Social indicators: Health

A. J. Culyer, R. J. Lavers and Alan Williams, *University of York*

This article was one of several presented at an international conference on social indicators held at Ditchley Park in April 1971. The conference was sponsored by the Social Science Research Council and attended by representatives from USA, Canada, France and United Kingdom, including representation from the Central Statistical Office. Current thinking on social indicators is proceeding in many and various directions. It could not yet be claimed that there is any consensus on the most desirable line of development even within particular social policy areas. This is especially true of the difficult field of health with which this article deals. Nevertheless it is part of the editorial policy of Social Trends *to keep in touch with developments in this important field. The article is therefore published not as representing a line of thought which has any especial status with public health authorities, government, or the medical profession but simply as an interesting example of the kind of substantive research which is going on in this field.*

I. Introduction

The general background to the current debate on social indicators was given in *Social Trends No. 1, 1970* by Professor C. A. Moser. In the present paper, attention is focused on the development of indicators in the particular field of health. We take it as self-evident that the choice of indicator is governed by the purposes for which it is intended to be used. These, it would appear, are broadly two. First, they are useful in recording the state and progress of groups of individuals. Second, they are useful in formulating policy: what targets shall we attempt to aim at? As is clear from the survey of some of the literature on social indicators contained in the appendix to this paper, it has not always been recognised that indicators designed to fulfil the first function will in general be different from those which are constructed with the second purpose in view. In particular, indicators which are to be of use in the formulation of policy must take account of the preferences of society and the costs involved in accommodating these preferences.

Since the policy functions will include the purely recording function, it is convenient to consider indicators in the context of the requirements of social policy decisions. This suggests that the decision taker requires three different kinds of indicator, each serving a different function and each complementary to the others but none sufficient alone for policy making. These three requirements are:

(A) A measure of the 'output' of social policies, e.g. the 'amount' of education, health, etc.

(B) A means of deriving the social valuation placed upon different 'outputs'.

(C) A measure of the technical possibility of increasing 'output'.

Together, adequate information on each of these measures is sufficient to form policy: (A) provides the units in which policy objectives are to be defined, (B) values increments (in terms of social worth) in each objective and (C) specifies what it is physically possible to do, for example, how much of one good thing must necessarily be sacrificed in order to obtain more of another.

Corresponding to each of these functions are three kinds of social indicator required in the field of health which we term as follows:

(A) Measures of the State-of-Health ('State' indicators)

(B) Measures of the Need-for-Health ('Need' indicators)

(C) Measures of the Effectiveness of health-affecting activities ('Effectiveness' indicators)

Each of these indicators can be used at a more or less aggregated level, for example to refer to society as a whole or to groups within society such as unmarried mothers, retired persons, specific social class or ethnic groups, or patients suffering from particular disorders. Each is necessary for proper planning, each raises important problems in conceptualising the kind of actual numbers needed and the extent to which what is practical corresponds to what is ideally required, and each raises problems of ambiguity of interpretation.

Need indicators are required in order to establish priorities. Not all needs can be met and some are more urgent than others. Essentially, a Need indicator would have to combine two elements: a social and humanitarian value upon an improvement in the community's health and the value of the other socially and compassionately desired programmes that would have to be gone without as a result of devoting more resources to health.

Effectiveness indicators are required in order to

sort out the effects of health services *per se* upon the community's health and to discover what inputs are required in order to achieve an objective stated in terms of the State indicator. Thus, one use for the Effectiveness indicator would be to demonstrate how by varying one such input the State indicator would respond during various time periods, or to show how different inputs may substitute for one another in promoting a given State or change of State. Essentially Effectiveness indicators provide the technical relationships between inputs and outputs. We do not underestimate the practical problems of discovering these relationships with any degree of precision.

Obviously, however, logically prior to both Need and Effectiveness indicators is the State indicator, since both of the former are variables—in principle one socially decided, the other technically determined—expressed as a function of the State indicator. If the problem of the State indicator cannot be solved, or what is a very similar thing, the problem of the *output* of health services, then no progress will be possible with the other two, for the objectives of policy, we believe, ought ultimately to be definable in terms of the state of the community's health.

In this paper we therefore turn our attention in the first place to the state-of-health (section II). Next we discuss the meaning of the need for health care (section III). We have very little to say at this stage, for obvious reasons, about the technology of medical care, though there is some discussion of this problem in the Appendix which surveys some of the literature on health indicators.

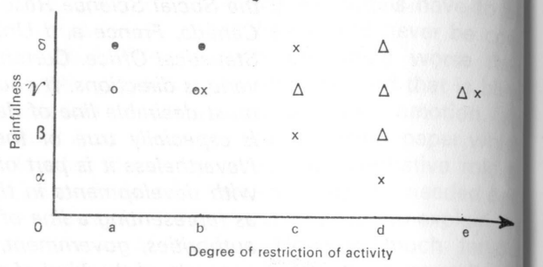
II. A proposed health indicator

In order to generate a state-of-health indicator which will also serve as an effectiveness measure, it will be necessary to devise an algorithm which will encompass both (a) medical data and judgements and (b) social judgements, with each expressed numerically in a standardised manner, yet clearly distinguished one from another. In this section of our paper we outline such an algorithm and point out some of its implications.

If we are to build up an index of health (or, in this case, of ill-health) we need to measure both intensity and duration. 'Intensity' is here interpreted as having two dimensions, 'pain' and 'restriction of activity', and for expository purpose the discussion here is conducted throughout in terms of these two dimensions only. In practice, however, it would probably be desirable to extend the number of dimensions to include other factors thought relevant such as 'distress'. The fundamental problem, however, of how to combine the relevant measures of ill-health into a single index can be illustrated without loss of generality in a two-dimensional example. The first step would therefore be to experiment with simple standardised descriptions of painfulness and of the extent to which activity is restricted, to see if there is any consensus among medical personnel as to how painful and how restricting particular conditions are, using these descriptive categories.

