

**Benchmarks of Fairness for Health Care Reform:
A Policy Tool for Developing Countries**

N. Daniels,¹ J. Bryant,² R.A.Castano,³ O.G. Dantes,⁴ K.S. Khan,⁵ S. Pannarunothai⁶

1. A New Tool for Policy Analysis:

We report here on progress we have made toward developing the benchmarks of fairness¹ into a policy tool that will be useful in developing countries for analyzing the overall *fairness* of health care reforms.

Fairness is a multi-dimensional concept, broader than the concept of equity.^{2,3,4} Fairness includes equity in health outcomes, in access to all forms of care, and in financing. Fairness also includes efficiency in management and allocation, since, when resources are constrained, their inefficient use means that some needs will not be met that could have been. For the public to be empowered to assure that its health is promoted, fairness must also include accountability. Finally, fairness also includes appropriate forms of patient and provider autonomy. The benchmarks facilitate an integrated examination of objectives that often involve trade-offs with each other. It requires looking across disciplinary boundaries in a systematic way.

When originally developed and presented in the United States, the benchmarks were provided with an ethical rationale that appealed to a theory of justice and health

¹ Professor, Dept. of Philosophy, Tufts University, Medford MA 02468 (email: ndaniels@emerald.tufts.edu).

² President, CIOMS, Geneva; Emeritus Professor of Community Health Sciences, Aga Khan University, Karachi, Pakistan.

³ Ministry of Health, Colombia (though not as a formal representative of the Ministry).

⁴ National Institute of Public Health, Cuernavaca

⁵ Associate Professor, Community Health Sciences, Aga Khan University, Karachi

care.^{5,1} The central idea in that theory is that disease and disability impair the range of opportunities open to individuals, and that a principle governing equality of opportunity provides a basis for regulating a health care system. The same theory can be extended to look beyond the point of delivery of health care to the social determinants of health as well.⁶

The objection might be raised that this liberal democratic, rights-based account is too culturally limited to provide an international framework for the benchmark approach. Nevertheless, in our work in four developing country sites, which differ among themselves quite considerably in political, cultural and religious backgrounds, we found great convergence on the benchmarks themselves without extensive discussion of an underlying ethical framework. Participants were introduced to the equal opportunity theory, but it played no explicit role in producing the agreement on the benchmarks, and there was no discussion of it in any detail. Because of our focus on fairness, we also avoided some culturally sensitive issues, such as abortion, euthanasia, and the use of human and fetal tissues or organs that have been highly controversial.

We did discuss the fact that the weight or priority given to different benchmarks might vary in different countries depending on some cultural beliefs. For example, more or less weight might be given to accountability or to patient and provider autonomy or even to the equity considerations underlying Benchmarks 1 and 2. But in our workshops, these variations were not significant. In any case, we deliberately refrained from weighting the benchmarks in a uniform way across countries. The primary feature worth nothing is the overwhelming agreement we reached though deliberation about the practical components of a fair reform.

⁶ Faculty of Medicine, Naresuan, Thailand

The benchmarks have relevance because there is rapid reform of health care systems around the world as a result of changes in economic and political systems, economic growth, or previous failures to meet population needs. External agencies have played a large role in offering incentives to undertake privatizing and decentralizing reforms. In all these contexts, however, reforms are usually debated without a systematic evaluation of their impact on the *fairness* of the resulting system. Privatizing and decentralizing efforts, for example, may aim at adding new resources and circumventing inefficient bureaucracies. The private sector, however, often competes with and weakens the public one, and it requires strong and efficient regulation if it is not to undermine equity. Though efficiency is included within fairness, promoting some kinds of efficiency without attention to other dimensions of fairness will not improve fairness and may undercut it. The benchmarks provide a framework for evaluating comprehensively the effects on fairness of these strategies and others.

The goal of the benchmarks is to force deliberation about the specific, interacting effects of the reforms being compared, not simply to produce a “report card” with numerical “grades.” Consequently, it is essential that a *rationale*, containing reasons and evidence, be provided for the score that is assigned to a proposal on each relevant criterion. These rationales provide the necessary objectivity needed to prompt thoughtful deliberation. Rationales might not be needed if we had included only criteria that were measurable magnitudes, such as the proportion of the population receiving some particular service or having some particular health status. Many salient and critical components, mechanisms, and processes contributing to fairness, such as those involved in accountability, are not so directly measurable, and satisfaction of criteria for them requires judgment. By insisting on rationales, subjectivity in these judgments is reduced..

Because the benchmarks map scores onto rationales, the tool can serve its purpose of promoting deliberation about fairness. Since the goal is to facilitate cross-disciplinary evaluation and deliberation, an analytic framework for thinking about diverse and competing values is more important than the semblance of rigorous measurement.

