

Indicators: how much complexity is desirable?

Indicadores: quanta complexidade é desejável?

Philip Musgrove

Principal Economist

Health, Nutrition and Population

The World Bank

Room G-7-015

1818 H Street NW

Washington, DC 20433

pmusgrove@worldbank.org

Paper prepared for Round-Table VI. Indicators: Methodological Aspects. V Brazilian Epidemiology Congress, Curitiba, PR, Brazil. March 23-27, 2002.

The opinions expressed here are entirely the author's, and do not necessarily represent the views of the World Bank, its Executive Directors, or the countries they represent.

Introduction

A great many specific topics could be treated under the heading of “methodological aspects of indicators” – purposes, definitions, data needs, methods of calculation or estimation, accuracy, sensitivity, specificity, potential uses, and so on. This paper approaches the subject not through any of these topics, but by looking at several indicators and trying to draw some conclusions from them about how to make up good indicators for moderately complex situations. That is, all the indicators considered involve at least two measurable variables; and they are sometimes built up from other indicators which are already generally accepted as adequately representing simpler concepts.

All the examples chosen, except the first, are ones with which I have had some personal involvement, in making them up, criticizing them, or both. I think some useful lessons can be drawn from these cases – lessons about what to do, or in some cases, what not to do. The principal lessons concern the right degree of complexity of an indicator – how many variables to base it on, how to relate them to one another, and how to assure that the result is understood by whoever needs to understand it. This is particularly important when the indicator is to be used directly to determine the allocation of resources, or indirectly to influence such decisions.

A non-epidemiological example: the “wind-chill index”

Anyone who has ever stepped outside on a cold, windy day knows that how cold he or she will feel does not depend only on the external temperature but also on how fast the wind is blowing. The question is how does that effect vary with wind velocity and with temperature? A formula for what might be called the “effective temperature”, but has actually come to be called the “wind chill” would contribute to public safety by reducing the risk of frostbite on exposed skin. The U. S. National Weather service introduced on 1 November 2001 a new version of the wind

chill index, replacing one that has been in use since 1945¹. The formula for the new index is

$$\text{Wind Chill} = 35.74 + 0.6215T - 35.75(V^{*0.16}) + 0.4275T(V^{*0.16})$$

Where **T** is the temperature in degrees Fahrenheit and **V** is the wind velocity in miles per hour, both measured or estimated at a height of five feet above the ground. Figure 1 shows the effective temperature for measured temperatures from 40 F (4.4 C) down to 45 below zero (- 43 C), and wind speeds from calm air (under 3 mph or 5 kph) to 60 mph (100 kph), which is nearly hurricane force. The chart also shows how long it would take for exposed skin to become frostbitten, which is the maximum safe time out of doors.

Converting to metric units changes the numerical parameters but not the way the two variables are related.

As with many other indicators, this one required specification of the way the component variables are to be measured (units, and height above ground). What is more important is that the index is built on a model of the human face – the area most exposed to frostbite – using modern theory about how

heat is transferred out of the body. This is the key improvement on the old index, which was based on the rapidity with which water in exposed containers freezes, under different combinations of temperature and wind speed. This theoretical modeling is what leads to the interaction of the two variables and to the exponent of 0.16 on the velocity. The effect of the wind is greater, the lower the temperature is; and at a given temperature, the effective temperature falls less rapidly than the wind speed increases because there is an upper limit to how fast heat is lost.

