

The Worst Study Ever?

Exposing the scandalous methods behind an extraordinarily influential 'World Health Report'

By Scott W. Atlas

THE WORLD HEALTH Organization's *World Health Report 2000*, which ranked the health-care systems of nearly 200 nations, stands as one of the most influential social-science studies in history. For the past decade, it has been the de facto basis for much of the discussion of the health-care system in the United States, routinely cited in public discourse by members of government and policy experts. Its most notorious finding—that the United States

ranked a disastrous 37th out of the world's 191 nations in "overall performance"—provided supporters of President Barack Obama's transformative health-care legislation with a data-driven argument for swift and drastic reform, particularly in light of the fact that the U.S. spends more on health than any other nation.

In October 2008, candidate Obama used the study to claim that "29 other countries have a higher life expectancy and 38 other nations have lower infant mortality rates." On June 15, 2009, as he was beginning to make the case for his health-care bill, the new president said: "As I think many of you are aware, for all of this spending, more of our citizens are uninsured, the quality of our care is often lower, and we aren't any healthier. In fact, citizens in some countries that spend

SCOTT W. ATLAS is a senior fellow at the Hoover Institution and professor of radiology and chief of neuro-radiology at the Stanford University Medical Center.

substantially less than we do are actually living longer than we do.” The perfect encapsulation of the study’s findings and assertions came in a September 9, 2009, editorial in Canada’s leading newspaper, the *Globe and Mail*: “With more than 40 million Americans lacking health insurance, another 25 million considered badly underinsured, and life expectancies and infant mortality rates significantly worse than those of most industrialized Western nations, the need for change seems obvious and pressing to some, especially

The study’s failings were plain from the outset and remain patently obvious; but they went unnoticed, unmentioned, and unexamined because *World Health Report 2000* was so politically useful. This object lesson in the ideological misuse of politicized statistics should serve as a cautionary tale.

when the United States is spending 16 percent of GDP on health care, roughly twice the average of other modern developed nations, all of which have some form of publicly funded system.”

In fact, *World Health Report 2000* was an intellectual fraud of historic consequence—a profoundly deceptive document that is only marginally a measure of health-care performance at all. The report’s true achievement was to rank countries according to their alignment with a specific political and economic ideal—socialized medicine—and then claim it was an objective measure of “quality.”

WHO researchers divided aspects of health care into subjective categories and tailored the definitions to suit their political aims. They allowed fundamental flaws in methodology, large margins of error in data, and overt bias in data analysis, and then offered conclusions despite enormous gaps in the data they did have. The flaws in the report’s approach, flaws that thoroughly undermine the legitimacy of the WHO rankings, have been repeatedly exposed in peer-reviewed literature by academic experts who have examined the study in detail. Their analysis made clear that the study’s failings were plain from the outset and remain patently obvious today; but they went unnoticed, unmentioned, and unexamined by many because *World Health Report 2000* was so politically

useful. This object lesson in the ideological misuse of politicized statistics should serve as a cautionary tale for all policymakers and all lay people who are inclined to accept on faith the results reported in studies by prestigious international bodies.

Before WHO released the study, it was commonly accepted that health care in countries with socialized medicine was problematic. But the study showed that countries with nationally centralized health-care systems were the world’s best. As Vincente Navarro noted in 2000 in the highly respected *Lancet*, countries like Spain and Italy “rarely were considered models of efficiency or effectiveness before” the WHO report. Polls had shown, in fact, that Italy’s citizens were more displeased with their health care than were citizens of any other major European country; the second worst was Spain. But in *World Health Report 2000*, Italy and Spain were ranked #2 and #7 in the global list of best overall providers.

Most studies of global health care before it concentrated on health-care outcomes. But that was not the approach of the WHO report. It sought not to measure performance but something else. “In the past decade or so there has been a gradual shift of vision towards what WHO calls the ‘new universalism,’” WHO authors wrote, “respecting the ethical principle that it may be necessary and efficient to ration services.”