We begin with a short history of the benchmarks approach, comment briefly on the benchmarks and their scoring, note some preliminary findings from their use, and conclude by explaining how the benchmarks supplement, rather than compete with, alternative ways of measuring equity and indexing health system performance.

2. History of the Benchmarks Approach

The original "benchmarks of fairness" were developed to assess and promote deliberation about comprehensive medical insurance reforms proposed in the United States in the first Clinton administration.^{1,7,8} These benchmarks focused heavily on desirable features needed in the reform of a technologically advanced but inefficient and inequitable system that lacked universal coverage. Despite this specific focus, the original benchmarks addressed basic questions that must be asked about *any* reform: does it reduce financial and nonfinancial barriers to access to public health measures and medical services exist? does it promote health care services appropriate to the needs of the population? does it distribute the burdens of paying for health protection fairly? does the reform promote clinical and administrative efficiency, so that health budgets produce value for money? does it make institutions publicly accountable for the decisions they make? how does it affect the choices people can exercise?

The benchmarks connect these quite general questions to specific operational criteria and measures. Reforms are then compared by scoring them against these criteria and by providing rationales for the scores. The rationales provide an objective basis for

deliberation. For example, the original benchmark on financial barriers ranked one reform higher than another if it closed more of the insurance gap or if it better provided for portability of insurance when workers changed jobs. This technique of evaluation is generalizable to reforms in other countries by expanding the criteria to include elements of system design crucial elsewhere.⁷

To adapt the benchmarks for use in health systems in countries at different levels of development, teams of collaborators from four countries, Colombia, Mexico, Pakistan, and Thailand were formed. During 1999, these teams held two week-long workshops in Cuernavaca (combining the Colombian and Mexican teams), Bangkok, and Karachi, with representation from each Asian site participating in the other Asian workshops. Members of the country teams had varying backgrounds, including: faculty members of interested universities, representatives of donor agencies supporting health care reform, members of health services research teams working on reform options, and persons involved in policy making at the national level.

Teams used each country as a “case study” for which appropriate benchmarks were developed. By successively reviewing the work of previous workshops across sites, the teams produced a modifiable schema of benchmarks appropriate to all of them as a final product.

In each workshop, the process included these steps: 1) seminar presentations and discussion about the salient problems facing each system, including a history and critical evaluation of recent reform efforts; 2) a seminar presentation and discussion about the original benchmarks and how they had been applied to U.S. reform efforts; 3) a discussion of whether new benchmarks were needed to address local issues that were not addressed by the original set, or by the provisional set developed by previous workshops;

4) a critical review and revision of each of the original benchmarks, or of the results provided by the preceding workshops; 5) an attempt to link the detailed discussion of system problems and reforms to specific criteria for each benchmark; 6) “testing” (including field testing in Thailand)¹⁰ of the provisional benchmarks by using them to score actual and proposed reforms in each country; 7) refinement and revision of the criteria in light of these scoring attempts; 8) development of specific plans for disseminating the benchmarks for actual use in each site. The Asian workshops included field trips to villages and urban slums to examine the delivery system and provide first-hand experience of the problems requiring reform

3. The Revised Benchmarks

There are nine benchmarks, each of which contains various operational criteria for evaluating specific aspects of the fairness of reform proposals (see Appendix). We highlight key features of each benchmark.

Benchmark 1, Intersectoral Public Health The rationale for this benchmark is that social determinants^{6,11} and other risk factors “upstream” from the point of health care delivery affect population health and its distribution. The first criterion in Benchmark 1 asks for estimates of the degree to which a demographically differentiated population incurs improvements in exposure to various risk factors as a result of the reforms under consideration. Obviously, not all reforms will touch on all or even many of these factors, but the comprehensive list is included because reforms would make a system more fair if they did eliminate inequalities in exposure to these factors. Where good information on these exposures does not exist, the criterion encourages gathering it.

The second criterion calls for developing an information infrastructure needed to measure and monitor health inequalities and to carry out research about the most

effective ways to reduce them. The third criterion evaluates reforms for their intersectoral focus and their involvement of communities and vulnerable groups in these efforts.

Country specific differences in problems and organization mean that intersectoral efforts must vary; use of the benchmarks requires country-specific adjustments. For example, it may be crucial to focus on violence reduction or accident reduction in some countries, and on clean water or other factors in others.

Benchmark 2, Financial Barriers to Access: Fairness requires reducing financial and non-financial barriers to access to needed services. Benchmark 2 recognizes the large “informal,” untaxable employment sector in many developing countries, often including sixty to ninety per cent of the population. Since workers and their families in the informal sector generally include the poorest part of the population, needed services must be provided in full or in large part through general tax revenues. The larger the informal sector, the larger the need for public financing, but the smaller the tax base to meet it

Benchmark 2 encourages a long term, intersectoral strategy aimed at moving as much of the population as possible into the formal sector, and then into insurance schemes that can be built on broadly based general tax revenues, social security payments, or employer-based contributions.