This example has nothing to do with epidemiology, which would be relevant only for estimating the incidence of hypothermia or frostbite. Nonetheless this case suggests some more general properties of good indicators. First, it has to be clear what one is trying to measure. What matters is not how cold it *is*, but how cold it *feels* to a human being. The old index, without the theoretical modeling that went into the revised version, gave a very different notion of how the effective temperature fell as wind velocity increased (Figure 2), and even showed an implausible *warming* after the speed passed 55 mph (88 kph). Second, even this carefully worked-out

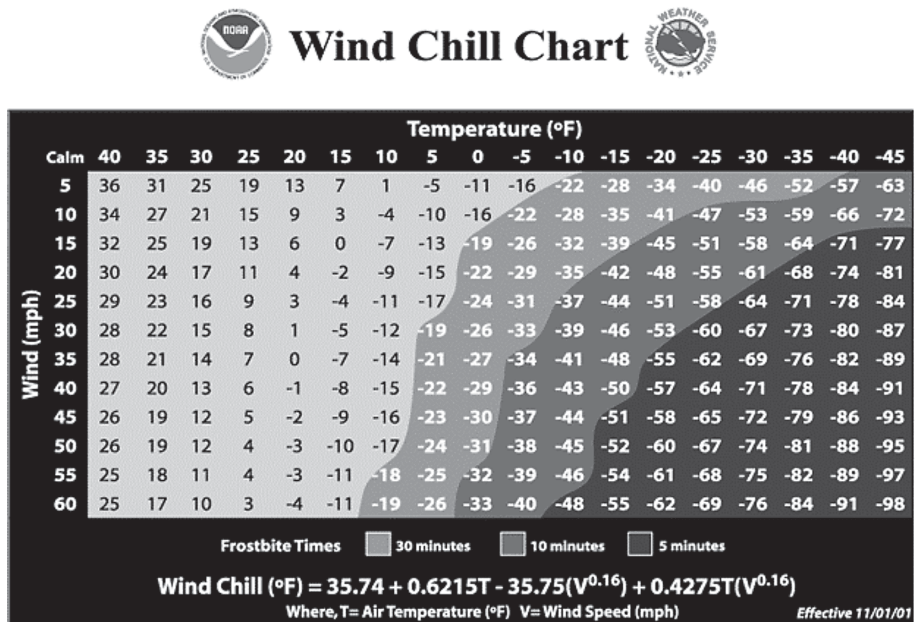


Figure 1 - How Temperature and Wind Velocity Determine the Wind Chill.

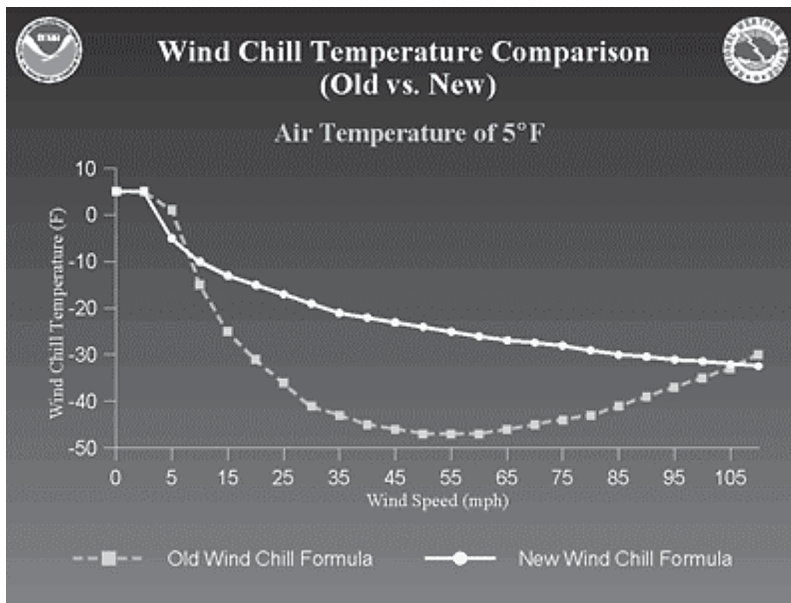


Figure 2 - It Matters How Wind Chill is Calculated

measure misses some important effects, such as those due to warming by sunlight, so further refinements may be undertaken. This may improve the accuracy of the indicator, at some cost to how easily it can be understood by someone deciding how warmly to dress; certainly the inclusion of a sunlight variable in the indicator will require making it as easy to understand as temperature and wind speed are. Third, there is no ideology, no extra-scientific assumption, in the construction of the wind chill indicator. This seems so obviously desirable as hardly to need mentioning, but indicators in the health field are not always free of assumptions that have little theoretical or evidentiary foundation. Finally, just because two or more variables are involved does not mean the only way to relate them is to add them together. The wind chill index includes a term linear in temperature, but not in velocity, and there is also an interactive term. For this particular indicator, linearity would lead to overestimation of the danger of frostbite at high wind velocities and would not take account of how the effect of the wind depends on the temperature. Both the question of ideology and that of linearity arise in the next example to be considered.