The initial descriptive stage may be represented as in Figure I below:

Figure I

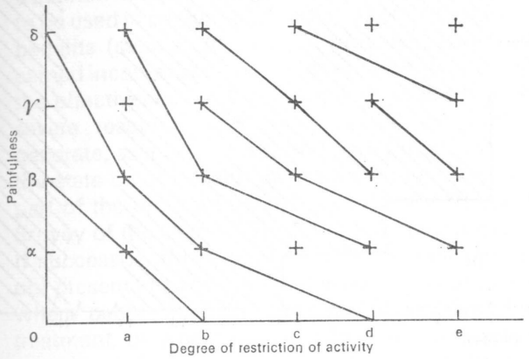


α, β, γ, and δ are simple descriptive statements concerned with painfulness (such as 'mildly uncomfortable', 'very uncomfortable', 'extremely painful' etc.) a, b, c, d and e are simple descriptive statements concerned with restriction of activity (such as light work only, confined to house and immediate vicinity, confined to house, confined to bedroom, confined to bed, etc.) [1] o, x and Δ each refer to different medical conditions or different combinations of medical conditions. For example, the medical condition Δ in Figure I is regarded by one observer as involving, for a patient suffering from it, degrees of painfulness and restricted activity described by the statements γ and d respectively. Of the other four observers who place condition Δ in the 'painfulness—restricted activity' space, two agree with the statement of the first observer on painfulness (but categorise the degree of restricted activity by statements c and e) and two agree with the statement of the first observer on activity-restriction (but regard associated painfulness as being better described by statements β and δ). Each o plotted on Figure I represents one expert's assessment of the most appropriate description of that condition in the categories offered (e.g. one says a, δ; another says a, γ; another says b, δ and yet another b, γ). Similarly each x represents corresponding judgements by other experts of the most appropriate descriptions of those conditions. The specification of medical conditions may, of course, have reference to age, social class and other attributes, and the degree of articulation would have to be such that patients suffering from each condition formed a relatively homogeneous group. If there is any consistency in these judgements (as there is in o and Δ in the example) some 'norm' will be indicated as the standard description for that condition; where no consensus exists (as with x in the example) it is likely that the condition under study needs to be more closely specified. However, supposing that we had each condition clearly ascribed to a pain

[1] A detailed 'Schedule for the Medical Assessment of Physical Disablement' illustrating the kind of thing required here is to be found in Andersen, 1964.

category (α, β, γ, etc.) and a restricted-activity category (a, b, c, etc.) we would now need to establish the trade-off between them (e.g. is the combination γa better or worse than the combination βc?). This pairwise comparison is essentially a *social* judgment and should be recognised as such, but may have to be made in practice by medical people. This first evaluative step is set out diagrammatically in Figure II.

Figure II



Each combination of a pain category (α, β, γ, etc.) and a restricted-activity category (a, b, c, etc.) is compared, and those that are regarded as approximately equivalent (in terms of the social-humanitarian benefit of avoiding them) are linked together by contour lines as indicated. In the example shown, the combinations (β, O); (α, a); and (O, b) are equivalent to each other, but better than (γ, O) and (O, c) (which may be equivalent to each other). Those in turn are better than the next group of equivalents (β, a); (α, b) and (O, d) and so on.

Despite the fact that describing the intensity of pain is notoriously difficult, and that interpersonal comparisons are bound to be rather arbitrary due, for example, to varying thresholds of pain, medical personnel can and do make such comparisons between stages and classes of condition, and such comparisons already have to be assimilated into judgements about 'acceptable' degrees of physical disability and pain at the diagnostic and therapeutic level when determining courses of treatment. It is therefore suggested that it should be possible to move to the second evaluative stage and construct (say) a 10 point scale of intensity of ill-health along the following lines:

0 = normal
1 = able to carry out normal activities, but with some pain or discomfort
2 = restricted to light activities only, but with little pain or discomfort
3–7 = various intermediate categories reflecting various degrees of pain and/or restriction of activity
8 = conscious, but in great pain and activity severely restricted
9 = unconscious
10 = dead [2]

Since it is intended to use these numbers as *weights*, and not simply as *rankings*, it is important to stress that society's judgements concerning the relative importance of avoiding one state rather than another are represented by the actual numbers attached to each respectively, e.g. state 2 is *twice as bad* as state 1, and state 10 is ten times as bad. This implication must not be shirked, and must be regarded as a statement about *health policy* (and is to be made by whoever is entrusted with that responsibility—e.g. 'the Minister'), not a technical statement about medical condition [3]. In terms of Figure II this would be represented by attaching numbers to each of the contour lines.

As to duration, this will be based on the outcome of scientific investigations, cast in statistical terms. For instance, recovery from a particular disease will follow one time-path (incorporating both intensity and duration) in 90% of the cases, another in a further 9%, and yet another in the remaining 1%. Chronic cases where no (or little) improvement in intensity is to be expected will have a duration equal to the life-expectancy of that class of individual, and the duration of the 'gain' from postponing death where successful treatment is possible will be similarly measured. A 'successful' treatment is *not only* one which reduces intensity and duration but could also be one that reduces intensity without affecting duration, or vice versa; or even that increases one at the expense of the other, providing the net outcome is to reduce the index number (a product of intensity *and* duration). The important sources of information here are the medical statisticians since it is purely empirical information that is required at this stage in the process.

There are many uses to which such an indicator could be put, but here we will concentrate on only two of them: first in a cost-effectiveness study, and secondly as a tool in measuring the state-of-health in a community.

Cost-effectiveness studies in the health field are plagued by the difficulties encountered in measuring the effects of various input changes on the health status of the clientele. An index of the kind we are suggesting might be used in the following manner for such studies.

