficult to get information on all relevant factors, and even more difficult to account for their expected effects on health. But some factors are deliberately excluded by the WHO analysis on the basis of paternalistic assumptions about the proper role of health systems. An earlier paper laying out the WHO methodological framework asserts, “Problems such as tobacco consumption, diet, and unsafe sexual activity must be included in an assessment of health system performance.”

In other words, the WHO approach holds health systems responsible not just for treating lung cancer, but for preventing smoking in the first place; not just for treating heart disease, but for getting people to exercise and lay off the fatty foods.

That approach is problematic for two primary reasons. First, it does not adequately account for factors that are simply beyond the control of a health system. If the culture has a predilection for unhealthy foods, there may be little health care providers can do about it. Conversely, if the culture has a pre-existing preference for healthy foods, the health care system hardly deserves the credit. (Notice the high rank of Japan, known for its healthy national diet.) And it hardly makes sense to hold the health system accountable for the homicide rate. Is it reasonable to consider the police force a branch of the health system?

Second, the WHO approach fails to consider people’s willingness to trade off health against other values. Some people are happy to give up a few potential months or even years of life in exchange for the pleasures of smoking, eating, having sex, playing sports, and so on. The WHO approach, rather than taking the public’s preferences as given, deems some preferences better than others (and then praises or blames the health system for them).

A superior (though still imperfect) approach would take people’s health-related behavior as given, and then ask which health systems do the best job of dealing with whatever health conditions arise. We could ask, for instance, which systems do the best job of treating cancer or heart disease patients. We could then rank nations according to disease-specific mortality rates or five-year survival rates. These approaches present challenges as well, as it can be difficult to control for all confounding factors. For example, better five-year survival rates may reflect earlier detection rather than better treatment or outcomes. Still, if the goal is to assess the efficacy of countries’ health care systems, it makes more sense to look at condition-specific success rates than indices (like the OA and OP) that fail to control for non–health-care factors like nutrition and lifestyle.

**Conclusion**

The analysts behind the WHO rankings express the hope that their framework “will lay the basis for a shift from ideological dis-
course on health policy to a more empirical one.\textsuperscript{16} Yet the WHO rankings themselves have a strong ideological component. They include factors that are arguably unrelated to actual health performance and some that could even improve in response to worse health performance. Even setting those concerns aside, the rankings are still highly sensitive to both measurement error and assumptions about the relative importance of the components. And finally, the WHO rankings reflect implicit value judgments and lifestyle preferences that differ among individuals and across countries. The WHO health care ranking system does not escape ideology. On the contrary, it advances ideological assumptions under the guise of objectivity. Those interested in objective measures of health system performance should look elsewhere.

Notes


4. Bear in mind that most nations finance the bulk of medical expenditures through health insurance, which results in a more uniform distribution of the burden of health spending.

5. To be precise, the FF measure uses the absolute value of the cubed difference, which means the value is always positive. Notice also that cubing puts an especially high weight on differences from the mean. Squaring differences is a much more common statistical approach to measuring dispersion. Cubing differences further reduces the scores of nations that rely less on government to finance medical care.

6. Rather than measuring inequality in DALE, Health Distribution measures inequality in infant mortality. Apparently, this change was made because of the better availability of data on differences in infant mortality.

7. Murray et al.

8. Though Murray et al. include a graphic showing the sensitivity intervals for different factor weights (their Figure 5, p. 8), they do not state specific bounds for those sensitivity intervals as they do for uncertainty intervals. Efforts to locate the data underlying that figure were unsuccessful. These estimates (and Figure 2 in this paper) represent the author’s best attempt to reproduce the intervals in Murray et al.


13. Geography may affect health because of climate effects. Jamison and Sandbu note that living in a tropical location appears to be associated with worse health outcomes (p. 1596).

14. Jamison and Sandbu were unable to obtain the data necessary to duplicate the WHO’s analysis for all 191 countries.

16. Ibid., p. 728.

---

**OTHER STUDIES IN THE BRIEFING PAPERS SERIES**

| 98. | *Doublespeak and the War on Terrorism* by Timothy Lynch (September 6, 2006) |
| 97. | *No Miracle in Massachusetts: Why Governor Romney’s Health Care Reform Won’t Work* by Michael Tanner (June 6, 2006) |
| 96. | *Free Speech and the 527 Prohibition* by Stephen M. Hoersting (April 3, 2006) |
| 93. | *Keep the Cap: Why a Tax Increase Will Not Save Social Security* by Michael Tanner (June 8, 2005) |
| 91. | *Medicare Prescription Drugs: Medical Necessity Meets Fiscal Insanity* by Joseph Antos and Jagadeesh Gokhale (February 9, 2005) |
| 90. | *Hydrogen’s Empty Environmental Promise* by Donald Anthrop (December 7, 2004) |
| 89. | *Caught Stealing: Debunking the Economic Case for D.C. Baseball* by Dennis Coates and Brad R. Humphreys (October 27, 2004) |
86. **School Choice in the District of Columbia: Saving Taxpayers Money, Increasing Opportunities for Children** by Casey J. Lartigue Jr. (September 19, 2003)