Indicator overload: trying to measure “health needs”

Just as it seems desirable to have an indicator of the total health situation of a population, involving a number of criteria and some difficult issues of judgment^{2,3}, it might seem helpful to have an aggregate indicator of the “health needs” of that population—that is, some way of summarizing not only how healthy or sick the population is, but what capacity the society has to take care of or improve its health and how well that capacity is being used. Some such reasoning must have appealed to the Pan American Health Organization in 1988, because when it published an “analysis of basic indicators”⁴, the technical document included an attempt to construct an “indicator of health needs”. Unfortunately, the attempt was a failure; and if the consultants hired for this task were not actually charlatans, at least they did not know what they were doing. There are many things wrong with the indicator they produced⁵, but three in particular are relevant to this discussion.

First, the index included no fewer than 20 different variables, grouped in five classes. This is simply overload, and makes it all but impossible to grasp intuitively what is being

measured. Some of the variables are purely demographic – fertility, growth rate of population, and shares of population under 15 or over 60; some are measures of health status – life expectancy at birth, infant and child mortality rates, and fraction of newborns who are underweight; others refer to health service utilization – immunization rates and hospital discharges; still others count health-related capital – hospital beds, doctors, dentists, nurses and auxiliaries per capita; and some refer to economic variables – income per head, health expenditure as a share of GDP and the share of health spending by the central government. It can certainly be argued that many or all of these variables would be relevant to decisions, say, about whether a country needed or could use more foreign assistance in health, which is currently the subject of considerable debate⁶. But that would require some conceptual or theoretical basis for how to relate needs as a function of one variable, to the levels of other variables. If immunization rates in a country are low because spending is low and there are too few trained personnel, the implications are very different from the case when there is no shortage of money or of nurses and auxiliaries and yet children go unprotected.

In the absence of any such model relating the variables, the index was constructed by linear combination. The complete formula – actually the first principal component to emerge from the analysis is:

$$\begin{aligned} \text{Index of needs} = & -0.09239(\text{Population} < 15) + \\ & 0.09221(\text{Population} > 60) - 0.07477(\text{Popula-} \\ & \text{tion growth rate}) - 0.08918(\text{Fertility rate}) + \\ & 0.06780(\text{Literacy rate}) + 0.08330(\text{Life Expect-} \\ & \text{ancy}) - 0.07413(\text{Infant Mortality}) - 0.06569 \\ & (\text{Child Mortality}) - 0.04743(\text{Low Birth-weight}) \\ & + 0.08231(\text{Beds per capita}) + 0.05923(\text{Doc-} \\ & \text{tors per capita}) + 0.05933(\text{Dentists per capita}) \\ & + 0.08943(\text{Nurses per capita}) + 0.08090(\text{Aux-} \\ & \text{iliaries per capita}) + 0.07542(\text{Hospital Dis-} \\ & \text{charges per capita}) + 0.04327(\text{Polio Immuni-} \\ & \text{zation}) + 0.05733(\text{DPT Immunization}) + \\ & 0.08397(\text{GNP per capita}) + 0.03225(\text{Health} \\ & \text{spending as \% of GNP}) + 0.08121(\text{Central} \\ & \text{Government \% of health spending}) \end{aligned}$$