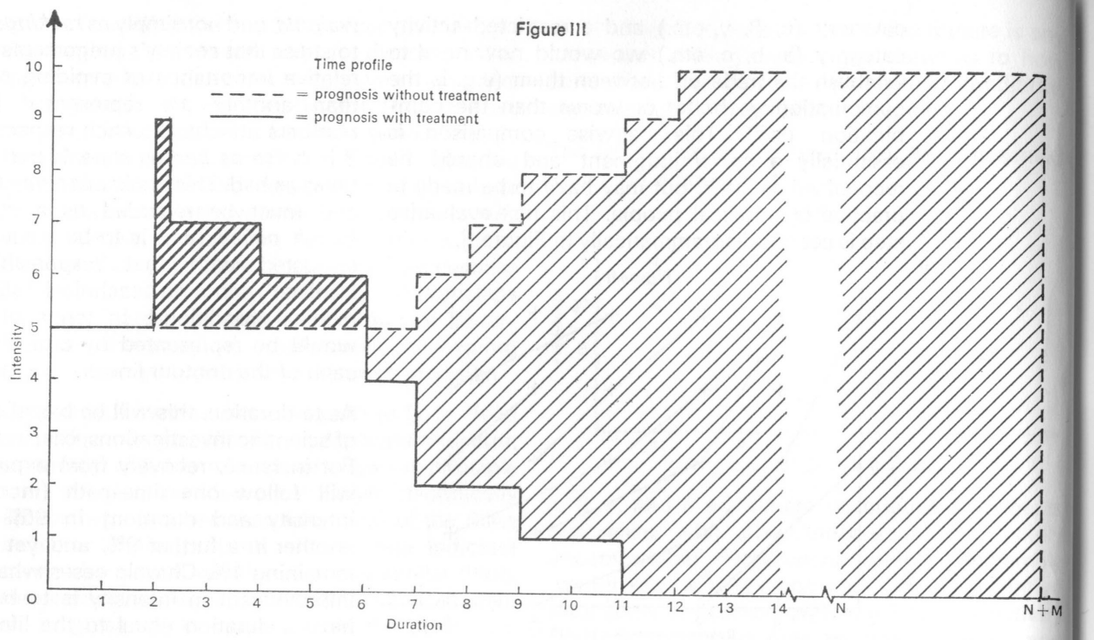
Figure III starts at a point of time O when the condition in question is diagnosed. In the illustrative example the first 2 weeks are spent in further observation, decision as to appropriate treatment, and waiting for therapeutic facilities to become available.

The prognosis without treatment (or with the best treatment other than that under considera-

[2] In this schedule, we have assumed that death is the worst state of all. In terms of Fig. II, it corresponds to a point on the horizontal axis through which a contour higher than any other passes. It may be, however, that a preferred ranking would place death on a lower contour than some other condition characterised by, say, a patient in an extremely painful moribund state.

[3] Magdeleine, *et al.*, 1967, in their work on a similar indicator fail to make this clear.

Figure III



tion) is represented by the broken line (– – – – – –), and may be described as a steady deterioration from approximately week 7, until death in week 12. This would be the standard prediction for this class of case. The average expectation of life for a person of that age/sex, etc. is represented as (N + M) which may be rather large if necessary (e.g. 50 years).

The prognosis with treatment is represented by the solid line (————), and may be described as 2 weeks of severe restriction of activity (in the pre-operative, operative and immediate post-operative phases) *plus*, possibly, considerable pain, with a steady improvement in condition during the ensuing 3 weeks, a convalescent phase from weeks 7 to 9, and a further 2 weeks taking it easy in a normal environment, after which the patient is completely normal (as far as this condition is concerned).

The index *score* (representing the 'effectiveness' of this treatment) would be the area lightly cross-hatched *minus* the area heavily cross-hatched, obviously including in the former the interval omitted in the horizontal scale as drawn. This particular example would obviously be a highly effective treatment if applied to people with long life-expectancy, less so the shorter life-expectancy. Both of the time profiles used would be derived from statistical analyses of clinical results, or experimental data if the former were lacking. It is up to the medical statisticians to provide these key data. A further sophistication which could be introduced if necessary would be to apply a discounting factor which would give less weight to future states of health compared with present states, and hence reflect the greater weight people seem to attach to the 'here and now' rather than to more distant prospects. In this way the use of the indicator would serve to narrow the area of uncertainty about the consequences of alternative patterns of resource allocation.

As a measure of a community's state-of-health the same categories could be used as a basis for a large-scale statistical survey, the object of which would be to measure both the intensity value and the duration of the various conditions affecting the population. Repeated periodically throughout the year (to allow for seasonal fluctuations) and from year to year (to establish trends) this would provide the kind of information required as a contribution to general social indicators, and would be free of many of the defects inherent in medical record-based statistics (though there would obviously be some advantage if these too included some simple standardised information of the kind here suggested).

Certain features of this system in general are noteworthy:

(a) In principle it enables preventive as well as therapeutic activities to be incorporated;

(b) although much more difficult in practice, in principle it can embrace mental illness;

(c) it treats one week of suffering at any particular intensity level as being equally undesirable irrespective of the identity of the patient. Other distributional assumptions are possible in principle, but would make the analysis much more complicated [4];

(d) it relates only to patients, and does not include infectivity, or the pain and suffering caused to others by the patient's condition. Neither of these shortcomings is insuperable in principle, but as a practical matter they will be difficult to overcome in the near future;

[4]This is a particular manifestation of a more general problem in compiling social indicators, which is that we may not wish to count each individual on a one-for-one basis when certain people register poor readings on many different indicators, e.g. health, education, crime, poverty.

(e) the *satisfaction* felt by patients themselves (or their friends and relatives) is not regarded as an independent consideration in this formulation, and to do so would raise such enormous difficulties for any health indicator that the matter is mentioned here only so that it is not lost sight of.

One important purpose of such an indicator is to facilitate cost-effectiveness studies, by providing a quantification of the purely humanitarian benefits to be used in conjunction with economic costs and benefits (such as the cost of providing care and earned income losses avoided) in order to improve the effectiveness of health services in the face of severe resource limitations. But it could also generate, as a by-product, improved indicators of the state of (ill) health of a community if used as part of the basic information matrix in a National Survey of the State of Health of the community [5]. If successful, this would fill an important gap in our present knowledge, for it would include cases where people had not presented themselves for treatment, or where those giving treatment were unaware of the patient's condition between episodes of treatment.