Benchmark 2 also specifies interim goals in both sectors. Because public resources are so scarce in the informal sector, a crucial issue is whether the most important services are available to all. Benchmark 2 encourages reforms to specify a basic package of services that all will receive by a specific target date, then to improve that package over time. For example, the 1995 Mexican reforms, funded by external loans, aim to provide universal access to a very modest package of services. By 1999,

over 90% of the population has access to them, and when 100% is reached in the near future, the Mexican government is obliged to finance this universal but modest package itself. In Colombia, the 1993 reforms aimed at a more comprehensive benefit package for the informal sector; the new constitution, however, created legal pressure, deriving from a right to life, to expand those benefits. It has not been possible with existing resources in Colombia to deliver that package universally. Neither reform, then, would meet fully the criteria specified in the benchmark.

In Thailand, the debate continues about whether to implement a defined minimum benefit package proposed in recent reforms, or whether to continue to rely on a type of public insurance (“Type B”) that allows providers discretion to negotiate what kinds of services will be available to those without any insurance. As a result of scoring Thai proposals using Benchmark 2, a specific research question emerged about the levels of unmet need in this population. Research on that issue should inform policy deliberation.

In Pakistan, the informal sector includes ninety per cent of the population. In *theory*, all people have access to an increasingly robust set of services, depending on the type of facility visited. In *reality*, many services, including drugs that are officially available, turn out not to be available in practice for various reasons (e.g., the existence of shadow providers, or the drug supply is not well prioritized or adequately funded), driving people to seek care from private sources. In scoring reforms, explicit attention is paid to the gap between intention and implementation.

Benchmark 2 concentrates on two goals of reform for the formal sector besides increasing the size of the sector: producing uniform and more adequate benefits across all groups of workers in it, and integrating the various schemes that involve these workers. In Thailand, for example, the long range reform plans call for considerable integration of

formal sector insurance plans through district fundholding and regulative controls, and eventual expansion of coverage to all family members, many of whom, except for dependents of civil service workers, are now not covered. In Pakistan, with only ten per cent of workers in the formal sector, the team focused on the need for developing from the start a plan that would lead to a well-integrated formal sector scheme, not a hodge-podge of private plans with little regulation or equity.

Benchmark 3, Non-financial Barriers to Access. The first criterion evaluates reforms for the measures they take to address the maldistribution of drugs, supplies, facilities and personnel common in all four countries. Where the reform relies on local fund-holding and decentralization, the criteria also examine specific goals and accountability for them (see Benchmark 8 as well).

The second criterion addresses gender barriers, which are especially important barriers to primary care in Pakistan, for example in the squatter slums of Karachi, where studies of children at high risk for death from diarrheal disease and pneumonia suggest that lack of maternal autonomy is a key risk factor. The benchmarks emphasize involving community political groups as an essential way to address these barriers, since simply providing services will not overcome them.

Two other criteria address other cultural and discriminatory barriers. The details of these problems will vary from country to country. The benchmarks revealed that too little explicit attention was paid to this issue in recent reforms in Mexico and Pakistan.

Benchmark 4, Comprehensiveness of Benefits and Tiering: The underlying rationale is that all people, regardless of class or ethnicity or gender, have comparable health needs and there are similar social obligations to meet them. Inequalities in the comprehensiveness and quality of care (“tiering”), especially where these have impacts

on health outcomes, violate equity constraints on system design. Some kinds of tiering are worse than others. It is less serious if a small but wealthy group does better than others, provided the others do well (e.g., private sector of insurance in the United Kingdom) than if a poor group is also given worse health benefits than the rest of society (e.g., failing to insure the working poor in the United States, or failing to deliver a minimal benefit package to the whole informal sector while the top 5% has excellent private insurance, as in Colombia). Some tiering is also unavoidable in systems with severe resource constraints and a large informal sector.

All teams focused on extensive differential treatment of people by class within a system, not only between the public and private sectors but within the public sector. Residents of Sultanabad, a squatter slum of Karachi, remarked that “the tradesman will do better than the laborer in a public hospital,” suggesting a widespread perception of tiering in the system, where the poor commonly wait four to five hours to be seen in a hospital, then to get five minutes with the doctor, whereas well-to-do patients can just walk into private sector services and be seen right away. Tiering exists in the benefit packages available to different subgroups in the formal sector in Thailand, Pakistan, Colombia and Mexico, as well. In Thailand, for example, civil service workers will have better access to hemodialysis than other formal sector workers. In Mexico and Pakistan, some multinational employers provide better coverage than the social security schemes, and the military in Pakistan has the best coverage of any group.