Second, a linear formulation implies that if more of a variable is good when there is little of it, it continues to be equally valuable when there is a lot of it. That is a reasonable assumption for immunization rates, say, but not for many of the other variables. Saturation is ruled out, and so is any recognition that the amount needed of one resource or activity may depend on how much there is available of another one. If doctors are good, more doctors are always better, even if there are not enough nurses to work with them, or not enough money to pay them adequately. Similarly, if it is considered a burden to have too many children, or too many elderly people, the ideal population structure necessarily consists only of working-age adults. And if spending more on health is desirable, that remains true even at levels of expenditure which would actually reduce a country's income and take resources away from other health-promoting uses. This is fundamentally a conceptual issue, but it also matters, when concocting any indicator, to have a clear understanding of the basic mathematics involved. There is probably no better example anywhere of the dangers to which the assumption of linearity – seemingly natural or innocuous in many circumstances – can lead.

In the third place, most of the variables used in this index are positive rather than normative; they measure something without making strong assumptions about what *ought* to be the case. One variable however is a marked exception: that is the share of health spending that passes through the central government. There is no basis for the ideological supposition that central government expenditure is universally better or worse than spending by other levels of government or the private sector. The assumption of linearity, and the positive sign in the equation, imply that health needs are least, or the country's situation is best, when the central government spends *all* the health money. Given everything else that was wrong with the index, this assumption makes little difference, but it illustrates the need to keep one's ideological views or hypotheses separate from the indicators that might be used

to test whether they are sound. It is a different matter to argue that the *worst* kind of health spending is out of pocket, and that public expenditure can protect against the impoverishing effects of such spending, especially when most people cannot afford meaningful private insurance^{7,8}. But even on that view, it does not follow that all public spending should come from the central government.

The point of this discussion is not to criticize PAHO for being misled into a nonsensical exercise. It is only to illustrate the dangers of trying to make complicated indicators for ill-defined notions, particularly without adequate conceptual and mathematical underpinning. Stringing so many variables together is almost sure to lead to an unsound or unintelligible outcome, even if every variable by itself is a valid indicator of something.

Probabilities as indicators: judging equity by inequalities

There are other ways to link variables than by linear combination (or even non-linear and interactive formulas, as with the wind-chill index). Sometimes the variables correspond to a natural progression of events or possibilities, so that indicators can be chosen for each stage. The next example takes this approach, which avoids the need to find a single indicator corresponding to a complex notion – in this case, the notion of equity in health.

In a simplified model of an episode of illness or injury, a person may or may not benefit from a preventive action **I** (immunization, for instance); may or may not become sick **S** or injured; and if sick or hurt,

may or may not receive treatment **T**. In the absence of treatment, he or she may or may not recover **R**; with treatment, he or she may or may not be cured **C**. If each of these events is regarded as all-or-nothing, abstracting from the severity of illness and from incomplete or inadequate treatment, cure and recovery, then each stage can be characterized by a probability. Equity may be considered a state of equal probability for everyone – equal chances of being protected, being sick, getting treated, and so on. (The probability that someone will get treatment despite *not* being sick or otherwise needing intervention, is not considered.) This will be an unconditional probability for **I**, and for all but the last of the subsequent stages, a probability conditional on what happened at the previous stage. The person's final state of health **H** then depends on the whole sequence of probabilities, as shown in Figure 3, modified from⁹ by the addition of the preventive stage **I**. Underlining indicates that the event in question does *not* happen; the probability of that is 1.0 minus the probability that it *does* happen. Arrows indicate the transitions from one stage to the next.

Good health status at the end of this sequence comes from not having been sick or hurt; from recovering without treatment, if sick; or from getting treatment that cures the problems or restores health. The probability of that outcome depends on the *effectiveness* of the preventive action, $P(\underline{\mathbf{S}}/\underline{\mathbf{I}})$ – $P(\mathbf{S}/\underline{\mathbf{I}})$, as well as on the *likelihood* or coverage of that action, $P(\underline{\mathbf{I}})$. Similarly, it depends on the effectiveness of treatment, $P(\underline{\mathbf{C}}/\underline{\mathbf{T}})$ – $P(\mathbf{C}/\underline{\mathbf{T}})$, as well as on the chance of being treated when in need, $P(\underline{\mathbf{T}}/\mathbf{S})$. Thus one les-