III. An exploration of the meaning of need

The purpose of this section is to explicate the various concepts that are commonly met in discussions of policy and to relate them to one another (leaving aside in this context the *practical* difficulties of compiling appropriate statistical series which are considered elsewhere in the Appendix) in order to concentrate initially on the purely *conceptual* difficulties. The general framework encompasses, we hope, all of the meanings of 'need' that are commonly encountered in discussing social indicators.

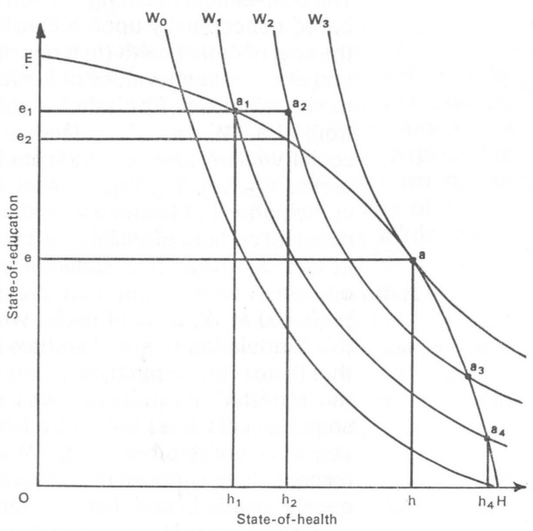
Need indicators are commonly expressed in terms of a *target* level which it has been decided that a particular state indicator should take. Alternatively, they may be expressed as the *difference* between a target and the current (or some projected) level of the indicator. These need indicators are thus expressed in terms of the state indicators.

This approach appears to us to be unsatisfactory since it is not clear how the target is decided. For example, if it is said that 'Society needs . . .' it is not clear whether what is meant implies that the speaker himself needs it, whether Society *ought* to have it in his opinion, whether a *majority* of members of Society want it, or all of them want it. Moreover, it is not clear whether it is 'needed' *regardless* of the social and humanitarian cost to Society.

We urge an alternative approach to the meaning of need, which is that the agent responsible for the decision (the 'Minister') should attach explicit *valuations* to a variety of levels of the State indicator; increments in these could then be compared with the incremental social cost of attaining any given level. Essentially, this procedure amounts to the calculation of an *intensity of need* measure which states the intensity with which 'Society' needs each of a variety of states-of-health.

Simplifying at the conceptual level, we assume that decisions regarding the meeting of social needs are taken by a single individual—the 'Minister'—who can fulfil more health needs only by forgoing known quantities of other desired entities that are needed, like education. For simplicity we assume that only the government provides these goods. The amount of health is measured by a State indicator and the amount of education by some analogous indicator. With a given budget and constant technical possibilities, he could provide either OE education *or* OH health *or* any combination of output of both education and health shown along the possibility boundary EaH in Figure IV [6].

Figure IV



Suppose that the 'Minister' has located the public sector currently at a_1 so that the community will be enjoying Oe_1 education and Oh_1 health. One possibility immediately available to the 'Minister' is to say that Oh_2 health is 'needed' (in the sense that it is a target level of the indicator to be aimed at) where h_2 is located at some arbitrary point to the right of h_1. (An alternative way of expressing the same idea would, of course, be to say that the community had $h_1\ h_2$ $(=Oh_2 - Oh_1)$ of unmet need, and one could if one wished define need to mean the difference between levels of the State indicator.) We term this concept of need

[5]For a brief description and methodological critique of such surveys see Linder, 1965. It may be that the planned General Household Survey will become a suitable vehicle for such an investigation in the UK (Moser 1970(b)).

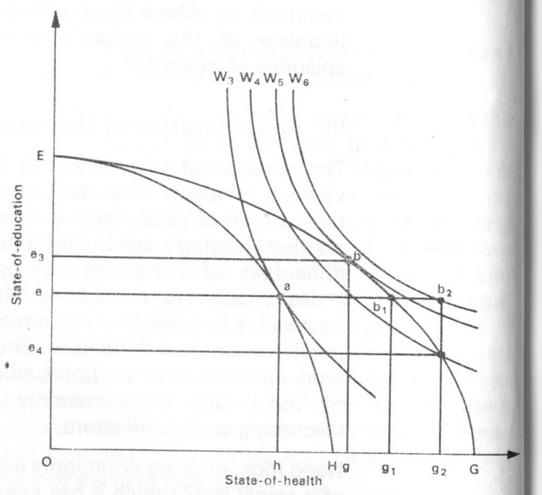
[6]In a more general analysis one could measure all non-health aspects of social well-being on the vertical axis instead of just education. The total resources of the community as a whole would then set a limit to health provision rather than the government's budget, and private as well as public provision could then be incorporated in the analysis.

'arbitrary' since it does not explicitly take account either of the fact that (out of a given budget) the need-for-health can be implemented only at some cost in terms of education forgone, or of the possibility that a more systematic consideration of the *relative social value* of education and health might enable the cost of forgone education to be weighed against the benefit of more health. This concept of need is thus based implicitly on the assumption that everything else must remain constant, in principle implying a movement to point a_2 which is, by definition, unattainable. Insistence on Oh_2 health implies a maximum of only Oe_2 education in actuality. This meaning for need is a less extreme form of the technocratic concept of need [7] which implies maximising the level of health that can be technically reached given current resources, viz., OH in Figure IV. OH implies zero education and hardly seems likely to be conducive to the well-being of society. It also serves to highlight the inadequacy of these kinds of need-statement: society might be said to 'need' something that would plainly be undesirable if the full consequences of its implementation were considered.