Benchmark 5, Equitable Financing, rests on the fundamental idea that financing medical services, as opposed to access to them, should be according to ability to pay. Three main funding streams are involved in most systems: tax-based revenues, premiums for insurance, and out-of-pocket payments. The benchmark divides the problem primarily

between tax-based and premium-based parts of the system, noting that in both contexts there is still out-of-pocket payments for care. Tax-based schemes are more equitable if they are more progressive in their structure. Premium-based schemes are more equitable if they are community-rated, rather than risk-rated. Risk-rating shifts the burden to those at higher risk of illness. The same inequity is involved in out-of-pocket contributions in both tax-based and premium-based systems. A good measure of progressivity must combine all financing streams.¹²

The substantial out-of-pocket costs for health care in all four collaborating sites was the main source of regressivity in financing and the main way of shifting burdens to the sick, rather than pooling them across the whole population. There are many pressures on systems to rely on and even increase cash payments for services.

Benchmark 6, Efficacy, Efficiency, and Quality of Care: The rationale for this and the next benchmark is that, other things equal, a system that gets more value for money in the use of its resources is fairer to those in need. (If resources were not limited, a lavish, inefficient system might still meet all (health care) needs (if not all others), but distributive justice and fairness arise as issues in the real world of resource limitations.) A key criterion in Benchmark 6 focuses on primary health care for community-based delivery. Reforms aimed at improving primary care must assure appropriate training, incentives, resource allocation, and community participation in the decisions affecting delivery. (The importance of public health measures is already captured in Benchmark 1.) Emphasis was placed on a population focus and on the need for the integration of different parts of the health system, such as referrals, and integration with other sectors. Community participation ideally involves an interactive relationship that goes beyond mere “outreach.”

The second main concern of Benchmark 6 is promoting evidence-based practice in all areas of services, including preventive, curative, and management practices. To advance evidence-based practice, the criteria call for the development of an information infrastructure and data base, as well as for health services research to support such evidence-based practice. The third main criterion concerns measures to improve quality of services in the system, including professional training, continuing education, credentialing and accreditation, and community participation in evaluating the quality of care.

In all systems we examined, there are problems with referral mechanisms and with the role of primary care gatekeepers. Dissatisfaction with primary care services leads many people to jump to higher level hospitals for primary care, leading to considerable inefficiency. Similarly, there is no control of efficacy or quality since people will often abandon the public sector primary care services for completely unregulated private sector services. Establishing good referral systems is a critical element in the efficiency of care, but the restrictions such systems involve also conflict with the kinds of choice or autonomy that are captured by Benchmark 9. To justify restrictions on autonomy, there must be high quality practitioners doing the diagnosis and referral, clear, accessible routes to higher levels of care, and education of all about the importance of such a system.

Benchmark 7, Administrative Efficiency, seeks efficiency in the management of the health care system. Each country team emphasized its own examples of inefficient management. Addressing these problems, however, also requires greater accountability, including transparency, in these systems; consequently, Benchmark 8 must work together with Benchmark 7 if real improvement is to result.

The criteria included in Benchmark 7 were constructed out of consideration of many examples for all four collaborating sites of sources of administrative inefficiency. Key areas of common concern, were various sources of administrative overhead (inappropriate technology acquisition, inefficient use of personnel, high transaction costs), costly forms of purchasing, cost shifting, and many types of abuse and fraud (shadow providers, drug sales and auto-referrals, inappropriate promotion of drugs and devices).

Some general points emerge that cut across the differences in the countries. In all of them, public sector practitioners receive very low pay, and this fact creates a context for many of the forms of abuse that create efficiency problems noted in this benchmark and accountability problems noted in the next one. The failure to have integrated financing schemes means that there are incentives to shift costs from one part of the system to others. In Thailand, where unions are weak, civil service work rules prevent efficiencies of manpower allocation. In Latin America, stronger unions and their work rules provide the same obstacle to reallocation of personnel.

In the public sector of all the systems, a common set of complaints is articulated: bureaucratic practices and corruption lead to great inefficiencies in the purchase of supplies and equipment, failures to enforce rules about personnel, favoritism and hiring on grounds other than competency, and other highly inefficient practices. In all these contexts, there is talk about “decentralization” as a solution, but decentralization is not a panacea unless there is careful planning and regulation to make sure decentralized units are aiming at similar goals.

In thinking about scoring reforms, Benchmarks 6 - 8 should play a key role in helping to think through the content of measures, such as the decentralization of public

bureaucracies and the establishment of district or other level budgeting of various revenue flows. The integrated Thai reform plan, and decentralization proposals in Pakistan, must be evaluated for the specific ways in which they achieve the goals highlighted in these benchmarks; the benchmarks aim at avoiding being lulled by a fashionable label or idea into ignoring the specifics of reform proposals.