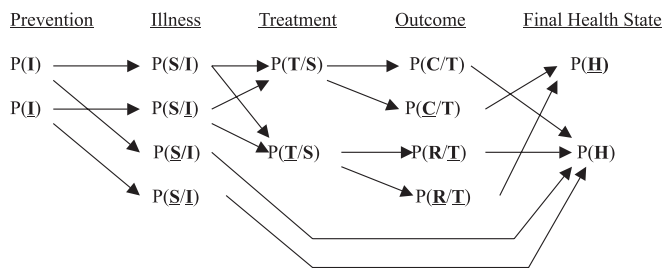


Figure 3 - Probabilities as Indicators of Equity in Health

son from this approach is that equity of final outcomes depends not only on the equity of intermediate steps but also on the effectiveness of those steps. If treatment is ineffective, for reasons of poor quality or some other cause, it does not matter much, so far as health outcomes are concerned, that some people get treated and some do not.

Effectiveness, as indicated, is measured by the difference of two probabilities. In order to calculate $P(\mathbf{H})$, probabilities are also sometimes multiplied. This can be done sensibly because the denominator of one probability is the numerator of the one before. But there are no simple linear combinations, and the underlying variables – the numbers of people given preventive care, or who get sick, or are treated, or recover – are never added together. Among other things, this means that a high probability of something desirable at one stage does not hide or offset a low probability elsewhere in the sequence.

A more significant lesson from linking probabilities in this way is that “the equity of health” is not one thing but a *series* of chances, with probabilities that may differ from stage to stage, and among people or population groups at each stage. The incidence of illness may be more or less equally distributed than the utilization of treatment, and may also vary with the equity of coverage of preventive activities. A health system may perform well at one stage – immunization, say – and quite poorly at another, such as treatment that actually resolves problems. And of course equity may look very different depending on which risk factor, disease, condition or treatment one considers. It is true that the end of the sequence is a single probability, $P(\mathbf{H})$, but differences in the chance of good health depend on differences at all prior stages, so equity cannot be summed up by inequalities in that one probability. In particular, how the health system performs in achieving equity is a question of the stages which are or should be under its control, namely $P(\mathbf{I})$, $P(\mathbf{T}/\mathbf{S})$ and $P(\mathbf{C}/\mathbf{T})$. Public health measures, especially information to the public, may also allow the health system to improve the likelihood of recovery with-

out treatment, $P(\mathbf{R}/\mathbf{T})$. These questions of how to judge the health system arise explicitly in the example discussed next.

One feature that is missing from this approach to equity is that of equity or fairness in how health care is financed. This is now seen as extremely important for its effect in preventing some people from receiving treatment because they cannot afford the cost, and also in its own right⁷. It is increasingly considered inequitable that some people should be impoverished by paying for health care, or because poor health destroys their capacity to earn income. From household data it is possible to construct indicators of the extent and severity of the impoverishing effects of recent spending on health care¹⁰. How much of current poverty is due to ill health is less easy to sum up in an aggregate indicator, partly because the ill health may be chronic or may have occurred long ago.

Estimating health sector “attainment” and “performance”

The World Health Organization has published indicators for five notions of what a health system appears to achieve (“attainments”), and for how well those achievements compare to an idea of what should be possible, given the society’s average level of schooling and the amount of money it spends on health (“performance”)⁷. These indicators have been widely criticized, particularly because they were made up without any consultation with the countries whose health systems were being judged, and also because the majority of the numbers were not based on any observation of the variables of interest but were imputed, based on other variables. This is not the place to take up the whole debate over WHO’s numbers and how they were derived; rather, the question is whether the indicators proposed illustrate some of the same issues raised by the other indicators discussed here.