The cost-benefit meaning of need implied above is based conceptually upon a consideration of both the cost of more health (in terms of less education) and of the relative values of health and education to the 'Minister'. The relative values are derived from the 'Minister's' preference function, or a social welfare function, indicated by the convex curves W_0, W_1, W_2, W_3, each of which connects combinations of health and education that yield a constant amount of social well-being. For example, a_1 and a_4 show two different combinations of education and health that are socially equally preferred at W_1 level of social welfare. Note that this formulation of social welfare function implies that there is no hierarchy of needs in the sense that the 'Minister' will wait until one need is satisfied at some specific level before he begins to be concerned about another need. We assume that it is reasonable to suppose that all needs are simultaneously needed, and hence that there is some increase in the level of health which would be regarded as off-setting a small reduction in the level of education, or vice versa. In other words we assume that needs are substitutable in the social welfare function. There is no reason to suppose that they will be substitutable at a constant rate, however. Indeed it can plausibly be argued that the more health the community enjoys relatively to the amount of education, the less it will value additional health relatively to additional education. This is what the convexity of the iso-welfare lines in Figures IV and V represents; namely, a changing rate of trade-off between education and health indicators.

The cost-benefit prescription is to maximise social well-being subject to the resource constraint and the technical possibilities, viz., to locate at a, where the slope of an 'iso-welfare' line (W_3) is the same as that of the possibility curve EaH. If social well-being is to be maximised, the cost-benefit concept that implies that Oh health is needed is clearly the best of these various approaches since it is the only conceptual framework that describes the nature of the ideal satisfactorily. Some of the dangers of the other concepts will be clear from inspection of Figure IV. Starting from a_1 on W_1 the technocratic concept takes us to H on W_0, a lower level of social well-being. The less extreme, but also arbitrary, concept such as h_2 could also result in a loss of social well-being. For example, suppose that the system were already located at a_3 lying between a and H on EaH, so that the target level h_4 would imply a location between a_3 and H. In this situation any further movement towards H must inevitably take the community away from the ideal rather than towards it.

Figure V



Note that in this figure we have assumed a uniform 50% rise in the technical possibilities for health, whatever its present level of output, and no increase in possibilities for education.

The arbitrary measure can, however, provide a correct guideline for policy under certain circumstances, though as we shall see these circumstances, while strictly definable in principle, are not obviously identifiable in practice, thus making reliance on the arbitrary measure hazardous. These circumstances can be conveniently investigated if we now assume that technical progress occurs in medicine, making it possible for the community to have more health without necessarily sacrificing education. In Figure V, this new development is indicated by the new possibility curve EbG [8]. For convenience, EaH may be viewed as applicable in, say 1970, with EbG applicable in 1985. If in 1970 the public sector were located at a (i.e. at the social optimum), a is now, in terms of 1985 possibilities, an *interior* position. Under these circumstances any rightward move along ea, so long as it did not go beyond b_1 would constitute an unambiguous social gain. Thus, the arbitrary measure of need, provided that it did not exceed the technical possibility of attaining the target without sacrificing any education (and provided also, of course, that it did not specify any *reduction* in health), would be a valid indicator of an improvement in social well-being [9].

It is, however, readily apparent that the arbitrary measure will not usually locate the most socially desired point b. The best it can do is to produce a movement to b_1 on the new possibility boundary and this must lie on an iso-welfare curve lower than that on which b is to be found. The implication of the 'Minister's' choice of b is that he judges it to be in the community's interest that the improvement in productivity in the medical sector be partly exploited in the form of released resources rather than greater output, these resources being transferred to the education sector, (e.g. releasing nurses to become teachers) [10]. This constitutes a further potential source of weakness in the arbitrary measures of need, in that they tend implicitly to assume that the fruits of technical progress in any sector are to be enjoyed only in that sector. A level of need defined as Og_2 is again undesirable, as was Oh_2 before, and for analogous reasons since it implies that b_2 is attainable when it is, in fact, not attainable. Similarly, OG would be even worse. Again, then, we find substantial reasons for preferring the cost-benefit definition of need at the conceptual level since it alone avoids the possibility of mistakes of principle whereas the other concepts, whatever their practical relevance, can lead to erroneous specifications of the problem (and hence, presumably, erroneous solutions of the problem as well). In practice, the difficulty with the cost-benefit notion of need is that it requires that the properties of the social welfare function be identified in the relevant range.

Summarising the principal themes of the argument this far, two major propositions have been advanced:

a) Not all changes in the state of health (SOH) of given absolute magnitude will be of the same value in terms of social well-being.
b) Even if we assume that any change in the SOH makes a *positive* contribution to social well-being, this is not necessarily to recommend its implementation as a target because its contribution, though positive, may be less than the losses incurred elsewhere as a necessary result of its implementation. These losses, or marginal social costs, can be expressed either in terms of the value of inputs required to increase the SOH indicator or in terms of the value of the other desired outputs foregone [11]. As the perceptive reader will have noted, the cost side of the calculation will rely heavily on the development of satisfactory Effectiveness indicators.

The implications of these propositions are broadly threefold. First, explicit valuation of the (marginal) trade-offs between SOH and other outputs is required. Effectively, this implies that the 'Minister' should take a range of different SOH levels and ask: *how intensely is each of these levels of the State indicator needed?* These valuations, or measures, of the intensity of need at each level, would be expressed in terms of some numeraire (e.g. education, money). At the analytical level they correspond to the slopes of the iso-welfare lines as they pass through EbG vertically above each SOH level. There is a strong case, we believe, for viewing these indicators of the intensity of need as the need indicators themselves.

Although they have not been graced with the title of 'need', some attempts to calculate such an index have been made. The most notable example in Britain is the Department of the Environment's use of explicit values of human lives expected to be saved by road improvement investments. This indicator presents formidable problems of quantification. Progress is however, being made toward their solution [12].

Second, the cost-benefit approach offers the most comprehensive method of coping fully with the problems raised in devising and using social indicators. In particular, the cost-benefit approach includes an explicit valuation of the (marginal) trade-offs in both consumption (Need indicators) and production (Effectiveness indicators) between SOH output and inputs and other outputs and inputs. It also provides methods for devising shadow prices where social institutions have failed to produce any or where they are known to have been produced only imperfectly.