The report went on to argue, even insist, that “governments need to promote community rating (i.e. each member of the community pays the same premium), a common benefit package and portability of benefits among insurance schemes.” For “middle income countries,” the authors asserted, “the policy route to fair prepaid systems is through strengthening the often substantial, mandatory, income-based and risk-based insurance schemes.” It is a curious version of objective study design and data analysis to assume the validity of a concept like “the new universalism” and then to define policies that implement it as proof of that validity.

The nature of the enterprise came more fully into view with WHO’s introduction and explanation of the five weighted factors that made up its index. Those factors are “Health Level,” which made up 25 percent of “overall care”; “Health Distribution,” which made up another 25 percent; “Responsiveness,” accounting

for 12.5 percent; "Responsiveness Distribution," at 12.5 percent; and "Financial Fairness," at 25 percent.

The definitions of each factor reveal the ways in which scientific objectivity was a secondary consideration at best. What is "Responsiveness," for example? WHO defined it in part by calculating a nation's "respect for persons." How could it possibly quantify such a subjective notion? It did so through calculations of even more vague subconditions—"respect for dignity," "confidentiality," and "autonomy."

And "respect for persons" constituted only 50 percent of a nation's overall "responsiveness." The other half came from calculating the country's "client orientation." That vague category was determined in turn by measurements of "prompt attention," "quality of amenities," "access to social support networks," and "choice of provider."

Scratch the surface a little and you find that "responsiveness" was largely a catchall phrase for the supposedly unequal distribution of health-care resources. "Since poor people may expect less than rich people, and be more satisfied with unresponsive services," the authors wrote, "measures of responsiveness should correct for these differences."

Correction, it turns out, was the goal. "The object is not to explain what each country or health system has attained," the authors declared, "so much as to form an estimate of what should be possible." They appointed themselves determinants of what "should" be possible "using information from many countries but with a specific value for each country." This was not so much a matter of assessing care but of determining what care *should* be in a given country, based on WHO's own priorities regarding the allocation of national resources. The WHO report went further and judged that "many countries are falling far short of their potential, and most are making inadequate efforts in terms of responsiveness and fairness."

CONSIDER the discussion of Financial Fairness (which made up 25 percent of a nation's score). "The way health care is financed is perfectly fair if," the study declared, "the ratio of total health contribution to total non-food spending is identical for all households, independently of their income, their health status or their use of the health system." In plain language, higher earners should pay more for health care, period. And people who become sick, even if that illness is due to high-risk behavior, should not pay more. According to WHO, "Financial fairness is best served by more, as well as

by more progressive, prepayment in place of out-of-pocket expenditure. And the latter should be small not only in the aggregate but relative to households' ability to pay."

This matter-of-fact endorsement of wealth redistribution and centralized administration should have had nothing to do with WHO's assessment of the actual *quality* of health care under different systems. But instead, it was used as the *definition* of quality. For the authors of the study, the policy recommendation preceded the research. Automatically, this pushed capitalist countries that rely more on market incentives to the bottom of the list and rewarded countries that finance health care by centralized government-controlled single-payer systems. In fact, two of the major index factors, Health Distribution and Responsiveness Distribution, did not even measure health care itself. They were both strictly measures of equal *distribution* of health and equal *distribution* of health-care delivery.

Perhaps what is most striking about the categories that make up the index is how WHO weighted them. Health Distribution, Responsiveness Distribution, and Financial Fairness added up to 62.5 percent of a country's health-care score. Thus, almost two-thirds of the study was an assessment of *equal-*

A matter-of-fact endorsement of wealth redistribution and centralized administration should have had nothing to do with WHO's assessment of the actual quality of health care. For the authors of the study, the policy recommendation preceded the research.

ity. The actual health outcomes of a nation, which logic dictates should be of greatest importance in any health-care index, accounted for only 25 percent of the weighting. In other words, the WHO study was dominated by concerns outside the realm of health care.

Not content with penalizing free-market economies on the fairness front, the WHO study actually held a nation's health-care system accountable for the behavior of its citizens. "Problems such as tobacco consumption, diet, and unsafe sexual activity must be

included in an assessment of health system performance," WHO declared. But the inclusion of such problems is impossible to justify scientifically. For example, WHO considered tobacco consumption equivalent, as an indicator of medical care, to the treatment of measles: "Avoidable deaths and illness from childbirth, measles, malaria or tobacco consumption can properly be laid" at a nation's health-care door."