One such issue is that of overloading an indicator, to the point where it is difficult or impossible to interpret, even if the underlying variables are related to the concept of inter-

est. The “performance” measure, which depends on estimating both the maximum possible attainment and the minimum that would be reached even if there were no health system, can be considered an example of overloading with assumptions: too much has to be assumed, relative to the amount of actual information in the calculation. In the case of the “responsiveness” indicator, imputing values based on regressions led to a formula which is overloaded both in variables and in the ways they are related to one another – not only linearly but also exponentially and in logarithms and powers, sometimes singly and sometimes interactively. The result – which WHO does not propose to repeat – is something shorter than the “health needs index” but even more complicated¹¹:

$$\begin{aligned} \text{Responsiveness} = & 1.123 + 0.0000423(\text{Health Spending per head}) + 0.0123(\text{Mean Schooling}) \\ & + 0.0142 \ln(\text{mean Schooling}) + 0.0000208(\text{GDP per head}) \\ & + 0.0239 \ln(\text{GDP per head}) + 0.0165(\text{Access Rate}) + 0.0417(\% \text{Private Health Spending}) \\ & + 0.0494(\% \geq 65^{**} - 0.139) (\text{Mean Schooling}^{**} 0.203) (\text{Access Rate}^{**} 0.059) \\ & + 0.0702(\text{Mean Schooling}^{**} 0.173) \text{EXP}[0.213(\text{Access Rate})] \\ & + 0.0310(\% \text{Poor}^{**} 0.562) (\text{Access Rate}^{**} - 0.462). \end{aligned}$$

Other criticisms apply to the indicator of “health status”, measured in disability-adjusted life years, and to the indicator of “financial fairness”. In the first case, the indicator does add useful information, but requires numerous somewhat arbitrary parameters to compare conditions of disability, just as for the estimation of the burden of disease¹². *The fewer arbitrary parameters, the better*, seems to be a good general rule in designing indicators; this point is elaborated below. In the second case, there is little justification for the assumption that fairness or equity means that every household should contribute exactly the same share of its non-subsistence income to financing health. Two of the inequality indicators are also more complex mathematically than may be reasonable¹³, making comparisons difficult – although WHO provides a justification for the

complexity, based on consulting preferences.

Granting all those criticisms, the WHO exercise nonetheless illustrates some virtues in the construction of indicators. First, it was thoroughly thought out what each indicator was meant to measure; there is a great deal of conceptual clarity behind them. Second, where data corresponding to the concepts did not exist, they were estimated—not only by imputation, but sometimes by construction from primary data sources for a limited number of countries. Simply using whatever numbers are already at hand is often the cheapest solution but not the best. Third, much thought went into the mathematical relations among variables or among the different values of a given variable. Where they were combined linearly, as in the “overall attainment” indicator, WHO provided a reasoned argument for not letting the marginal importance of one variable depend on the level of that variable, or any other. The weights or parameters applied to the different attainments were also derived from a consultative process rather than being specified arbitrarily. The technical failings of the resulting indicators, and the ethical failings of how they were produced and launched on the world, should not obscure these virtues. Fortunately, the resulting ranking of health systems had no consequence for what is spent, or done, in any system. Some examples are taken up next, in which indicators *do* have such consequences.

Indicators for resource allocation: sharing public money

The indicators discussed so far might be, or might have been, used to help make decisions – even the wind chill index is designed to help one decide whether to go out, how warmly to dress and how long to stay exposed – but they do not directly drive those decisions. In particular, they do not determine what gets spent, by whom or for what. Sometimes, however, governments need indicators which will do exactly that: assign resources to end uses or to other levels of government. These are usually called *formu-*

las rather than *indicators* for that reason, but some of the same problems arise as for descriptive indicators, and some of the same principles apply for their construction. They are particularly important in federal systems, where funds may be raised primarily by the central government but spent by sub-national governments, which may also contribute from their own revenues. The United States, Canada and Brazil provide examples of formulas for transferring federal funds for health to states or provinces¹⁴ for general use or for specific programs of financing health care for part or all of the population or for investment in physical capacity.