Finally, these more ambitious methods are predicated upon the existence of a State indicator though the latter cannot be regarded in any way as an adequate *substitute* for them. Essentially, the State indicator provides the dimensions of 'output' in which the problem for solution is measured. We emphasise that we regard the calculation of the State indicator in this light—as a necessary but limited (though difficult) first step.

One possible conclusion derived from this

[7] As exemplified in the statement by Cohen, 1968, quoted on page 38.

[8] Note that another use of this technique of analysis would enable examination of an increase in the government's budget. Under these circumstances the curve EaH would shift *upwards* as well as to the right, indicating that more of everything could now be had even if no technical progress at all had occurred (of course, something in the private sector would have been sacrificed).

[9] The same point applies if the public sector is presently located at an interior combination due to inefficiency in the use of inputs. Here the target should not exceed the improvement in performance predicted from the more efficient resource use got as a result of (say) a cost-effectiveness study. We shall not, however, pursue the matter of technical inefficiency any further in the present context.

[10] In the extreme we might envisage the social optimum being vertically above a, implying that the benefits from technical progress in medicine were reaped wholly in the form of increased levels of education.

[11] Under certain conditions these two alternative ways of expressing the cost amount to the same thing. For cases where they diverge, there are means of selecting the appropriate measures.

[12] Some literature in this area is: Dublin and Lotka, 1946; Dawson 1967; Schelling, 1968; Fromm, 1968; Jones-Lee 1969.

discussion may be that the word 'need' ought to be banished from discussion of public policy, partly because of its ambiguity but also because we believe it generally true that the word is most frequently used in the 'arbitrary' senses of which we have been rather critical. We hope to have shown here that there *is* a meaningful and useful, if unconventional, concept of need. The calculation of a Need indicator based upon our concept is perhaps, however, a task to be tackled only when it seems clear that the operationality of a State indicator such as that proposed elsewhere in this paper is feasible. Speculative horses are perhaps best backed one at a time.

IV. Conclusions

The chief aim of this paper has been to devise a conceptual scheme with particular reference to health, which will assist in the construction of social indicators designed to achieve the objectives of measuring the quality of life, estimating the effects on social phenomena of the actions of government, commercial and voluntary organisations and of individuals, and measuring the magnitude of social problems, the rate at which these are changing and the manner in which they are inter-related. These general objectives in the construction of social indicators are discussed, for example, in Allen, 1968; Bauer, 1966; Cohen, 1968; U.S. Department of Health, Education and Welfare, 1969; and Moser, 1970 (a) and (b).

In a number of respects, however, we find ourselves in disagreement with contributors to the recent debate. It should by now be clear, for instance, that we do not regard some of the indicators reviewed by Biderman, 1966, as useful social indicators, on the grounds that the goals whose achievement they are designed to monitor relate to inputs (e.g. 'more hospitals, clinics, nursing homes'). State indicators, as well as being free of any input content, should also in our view be distinguished from Need indicators. The statement of Cohen, 1968, that 'there is a need for statistics which indicate clearly and precisely present conditions in our society, *including,* for example, the magnitude of existing social problems and their rate of change' (our italics), fails to make this distinction clear, for while State indicators may serve adequately to indicate present conditions, the magnitude of social problems can properly be measured only by taking into account society's preferences and the opportunity cost of satisfying them. Need indicators, that is to say, should not be subsumed under State indicators, since to confuse them leads to such unhelpful statements as 'we are still far from assuring every American the right to the best health care that modern medical science makes possible' (Cohen, 1968).

We have also argued that policy values must inevitably enter into indicators, both 'internally' (e.g. relative valuations of different dimensions of health within a State indicator) and 'externally' (e.g. the value of a change in a State indicator relative to a change in some analogous indicator of education). We do not therefore regard the characterisation of some indicators as 'evaluative' as opposed to 'informative' (Moser, 1970 (a)), or of their uses as 'scientific' as opposed to 'unscientific' (Biderman, 1966), as being helpful in this respect. As Moser, 1970 (b) has pointed out, 'in practice the weighting of the components of the quality of life according to some set of values is taking place all the time.' In terms of the conceptual scheme we have used, even if society's preference structure remains constant the value of a given change in a State indicator will depend on the level of the indicator as well as on the levels of other indicators, and even if a high value is placed on a given change in a State indicator it would be necessary to identify a Need indicator along the lines we have suggested (ie. incorporating social values) before the change in the State indicator could be adopted as a goal.

What we have proposed both in the way of conceptualisation and potential measurement may appear to be unduly abstract and academic to some. Our defence against such charges is that unless the conceptual framework which defines the purposes and explores the assumptions implied in the use of social indicators is made absolutely clear, the danger is that a new statistical giant with an insatiable appetite for data of all kinds is likely to be the firstborn—and probably stillborn—offspring of the social indicators movement. Our conceptualisation is predicated firmly on the kinds of information that appear to be *necessary* for national policy formation and we have attempted to indicate how the first operational step towards implementing these ideas may be taken. We do not underestimate the difficulties that remain. Even the implementation of our limited first step will involve a long, sustained and heavy programme of study and research. But the payoff is potentially considerable and for this reason we recommend that some group interested in community or social medicine be encouraged and financed to conduct a pilot study along these lines to see how feasible the proposals are.

Appendix. The present state of the art

The purpose of this Appendix is to review some of the literature on health indicators and to comment on some of the relevant available data. Many of the studies to which reference will be made were not designed specifically for the construction of health indicators, but are of obvious relevance.

(a) Indicators of State-of-Health

In general it is necessary to distinguish State indicators proper, which relate to the output dimension, from indicators of environmental conditions and of health service provision, which are measurements on the dimension of input.

Indicators of environmental conditions, e.g. the quality of the water supply, give no direct information on the State-of-Health in a population and on changes therein, even though they may appear to do so.