85. **Smallpox and Bioterrorism: Why the Plan to Protect the Nation Is Stalled and What to Do** by William J. Bicknell, M.D., and Kenneth D. Bloom (September 5, 2003)

84. **The Benefits of Campaign Spending** by John J. Coleman (September 4, 2003)


82. **Failing by a Wide Margin: Methods and Findings in the 2003 Social Security Trustees Report** by Andrew G. Biggs (April 22, 2003)


80. **States Face Fiscal Crunch after 1990s Spending Surge** by Chris Edwards, Stephen Moore, and Phil Kerpen (February 12, 2003)

79. **Is America Exporting Misguided Telecommunications Policy? The U.S.-Japan Telecom Trade Negotiations and Beyond** by Motohiro Tuschiya and Adam Thierer (January 7, 2003)

78. **This Is Reform? Predicting the Impact of the New Campaign Financing Regulations** by Patrick Basham (November 20, 2002)

77. **Corporate Accounting: Congress and FASB Ignore Business Realities** by T. J. Rodgers (October 25, 2002)

76. **Fat Cats and Thin Kittens: Are People Who Make Large Campaign Contributions Different?** by John McAdams and John C. Green (September 25, 2002)

75. **10 Reasons to Oppose Virginia Sales Tax Increases** by Chris Edwards and Peter Ferrara (September 18, 2002)

74. **Personal Accounts in a Down Market: How Recent Stock Market Declines Affect the Social Security Reform Debate** by Andrew Biggs (September 10, 2002)

73. **Campaign Finance Regulation: Lessons from Washington State** by Michael J. New (September 5, 2002)

72. **Did Enron Pillage California?** by Jerry Taylor and Peter VanDoren (August 22, 2002)

70. Farm Subsidies at Record Levels As Congress Considers New Farm Bill by Chris Edwards and Tad De Haven (October 18, 2001)

69. Watching You: Systematic Federal Surveillance of Ordinary Americans by Charlotte Twight (October 17, 2001)

68. The Failed Critique of Personal Accounts by Peter Ferrara (October 8, 2001)

67. Lessons from Vermont: 32-Year-Old Voucher Program Rebuts Critics by Libby Sternberg (September 10, 2001)

66. Lessons from Maine: Education Vouchers for Students since 1873 by Frank Heller (September 10, 2001)

65. Internet Privacy and Self-Regulation: Lessons from the Porn Wars by Tom W. Bell (August 9, 2001)

64. It’s the Spending, Stupid! Understanding Campaign Finance in the Big-Government Era by Patrick Basham (July 18, 2001)

63. A 10-Point Agenda for Comprehensive Telecom Reform by Adam D. Thierer (May 8, 2001)


61. Disparate Impact: Social Security and African Americans by Michael Tanner (February 5, 2001)

60. Public Opinion and Campaign Finance: A Skeptical Look at Senator McCain’s Claims by David M. Primo (January 31, 2001)


58. Will the Net Turn Car Dealers into Dinosaurs? State Limits on Auto Sales Online by Solveig Singleton (July 25, 2000)

57. Legislative Malpractice: Misdiagnosing Patients’ Rights by Greg Scandlen (April 7, 2000)

56. “We Own the Night”: Amadou Diallo’s Deadly Encounter with New York City’s Street Crimes Unit by Timothy Lynch (March 31, 2000)
55. The Archer-Shaw Social Security Plan: Laying the Groundwork for Another S&L Crisis by Andrew G. Biggs (February 16, 2000)
   (December 8, 1999)

54. Nameless in Cyberspace: Anonymity on the Internet by Jonathan D. Wallace
   (December 8, 1999)

53. The Case against a Tennessee Income Tax by Stephen Moore and Richard Vedder (November 1, 1999)

52. Too Big to Fail? Long-Term Capital Management and the Federal Reserve by Kevin Dowd (September 23, 1999)

51. Strong Cryptography: The Global Tide of Change by Arnold G. Reinhold
   (September 17, 1999)

50. Warrior Cops: The Ominous Growth of Paramilitarism in American Police Departments by Diane Cecilia Weber (August 26, 1999)

49. Mr. Smith, Welcome to Washington by Roger Pilon (July 30, 1999)

48. The U.S. Postal Service War on Private Mailboxes and Privacy Rights by Rick Merritt (July 30, 1999)

47. Social Security Reform Proposals: USAs, Clawbacks, and Other Add-Ons by Darcy Ann Olsen (June 11, 1999)

46. Speaking the Truth about Social Security Reform by Milton Friedman (April 12, 1999)


44. The Costs of Reducing Carbon Emissions: An Examination of Administration Forecasts by Peter VanDoren (March 11, 1999)

43. The Perils of Government Investing by Michael Tanner (December 1, 1998)


41. Term Limits and the Republican Congress: The Case Strengthens by Aaron Steelman (October 28, 1998)