From a political standpoint, of course, the inclusion of behaviors such as smoking is completely logical. As Samuel H. Preston and Jessica Ho of the University of Pennsylvania observed in a 2009 Population Studies Center working paper, a "health-care system could be performing exceptionally well in identifying and administering treatment for various diseases, but a country could still have poor measured health if personal health-care practices were unusually deleterious." This takes on additional significance when one considers that the United States has "the highest level of cigarette consumption per capita in the developed world over a 50-year period ending in the mid-80s."

At its most egregious, the report abandoned the very pretense of assessing health care. WHO ranked the U.S. 42nd in life expectancy. In their book, *The Business of Health*, Robert L. Ohsfeldt and John E. Schneider of the University of Iowa demonstrated that this finding was a gross misrepresentation. WHO actually included immediate deaths from murder or fatal high-speed motor-vehicle accidents in their assessment, as if an ideal health-care system could turn back time to undo car crashes and prevent homicides. Ohsfeldt and Schneider did their own life-expectancy calculations using nations of the Organisation for Economic Co-operation and Development (OECD). With fatal car crashes and murders included, the U.S. ranked 19 out of 29 in life expectancy; with both removed, the U.S. had the world's best life-expectancy numbers (see table above).

But even if you dismiss all that, the unreliability of *World Health Report 2000* becomes inarguable once you confront the sources of the data used.

NATIONAL LIFE EXPECTANCY: 1980 TO 1999 (with and without fatal injuries)

RANKING	ACTUAL MEAN LIFE EXPECTANCY (INCLUDING FATAL INJURIES)		STANDARDIZED MEAN LIFE EXPECTANCY (WITHOUT FATAL INJURIES)	
	OECD NATIONS		OECD NATIONS	
1.	Japan	78.7	United States	76.9
2.	Iceland	78.0	Switzerland	76.6
3.	Sweden	77.7	Norway	76.3
4.	Switzerland	77.6	Canada	76.2
5.	Canada	77.3	Iceland	76.1
6.	Spain	77.3	Sweden	76.1
7.	Greece	77.1	Germany	76.1
8.	Netherlands	77.0	Denmark	76.1
9.	Norway	77.0	Japan	76.0
10.	Australia	76.8	Australia	76.0
11.	Italy	76.6	France	76.0
12.	France	76.6	Belgium	76.0
13.	Belgium	75.7	Austria	76.0
14.	United Kingdom	75.6	Netherlands	75.9
15.	Germany	75.4	Italy	75.8
16.	Finland	75.4	United Kingdom	75.7
17.	New Zealand	75.4	Finland	75.7
18.	Austria	75.3	New Zealand	75.4
19.	United States	75.3	Czech Republic	75.1
20.	Denmark	75.1	Ireland	75.0
21.	Ireland	74.8	Spain	74.9
22.	Portugal	73.9	Slovak Republic	74.4
23.	Czech Republic	72.2	Greece	74.4
24.	Slovak Republic	71.6	Portugal	74.3
25.	Poland	71.5	Hungary	74.3
26.	Korea	71.1	Korea	73.3
27.	Mexico	70.9	Poland	73.2
28.	Hungary	69.7	Mexico	72.8
29.	Turkey	64.4	Turkey	72.0

In the study, WHO acknowledged that it "adjusted scores for overall responsiveness, as well as a measure of fairness based on the informants' views as to which groups are most often discriminated against in a country's population and on how large those groups are" [emphasis added]. A second survey of about 1,000 "informants" generated opinions about the relative importance of the factors in the index, which were then used to calculate an overall score.

Judgments about what constituted "high quality" or "low quality" health care, as well as the effect of inequality, were made by people WHO called "key

informants.” Astonishingly, WHO provided no details about who these key informants were or how they were selected. According to a 2001 *Lancet* article by Celia Almeida, half the responders were members of the WHO staff. Many others were people who had gone to the WHO website and were then invited to fill out the questionnaire, a clear invitation to political and ideological manipulation.