Similarly indicators of health service provision, such as hospital beds or doctors per thousand of population, while they abound in e.g. the annual reports of the Department of Health and Social Security and other published sources of data, are not State indicators. Although input measures have been proposed as indicators of level of health in, for example, WHO, 1957, indicators of State-of-Health need to be quite free of any input content in order that the effects on them of varying input combinations may be subsequently estimated. We also note that the State indicators considered in this section—measures of mortality, morbidity, restriction of activity and composite measures—are indicative of the degree of ill-health rather than of the level of positive health of a population.

(i) *Mortality measures*

Data on mortality, due to their availability and comparability across time and space, have been traditionally used to indicate levels of health, and form the logical starting point of such general reviews as that by Sullivan, 1965. However, neither comprehensive (crude and standardised death rates, expectation of life at a particular age, or the ratio of deaths over 50 years of age to all deaths suggested by Swaroop and Uerema) nor specific indicators of mortality (infant and neonatal rates, rates for infective and parasitic or the degenerative diseases) are direct indicators of the level of health of the living members of a population. As noted by Acheson, 1968, however, they might be of more relevance in situations where ill-health is due largely to a few fatal diseases whose evolutionary time-scale is relatively brief, e.g. in under-developed countries.

(ii) *Morbidity measures*

Simple measures of morbidity which might be used as indicators of health with respect to particular disorders, such as the incidence (flow rate of new cases over time) and point-prevalence (stock rate at a particular time) often involve difficulties of interpretation: Fletcher and Oldham, 1964, for example, point out that diseases with a high prevalence may have a high or low incidence combined with a (respectively) high or low case fatality or recovery rate.

More complicated actuarial measures require data on age-specific first diagnoses: morbid risk, the probability of contracting a disease in the age interval x to x + 1; Strömgren's period of risk, the probability of a newly-born who survives to age x contracting a disease; and the expectation, or risk of a newly-born contracting the disease (Moran, 1969). Even apart from the problems created by the absence of agreed diagnostic criteria, data on first diagnoses are not available for human populations except as revealed by sample surveys of general health among the community in a particular area or of particular disorders. Butterfield, 1968, for example, found that only 5% of a sample of 2,165 from the electoral roll of Bermondsey and Southwark in 1965 were without symptoms over a 2-week period, and that for all categories of symptoms on which information was obtained, only about 11% of those suffering symptoms were making use of some agency of the N.H.S. Similar studies reported by Butterfield recorded the proportion of 'completely healthy' persons as 12%. The only national survey of sickness among a sample of the population of England and Wales covered the period 1943–52, and is described in Linder, 1965.

In practice, however, morbidity usually refers to numbers suffering from illness and to rates as revealed by data on the use of facilities such as hospital in-patient care and G.P. services (Department of Health and Social Security, 1969). Providing the relationship between morbidity in this sense and total morbidity in the community remains constant through time and space, such data could legitimately be used to construct State indicators for comparable periods or regions. Except in the case of groups of disorders for which treatment is indicated to be essential, in-patient data are an imperfect guide to morbidity in the community, and although information on numbers on the waiting lists of hospitals is also available, these numbers depend considerably on the referral practice of G.P.'s, which in turn are in part a function of available hospital facilities. The addition of numbers on the waiting list to throughput (i.e. deaths and discharges) would therefore fail to provide unambiguous indicators of the state of health of a community. Moreover, data such as those contained in the Hospital In-patient Enquiry on a 10% sample of discharges-and-deaths (undifferentiated) of in-patients from non-psychiatric departments of hospitals give no indication of the relative numbers of first and subsequent admissions involved. Data on morbidity as revealed by information on the usage of facilities other than hospitals are not generally available and only collected with considerable effort (Logan and Cushion, 1968; Ashford and Pearson, 1970). State indicators constructed from hospital 'throughput' data alone might therefore record a misleading movement if inputs to the system change: e.g. the future planned reduction of psychiatric beds to around 0·5 per thousand from a figure of around 2·6 (Bransby, 1969) would be reflected as an improvement in any such State indicator. Even for disorders where hospital treatment was indicated as essential, such State indicators would have to be supplemented by data on mortality and disability before any inferences could be drawn about changes in them. Of the currently available time-series data on morbidity, the only ones suitable for the construction of State indicators would seem to be

notifications of some of the acute infectious diseases, and these are of little relevance to the overall health of the community.

In addition to the difficulties associated with the choice of morbidity measures (e.g. incidence or prevalence rates) and the coverage of the data from which these are calculated (total population or users of health facilities only) the construction of satisfactory State indicators involves the problems of selection and aggregation. That is to say, how are morbidity measures for different disorders to be combined into a single index useful for inter-temporal and cross-section comparisons? In the formulation adopted by Draper, 1963, for example, the problem becomes one of selecting the values of the weights in a linear combination of age–and sex–specific morbidity measures. It is important to note that even when these weights are generated by some ostensibly neutral technical procedure, such as the regression of air pollution levels on the prevalence of a number of diseases thought to have this aetiological factor in common, at heart their selection is a policy decision (in this example the decisions about the importance of air pollution, the functional form of the regression equation, and the regressor diseases to be included).

(iii) *Measures of restriction of activity*

The use of measures of the extent to which activity is restricted in a population, e.g. the number of days of restricted activity in a year as State indicators, is discussed by Sullivan, 1965. As with measures of total morbidity in the population, these indicators could be constructed from information gained from such inquiries as the Survey of Sickness in England and Wales, 1943–52, which included questions on interference with usual activities, days kept indoors etc. (Linder, 1965). Such measures would not be without ambiguity as State indicators, however, since whether or not a morbid condition restricts activity depends partly on the occupation, marital status, personality, etc. of an individual.