Benchmark 8, Democratic Accountability and Empowerment, emerged in all four countries as critically important, since without these forms of accountability, reforms are unlikely to succeed in other areas as well. The rationale for including accountability is that health systems are responsible for the improvement of population health in an equitable manner, and those affected by decisions and policies that affect well-being in such fundamental ways must have an understanding of and ultimate control over that system. Such control is not exercisable without accountability for reasonableness^{13,14} in decision-making about allocation and other matters. Such accountability includes transparency, including that provided by global budgeting, fair appeals processes, adequate privacy protection, and measures to enforce compliance with rules and laws. None of the criteria are ultimately effective without a strengthening of civil society, so that people understand the problems of the health sector and are empowered to seek improvements in it.

One important criterion, originally proposed in the Latin American workshops, evaluates reform for its attempt to stimulate the growth of advocacy groups, clearly an intersectoral matter. This criterion is important because of the crucial role such groups play, in countries with developed democratic traditions, of compelling public authorities to rectify problems in both public and private sectors. In Pakistan and Thailand, this idea was expanded into the criterion “strengthening civil society” which now has two

components, establishing an enabling environment for advocacy groups and stimulating public debate about health policy measures. Many aspects of this benchmark go beyond merely holding institutions in the health sector accountable to the public; they actually increase the power of the public to act to remedy problems.

Benchmark 9, Patient and Provider Autonomy, is the benchmark that most directly addresses a culturally variable issue. How important is autonomy or choice? In some market-based approaches, informed choice is necessary if quality is to be improved and true preferences met. But how much choice, and what kinds of choices? Similarly, provider autonomy is much sought by professionals, but that is often seen by many planners as an obstacle to efficient use of services, since professionals and provider institutions are influenced by incentives to utilize what they can supply.

For these reasons, it is important to emphasize how Benchmark 9 may conflict with other benchmarks (and others with each other) and that reasonable people in the same or different cultures may disagree about weightings. Consequently there may be no one fairest system, but many fair designs. Benchmarks allow for cultural and other variation, but encourage maximum deliberation about grounds for designs that value some benchmarks over others.

A clear example of the conflict between Benchmark 9 and other benchmarks involves referral systems and the restrictions on patients they involve. Benchmark 6, for example, may approve of restrictions on autonomy in order to achieve a primary care focus and the efficiency that results from letting primary care physicians filter access to other levels of care, but Benchmark 9 is driven by the concern for a loss of choice. Similarly, choice of alternative providers will undermine efficiency and quality if there is no adequate credentialing and evidence-based assessment of alternative forms of

treatment. Practitioner autonomy may be essential if the practitioner is to address the health care problems of individual patients, but this presupposes high levels of competency and knowledge of appropriate evidence-based, population-oriented practices as a baseline.

4. Scoring and Uses of the Benchmarks

We have adapted the benchmarks for use in evaluating competing reform proposals within a country, and the discussion of scoring focuses on that use. It may be possible to use the benchmarks to make some international comparisons of fairness across systems,¹⁴ but we have ignored such an application in the work reported on here. If our goal had been to provide inter-country comparisons, it would have been necessary to validate the scale in a way that is not crucial for evaluating reforms. In evaluating reforms, progress is made if people can agree on what they think the current limitation of the system is on a particular dimension and then agree about the degree to which a specific reform would improve or worsen that aspect of the system. Disagreements about these aspects of scoring will make the deliberative process about the merits of reforms better, which is the ultimate goal of the benchmarks anyway. It is crucial to understand this purpose of scoring in order to see why our adaptation has adopted a particular approach to it.

In the original use of the benchmarks to evaluate competing U.S. reforms, a scoring system was adopted that took the status quo as a “0,” assigned a maximal positive outcome a “5,” and maximal regression from the status quo a “-5.” Though the U.S. authors emphasized that these numbers reflect ordinal rankings on complex dimensions, not cardinal measures of some quantity, numbers invite confusion. Our Latin American teams therefore used ordinal notations (0 to five “pluses” or “minuses”) to emphasize that the heuristic of scoring was primarily aimed at forcing the clear

presentation of rationales. The Asian teams were comfortable with the convenience of numbers, carefully explaining that they were ordinal, and emphasizing the centrality of rationales. All agreed on the primary point: the scoring exercise is a heuristic device aimed at generating clear rationales, and we agreed to leave the choice of symbols to country teams using the tool.

The point of this method of scoring is to see how well particular reform proposals, as compared to alternatives, fare on the many aspects of fairness captured by the benchmarks. Some proposals will be stronger on some dimensions than others, but weaker on others. Where these do not represent true tradeoffs, it may be possible to formulate policies that are true improvements overall. Where tradeoffs are being made, because of the requirement that rationales accompany all scoring, the framework stimulates discussion of the competing values underlying the alternatives.