Another problem emerges in regard to the references used by the report. Of the cited 32 methodological references, 26 were from internal WHO documents that had not gone through a peer-review process. Moreover, only two were written by people whose names did not appear among the authors of *World Health Report 2000*. To sum up: the report featured data and studies largely generated inside WHO, with no independent, peer-reviewed verification of the findings. Even these most basic requirements of valid research were not met.

The report’s margin of error is similarly ludicrous in scientific terms. The margin for error in its data falls outside any respectable form of reporting. For example, its data for any given country were “estimated to have an 80 percent probability of falling within the uncertainty interval, with chances of 10 percent each of falling below the low value or above the high one.” Thus, as Whitman noted, in one category—the “overall attainment” index—the U.S. could actually rank anywhere from seventh to 24th. Such a wide variation renders the category itself meaningless and comparisons with other countries invalid.

And then there is the plain fact that much of the necessary data to determine a nation’s health-care performance were simply missing. The WHO report stated that data was used “to calculate measures of attainment for the countries where information could be obtained...to estimate values when particular numbers were judged unreliable, and to estimate attainment and performance for all other Member States.”

According to a shocking 2003 *Lancet* article by Philip Musgrove, who served as editor in chief of part of the WHO study, “the attainment values in WHO’s *World Health Report 2000* are spurious.” By his calculation, the WHO “overall attainment index” was actually generated by complete information from only 35 of the 191 countries. Indeed, according to Musgrove, WHO

had data from only 56 countries for Health Inequality, a subcategory; from only 30 countries for Responsiveness; and from only 21 countries for Fair Financing. Nonetheless, rankings were “calculated” for all 191 countries.

Musgrove gives specific examples of overtly deficient data that was directly misused by WHO. He stated that “three values obtained from expert informants (for Chile, Mexico, and Sri Lanka) were discarded in favor of imputed (i.e. estimated) values” and that “in two cases, informants gave opinions on one province or state rather than the entire country.” Musgrove wrote to Christopher Murray, WHO’s director of the Global Program on Evidence for Health Policy at the time, on August 30, 2000, about the study’s handling of the missing statistics: “If it doesn’t qualify as manipulating the data, I don’t know what does...at the very least, it gravely undermines the claim to be honest with the data and to report what we actually find.”

IF *World Health Report 2000* had simply been issued and forgotten, it would still be a case study in how to produce a wretched and unreliable piece of social science masquerading as legitimate research. That it served so effectively as a catalyst for unprecedented legislation is evidence of something more disturbing. The executive and legislative branches of

The executive and legislative branches used WHO’s document as an implicit Exhibit A to justify imposing radical changes to the health-care system, even in the face of objections from the American people. To blur the line between politics and objective analysis is to do violence to both.

the United States government used WHO’s document as an implicit Exhibit A to justify imposing radical changes to America’s health-care system, even in the face of objections from the American people. To blur the line between politics and objective analysis is to do violence to them both.

Despite the compelling studies that undermine this erroneous document, many government officials, policymakers, insurers, and academics have continued to use it to justify their ideology-based agenda, one that

seeks centralized, government-run health care. Donald Berwick, later up for the position of Obama's director of the Center for Medicare & Medicaid, used its questionable data to make a grand case against the U.S. health-care system in 2008: "Even though U.S. health-care expenditures are far higher than those of other developed countries, our results are no better. Despite spending on health care being nearly double that of the next most costly nation, the United States ranks thirty-first among nations on life expectancy, thirty-sixth on infant mortality, twenty-eighth on male healthy life expectancy, and twenty-ninth on female healthy life expectancy." (This last bit came from updated 2006 WHO data.)

What we have here is a prime example of the misuse of social science and the conversion of statistics from pseudo-data into propaganda. The basic principle, casually referred to as "garbage in, garbage out," is widely accepted by all researchers as a cautionary dictum. To the authors of *World Health Report 2000*, it functioned as its opposite—a method to justify a preconceived agenda. The shame is that so many people, including leaders in whom we must repose our trust and whom we expect to make informed decisions based on the best and most complete data, made such blatant use of its patently false and overtly politicized claims. ▶