Although the formulas were devised for somewhat different purposes, it is instructive to compare them. In the United States, the formula governs the federal share of financing of the Medicaid program of health care for the poor. Federal or national expenditure in state s is $CNs = \phi s * Qs * \pi s$, while the state government's expenditure is $CSs = (1 - \phi s) * Qs * \pi s$, where C refers to cost or expenditure, Q is the total quantity of services provided to the beneficiary population, and π is the price of those services. Both quantities and prices vary among states, which have some leeway in determining who can register as a beneficiary and in negotiating prices with private providers. The parameter ϕ , which also varies among states, is the federal share: it depends on a state's per capita income, and ranges from one-half in the richer states to five-sixths in the poorest ones.

In Canada, under the previous federal/provincial arrangement for financing health care from earmarked funds, public expenditure in a province was $Cs = Q * \pi * Ps - Qas * \pi * As$. Q are quantities *per capita* and π are prices of the public program. P is the entire population of the province, while A is the population covered by private insurance. Qa is the quantity of services provided for those insured. The public expenditure was then the cost of the public coverage for everyone, less the expense covered by insurance. Each province was required to spend on health, a fraction β of its revenues Rs (own tax revenues plus non-tied federal transfers).

This way the provincial contribution was $CSs = \beta * Rs$, and that of the federal government was $CNs = Cs - CSs$.

Brazil is now starting a transition to a system like that of Canada in one respect, which is that states and municipalities are required by a constitutional amendment to dedicate, at a minimum, a specific fraction of their revenues to health, that fraction to grow as the change is implemented. In principle, that means federal funds can be re-directed to benefit poorer states and municipalities more than in the past. Before this amendment was adopted, another formula was employed to allocate federal resources (from a combined World Bank-IDB loan) for investment among the states, with the same aim of protecting the poorest ones. The states were not to be guaranteed any particular amount of investment, since what they actually received would depend on the number and quality of proposals they could submit; but funds would be reserved for them and not simply be allocated on a first-come, first-served basis for the country as a whole. The formula chosen was $Fs/F = \alpha * (Ps/P) + (1 - \alpha) * (Ds/D)$, where F is the investment fund, P is population, and D is the gap between what a state has received in federal recurrent funding, and what it would get if every state got as much per capita as the state receiving the largest transfers per head. The fraction of the total fund assigned as a ceiling to a state, Fs/F , depends on the state's share of the country's population and on its share of the total deficit or gap in recurrent federal health expenditure, compared to the best-off state. The parameter α determines the relative weights of these two shares. Brazilian law required $\alpha \geq 0.5$; the value chosen was 0.7. (The U. S. and Canadian parameters ϕ and β were not legally constrained to any particular range of values.)

Despite the differences in the detailed formulas, all these indicators have in common three properties that are likely to be valuable for indicators in general. First, the variables on which they are based are relatively easy to measure and well understood. Population, income and the quantity of services to be pro-

vided may all be estimated with error, but it is clear what they all mean, and the errors are likely to be small. The other variables – state or provincial revenues, prices paid to medical care providers, and past central government spending – are even easier to measure precisely. Second, several of these variables can be changed by government action at either the central or the state level, but none of them can readily be manipulated to the advantage of the central government or that of the recipient state or province. Once the formula is determined, cheating is essentially impossible; and there are no perverse incentives to distort the distribution of funds from what it is intended to achieve.

Third, each indicator includes only *one* arbitrary parameter (ϕ , β and α , respectively). This makes the formula much easier to understand than one with several arbitrary numbers, which in turn contributes to its being politically acceptable. More sophisticated-seeming formulas, including variables such as standardized mortality rates, were proposed in the Brazilian case. While there might have been some gain in matching transfers to needs for health spending, such complications would have put at risk the necessary understanding and acceptance¹⁵ and would have led in the direction of the unintelligible “health needs index” discussed above. All three formulas had inescapably to include at least one parameter that could only be determined or negotiated politically, but there was no need for more than one.