5. Some Preliminary Findings

Our scoring exercises in Colombia, Mexico, Pakistan, and Thailand showed that the adapted benchmarks could reveal: a) places at which proposed reforms were insufficiently detailed or vague about mechanisms to reveal their effects; b) problematic assumptions about how goals of reform would be achieved; c) empirical issues that would have to be resolved in order to determine the likely success of implementing a reform. As a result, we were able in both Pakistan and Thailand to construct practical lists of such issues to be brought before groups considering the proposals for implementation. We also noted the large gap that often loomed between the intention underlying a reform and the results of implementation. We were able to highlight this gap by scoring the both the intention and implementation of a proposal, where we had evidence about implementation.

In Thailand, the benchmarks were “field tested” by asking people to use them for evaluation of national reforms under consideration (and partly implemented) as well as for changes at the provincial level over a two year period. Results are reported fully elsewhere.¹⁰ This exercise has led to the proposal that the benchmarks be deployed for use in evaluating current national proposals for system reform and for use in evaluating plans made by provincial health officers, who will have more autonomy under proposed reforms. In Pakistan, plans exist to incorporate the benchmarks into training programs for provincial and district health officers, as well as into medical school curricula. Plans also exist to have a public health network of academic centers promote the use of the benchmarks at national and provincial levels. There are more ambitious plans to involve regional WHO organizations in the broader adaptation and dissemination of the tool.

All participants in the four sites agreed that a useful format for presenting the final product will be an interactive computer program that allows policy analysts and broader community groups to draw on a database of similar reforms and their outcomes. Properly designed, such a tool would allow concentration on selected benchmarks with the option of ignoring those less relevant. This flexibility would allow it to be used at different levels within a system -- not just for comprehensive national reforms, but for more specific reforms at the provincial or district level. PAHO has already expressed interest in posting such a tool on its website as one item in a policy “toolbox (We note that such a program would have a much more specific function than Michael Reich’s Policy Maker; Policy Maker provides analytic techniques for evaluating the implementability of any kind of policy, but it lacks the detailed framework for assessing fairness that is included in the benchmarks. The approaches supplement but do not replace each other.)

6. The Benchmarks vs Other Measures of Equity and Health System Performance

By way of conclusion, we emphasize that the benchmarks supplement or complement, rather than compete with, various other efforts to monitor equity in health systems or to index health system performance across countries. Consider, for example, the WHO-sponsored efforts to develop measures for monitoring health inequities across demographic groups and for setting goals and targets for reducing these inequities.⁴ Some new approaches to measuring health inequalities may better highlight subgroup differences.^{16,17} Some of these measures could be incorporated into the benchmark approach; in addition, since setting targets requires evaluating how reforms will affect a system, the benchmarks will prove a useful supplement to such an approach.

The benchmarks, it should be noted, refrain from passing judgment on an important issue of distributive justice that is connected with some efforts at measuring population health and its distribution. It is commonly noted that strategies that aim at reducing the aggregate burden of disease in a population sometimes conflict with strategies aimed at reducing inequalities in health status. On the assumption that reasonable people may disagree about how to resolve such conflicts, the benchmarks refrain from making an overall judgment on this issue and instead insist on fair procedures for making them within a country. Specifically, Benchmark 8 requires that reforms put into place procedures for making resource allocation decisions in a way that is transparent and publicly accountable to deliberative democratic processes.

As noted earlier, the benchmarks attempt no uniform scaling of fairness across systems. Instead, we adopt the status quo as a baseline for purposes of evaluating intra-country reforms. Suppose, however, that WHO develops an index that measures health

system performance across countries that includes health status, responsiveness, and fairness in financing.¹⁸ Such an index would clearly add something to the benchmarks, for it would help focus attention on areas where reform was clearly needed. At the same time, such an index would only supplement, not replace, the use of the benchmarks in evaluating intra-country reforms, for any country stimulated by the cross-country index to undertake further reforms would still benefit from the many-dimensional tool for evaluating reforms that we propose.

To date, the benchmarks have been used in only a preliminary way to evaluate reform proposals or recent reforms. Supasit and Samrit (2000) report on field tests in which the benchmarks were used to evaluate proposed national reforms and recent provincial reforms that establish the benchmarks can be used for scoring reforms in Thailand. Teams in Colombia, Mexico, and Pakistan have also carried out scoring exercises aimed at testing the usefulness of the specific criteria. In each setting, the benchmarks have shown that they can stimulate deliberation about the mechanisms through which proposed reforms operate, force greater specification of reform measures, and help to frame research questions that can bring evidence to bear on choices. A fuller evaluation of the approach must await pending efforts to deploy the benchmarks more widely.

Acknowledgment: The Rockefeller Foundation and its Global Health Equity Initiative funded our project; we thank the National Institute of Public Health in Mexico, the National Institute of Health, in Thailand, and the Community Health Sciences Department of Aga Khan University for hosting our workshops. We also wish to thank the members of the collaborating country teams and our hosts at the various site visits we made. Daniels also was supported by a Robert Wood Johnson Investigator Award.

APPENDIX

Benchmark 1: Intersectoral Public Health

- I. Degree to which reform increases percent of population (demographically differentiated where relevant and possible) receiving the following:
 - Basic Nutrition
 - Housing
 - crowding
 - homelessness
 - physical adequacy
 - Environmental Factors
 - Clean water (and water treatment)
 - Sanitation (vector control)
 - Clean air
 - Reduced exposure to workplace and environmental toxins
 - Education and Health Education
 - Literacy
 - Basic education
 - Health Literacy
 - Nutritional education
 - Sex education and promotion
 - Substance abuse education and promotion
 - Anti-smoking
 - Anti-drug and alcohol abuse
 - Public Safety and Violence Reduction
 - Vehicular accident reduction
 - Violence reduction (homicide, rape)
 - Domestic abuse (women, children)
- II. Development of information infrastructure for monitoring health status inequalities
 - Provision for regular measurement of health status inequalities, using appropriate indicators
 - Research into interventions most likely to reduce health status inequalities
- III. Degree to which reform has actively engaged intersectoral efforts at local, regional, and/or national level to improve social determinants of health, and the degree to which vulnerable groups have been involved in defining these efforts.

Benchmark 2: Financial Barriers to Equitable Access

- I. Nonformal Sector Coverage
 - Universal access to the most appropriate package of basic services, and improvement of packages over time
 - Examples of packages of varying scope:
 - Mexican 12 interventions (a minimal package)
 - PAHO's Primary Care package (a slightly more extensive package)
 - Colombia's Basic Benefic Package/Subsidized Regimen or Thai package
 - Catastrophic coverage

(unclear just where Pakistan package fits, but probably below Colombia, through public facilities).

Drug Coverage

Medical Transportation Costs

Portability of coverage (geographical, employment status)

II. Insurance for Formal Sector

-- encourages moving populations from informal to formal sector

Reduction of the following obstacles to enrolling people in the formal sector:

corruption and enforcement-- of tax requirements, mandatory enrollment

worker resistance to enrollment

small employer resistance

Family coverage for enrolled workers

Drug coverage

Medical Transportation Costs

Producing uniform benefits across all groups of workers

Integrating various schemes involving those workers

Benchmark 3: Nonfinancial Barriers to Access

I. Reduction in geographical mal-distribution of:

Facilities and services

Personnel (mix and training)

Supplies

Drugs

Clinic hours (appropriate to village routines, work schedules)

Transportation for medical purposes

II. Gender

Status in family regarding decision-making

Mobility

Access to resources

Reproductive autonomy

Gender sensitive provision of services, involvement of community political groups to address gender barriers

III. Cultural

Language

Attitude and practices relevant to disease and health

Uninformed reliance on untrained traditional practitioners (such as some healers, midwives, dentists, pharmacists)

Perception of public sector quality

IV. Discrimination by race, religion, class, sexual orientation, disease, including

1. stigmatization of groups receiving public care.

Benchmark 4: Comprehensiveness of Benefits and Tiering

I. All effective and needed services deemed affordable, by all needed providers

No categorical exclusions

- II. Reform reduces tiering and achieves more uniform quality
 - Integrates services to the poor and others

Benchmark 5: Equitable Financing

- I. Is financing by ability to pay?
 - If tax based-scheme
 - How progressive (by population subgroup)?
 - How much reliance on cash payments (by subgroup)?
 - If premium-based
 - Is it community-rated (by subgroup)?
 - Reliance on cash payments (by subgroup)?
 - Out-of-Pocket payments contribute to both
 - Main source of shifting burdens to the sick

Benchmark 6: Efficacy, Efficiency and Quality of Health Care

- I. Primary Health Care Focus
 - PHC training for community based delivery
 - Population-based
 - Community participation
 - Integration with rest of system (referrals)
 - Intersectoral integration (social and environmental determinants)
 - Incentives
 - Appropriate allocation of resources to PHC
 - Interactive community participation, including vulnerable subgroups
 - Referral mechanisms
 - Primary health care gatekeepers
 - By-passing primary health care sites
 - Respect for autonomy
- II. Implementation of Evidence-based Practice
 - Health policies
 - Public health and clinical prevention
 - Therapeutic interventions
 - Incentives for clinical guidelines
 - Evidence-based evaluation of methods for managing utilization of services
 - Information infrastructure and database
 - Evidence-based research on clinical and public health measures
 - Health services research on patterns of care
 - Population health needs and utilization rates, including variation studies
 - (with demographic differentiation)
- III. Measures to improve quality
 - Regular assessment of quality, including satisfaction, with surveys or community group involvement as appropriate
 - Accreditation of Plans and Hospitals
 - Professional Training
 - Curriculum focused on fair design of system