In Conclusion

The conclusions from this comparison of five different kinds of indicators are few and simple. Three suggestions seem worth emphasizing. First, avoid overloading an indicator with variables, as occurred with the “index of health needs” and the imputed “re-

sponsiveness” indicator of WHO. The fact that a variable bears some relation to the concept one wants to measure is necessary but not sufficient for it to be included in a formula. As more and more variables are incorporated, the problems of how to relate them increase, and the intelligibility of the result suffers. Second, ensure that the mathematical relations among the variables agree with any sound, applicable theory. In particular, do not just add variables together linearly if the real effects are non-linear or interactive. And if the phenomenon of interest actually occurs in a logical or temporal sequence, as illustrated by the indicators of equity in health, it is preferable to develop an indicator for each stage and specify how they are related – rather than trying to collapse all the stages into a single indicator corresponding to the final stage. Third, while it may be unavoidable to include one or more arbitrary parameters in the equation, keep their number to a minimum. When everything can be objectively measured or estimated, it may be possible to have no arbitrary constants at all, as in the wind-chill index. When there is a political decision to make which cannot be determined solely from the scientific evidence, it may still be possible to limit the formula to just one arbitrary number, as in the United States, Canadian and Brazilian rules for allocating federal government resources to sub-national governments. Simplicity is valuable for assuring that the indicator captures the main variables and effects of interest, and even more so for making the result comprehensible and acceptable to non-experts. None of these rules of thumb guarantees the definition and estimation of a good indicator for a complex concept, but they may help prevent the construction of measures that do not make sense or are needlessly difficult to interpret.

References

1. “National Weather Service Implements a New Wind Chill Temperature Index, from www.noaa.gov.
2. Sadana R. Measuring Health: Concepts. In: Murray L, Frenk J (eds.). *Health System Performance: Concepts, Measurement and Determinants*. (chapter 5). Geneva: World Health Organization; January 2001.

3. Murray CJL, Salomon J, Mathers CD. Summary Measures of Population Health. In: Murray CJL, Frenk J (eds.). *Health System Performance: Concepts, Measurement and Determinants*. (chapter 4) Geneva: World Health Organization; January 2001.
4. Pan American Health Organization, *Los Servicios de Salud en las Américas: Análisis de Indicadores Básicos*. Cuaderno Técnico No. 14. Washington, DC: PAHO, 1988.
5. Musgrove P. a Ruggiero AL, The 'Índice de Necesidades de Salud'.... Washington, DC: internal PAHO memorandum; 5 May 1989.
6. World Health Organization. *Macroeconomics and Health: Investing in Health for Economic Development*. Report of the Commission on Macroeconomics and Health. Geneva: WHO; 20 December 2001.
7. World Health Organization. *World Health Report 2000. Health Systems: Improving Performance*. Geneva: WHO; 2000.
8. Musgrove P, Zeramdini R, Carrin G. Basic Patterns in National Health Expenditure, *Bull World Health Organ* 2002; 80(2). p. 134-142.
9. Musgrove P, Measurement of Equity in Health, *World Health Stat Q* 1986; 39(4).
10. Wagstaff A, van Doorslaer E. *Catastrophe and Impoverishment in Paying for Health Care: With Applications to Vietnam 1993-98*. Unpublished paper. Washington, DC: the World Bank; February 2002.
11. Musgrove P. *Judging Health Systems: Reflections on WHO's Methods*. Unpublished article. 5 February 2002.
12. Murray CJL, Lopez AD. (eds.). *The Global Burden of Disease*. Cambridge: Harvard University Press; 1996.
13. Musgrove P. *Measuring Health System Attainment: Reflections on WHO's Indicators*, Unpublished article. 5 February 2002.
14. Musgrove P. *Compensatory Finance in Health: in search of Geographic Equity in a Federal System*, translated and revised version of a paper presented to the Third International Seminar on Fiscal Federalism. La Plata, Argentina, 24 April 1998.
15. Musgrove P. *Equitable Allocation of Ceilings on Public Investment*, Human Capital Development Working Paper No. 69. Washington, DC: The World Bank; August 1996.