Continuing education

Benchmark 7: Administrative Efficiency

- I. Minimize administrative overhead
 - Inappropriate technology acquisition
 - Purchase
 - Maintenance
 - Training
 - Excessive marketing costs (hospitals or plans)
 - Efficient use of personnel
 - Reduction of excess
 - Appointments and promotions based on competency
 - Appropriate economies of scale
 - Adequate risk pools for insurers
 - Reduction of duplicate structures, including integration of vertically organized programs
 - Minimize transaction costs
 - Enrollment and disenrollment costs
 - Inappropriate shifting and transfers of personnel or patients
 - Minimize loss of needed personnel from system as a whole (brain drain)
 - Oversupply of some services in some areas
- II. Cost-reducing purchasing
 - Reduce price variation
 - Drug cost reduction through large scale purchasing
 - Reliance on (quality) generics where possible
- III. Minimize cost shifting
 - Cost shifting from PC to Tertiary
 - Cost shifting to patients
 - Cost shifting to public sector or insurance from other types
 - Cost shifting between schemes
- IV. Minimize abuse and fraud and inappropriate incentives
 - Shadow providers, partial and total
 - Practitioner auto-referral
 - Drug sales at profit by rural doctor
 - Billing practices
 - Un-credentialed practitioners in rural areas (also a problem in urban areas in Karachi)
 - Vehicles and other perks
 - Inappropriate promotion of drugs and devices
 - Appropriation of public resources for private practice

Benchmark 8: Democratic Accountability and Empowerment

- I. Explicit, public, detailed procedures for evaluating services with full public reports
 - Use reports
 - Performance reports
 - Compliance reports

- Use of adequately qualified consultants
- II. Explicit deliberative procedures for resource allocation with transparency and rationales for decisions based on reasons all “fair minded people” can agree are relevant
- III. Global budgeting
- IV. Fair grievance procedures
 - Legal procedures (malpractice)
 - Non legal dispute resolution procedures
- V. Adequate privacy protection
- VI. Measures for enforcement of compliance with rules and laws
- VII. Strengthening civil society
 - Enabling environment for advocacy groups
 - Stimulating public debate, including participation of vulnerable groups

Benchmark 9: Patient and Provider Autonomy

- I. Degree of consumer choice
 - Of primary care providers
 - Of specialty care providers
 - Of alternative providers
 - Of procedures
- II. Degree of practitioner autonomy

References

1. Daniels N, Light D, Caplan R. *Benchmarks of Fairness for Health Care Reform*. New York, Oxford University Press, 1996.
2. Whitehead, M. The concepts and principles of equity and health. *International Journal of Health Services*, 1992, **22**: 429-45.
3. Dahlgren G, and Whitehead M. *Policies and strategies to promote social equity in health* . Institute of Future Studies, Stockholm, 1992.
4. Braveman P. *Monitoring equity in health: a policy-oriented approach in low- and middleincome countries*. World Health Organization, Geneva, 1999.
5. Daniels N. *Just Health Care*. New York, Cambridge University Press, 1985.
6. Daniels N, Kennedy B, Kawachi I. Why justice is good for our health: the social determinants of health inequalities. *Daedalus* 1999, **128**:4:215-52.
7. Brock D, and Daniels N. Ethical foundations of the Clinton administration's proposed health care system. *JAMA*, 1994, **271**: 1189-1196.
8. Daniels N. *Seeking Fair Treatment: From the AIDS Epidemic to National Health Care Reform*. New York, Oxford University Press, 1995.
9. Daniels N, and Bryant J. Parametros de justicia y monitoreo de la equidad: Apoyo a un programa de los OMS. *Salud y Gerencia*, 1998,**16**(enero-junio):1,7-12.
10. Pannarunothai S, Srithamrongsawat S. Benchmarks of fairness for health system reform: the tool for national and provincial health development in Thailand. *Human resources for health development journal*, 2000 (April) (in press).
11. Sen A. Health in reflection. *Bulletin of the World Health Organization*, 1999, **77**:619-23.
12. Kakwani N. Equity in health in Thailand. Bangkok, March 1999 (unpublished ms).
13. Daniels N, Sabin J. Limits to health care: fair procedures, democratic deliberation, and the legitimacy problem for insurers. *Philosophy & public affairs*, 1997, **26**: 303-50.
14. Daniels N, Sabin J. The ethics of accountability in managed care reform. *Health affairs*, 17:5: 50-64.
15. Caplan RL, Light DW, Daniels N. Benchmarks of fairness: a moral framework for assessing equity. *International journal of health services*, 1999, **29**: 853-69.
16. Gakidou EE, Murray CJL, Frenk J. Defining and measuring health inequality. *Bulletin of the World Health Organization*, 2000, **78**: 42-54,

17. Gwatkin DR. Health inequalities and the health of the poor. *Bulletin of the World Health Organization*, 2000, **78**:3-18.

18. The Secretariat, *Trends and Challenges in World Health*, 2000, Geneva: World Health Organization.