Published data relating to restriction of activity are of two kinds: the first takes the form of numbers of permanent registered disabled at points of time, which do not necessarily reflect accurately the prevalence of disability in the community, even assuming that the categories eligible for registration are the only ones relevant. Taylor and Fairrie, 1968, for instance, found that although 11% of the male population of working age in an area were eligible for registration, less than one-third of these were in fact registered. Due to the effects of degree of information available and social attitudes, moreover, this proportion is unlikely to remain constant across space and time. Secondly, for the insured population only, series are available on spells and days of certified incapacity by diagnostic category over time, and on total numbers incapacitated at points of time. As with data on discharges of hospital in-patients, several spells of incapacity may be accounted for by one or more persons. Furthermore, the use of such data to construct State indicators is bedevilled by the intervention of factors other than level of health, such as job satisfaction and unreported sickness absence.

(iv) *Composite measures*

Many of the inadequacies in the separate use of measures of mortality, morbidity or restriction of activity as bases for the construction of State indicators arise from the interdependent nature of these phenomena: an indicator based on the mortality experiences alone of a population, for example, might record an improvement at the expense of increased morbidity and/or a greater degree of restricted activity among the population. Two earlier indices designed to overcome such difficulties are described by Sullivan, 1965, and include morbidity or restricted activity components as well as a mortality component: the index proposed by Chiang is a weighted mean of age-specific functions of the death rate and the average yearly duration of illness, and Sanders' method involves a modified life-table calculation which takes account of the age–and sex-specific risks of 'functional inadequacy'.

A simplified development of Sanders' proposal, the expectation of healthy life, has been adopted more recently in U.S. Department of Health, Education and Welfare, 1969. It is worth noting that the implication of using as a State indicator the expectation of life at birth adjusted for expected number of bed-disability days is that society is indifferent between the loss of one day of life and a day confined to the sick-bed: a principle enshrined in the couplet 'If you're stuck in bed, You might as well be dead'. A further practical aspect of the expectation of healthy life is that if data on the use of health facilities are used as the basis for estimating expected bed-disability days, the resulting measure will be partly a function of the level of inputs: if total available hospital beds were reduced, for example, the measure would be likely to register an increase, i.e. improvement in State-of-Health.

A rather different approach to the construction of a composite measure is adopted by Magdeleine *et al.*, 1967, who make use of data on the nature of perceived morbidity and likely duration of restricted activity among a sample of almost 4,000 male subjects. Here, an individual's position on a seven-point ordinal scale measuring degree of morbidity is determined by three aspects of the disorders from which he is suffering: the probable evolutionary character, degree of invalidity and risk of death involved. The evolutionary character encompasses the expected duration of the disorder as well as its prognosticated course (in terms of deterioration or improvement) and although each of the three dimensions is intended to be independent, the distinction between evolutionary character and degree of invalidity is not always clear from the examples given. The ordinal scale derived is then converted to an interval level of measurement by assuming that degree of morbidity is a continuous variable and that each interval from the values 1 to 5 is of unit length, while the extreme intervals have infinite lower and upper values.

Magdeleine's indicator of mortality is not exactly a State indicator which measures the health status of a group at a point in time or over some fixed period, since its components are partly prospective (prognosticated course of, and risk of death attaching to, the disorder) and the prospective period envisaged will vary from one disorder to another. These prospective components, moreover, are valued by reference to the *expected* course of the disorders suffered, rather than by recording the progress of the individuals suffering from them. Any attempt to construct State indicators on the basis of data on point or short-period morbidity of a once-for-all kind, of course, has to resort to such devices in order to encompass as many relevant factors as possible. As should be clear from Section II above, however, the use of longitudinal data would enable a more satisfactory approach to be adopted by treating death as a point on the invalidity dimension and by measuring the duration of invalidity suffered by individuals directly.

(b) Indicators of Need-for-Health

The development of Need indicators which are functions of the current values assumed by State indicators and a set of target values of these latter is clearly contingent on the construction of satisfactory State indicators. In some contexts, the sense in which the word 'need' is used comes close to having this meaning for particular aspects of health: if the proportion of females of child-bearing age who are immune to rubella is regarded as a simple State indicator, for example, and a target of complete immunity is decided upon (as by the Department of Health and Social Security recently) then the proportion of susceptibles is an implicit Need indicator in our sense. (Frequently, however, by 'need' may be meant the target value of the State indicator itself).

More commonly, needs have been discussed in relation to the provision of health services, i.e. in the context of the input dimension. Bailey, 1961, for instance, discusses the estimation of the need for hospital beds in particular areas subject to given lengths of in-patient stay and turnover intervals, and measures such as the payment of additional amounts to G.P.'s practising in certain districts are designed to meet the need for a given ratio of doctors to population to be maintained.

(c) Effectiveness Indicators

Effectiveness also is frequently interpreted in terms of the extent to which quantities of inputs to the system of health provision reach specified levels. The investigation described in Nuffield Provincial Hospitals Trust, 1960, for example, evaluates the effectiveness of hospital casualty departments by reference to such variables as the relative number of hours during which consultant cover is provided. The health problem index described by Parker, 1967, is designed to measure the effect of devoting resources to particular diseases or aspects of health, and would therefore seem to serve as a measure of effectiveness in the sense of this paper. The arguments of the index, however, include relative length of in-patient stay, total in-patient bed-days and total out-patient visits over a period, as well as measures of mortality and restricted activity. Since the former three arguments are partly input measures, our criterion of a satisfactory indicator is vitiated. As an example of the paradoxical results of applying the health problem index, which is an increasing function of all the above mentioned arguments, we note that an increased allocation of resources to the hospital service would probably result in an increase in the index, i.e. a more serious health problem, by virtue of the effect on in-patient bed-days and out-patient visits.

A more general approach to the problem of measuring effectiveness, or rather ineffectiveness, is taken by Packer, 1968, who adopts as a starting point a measure of ineffectiveness for the individual which is given by a weighted sum of the estimated time spent in a number of states of restricted activity, the weights to be determined by social policy, and goes on to discuss ways in which these individual measures may be aggregated. The resulting measure, for example, might be a function of the number of individuals in various states, the time spent in these states, and the weights attached to them. Resource allocation among different programmes (e.g. the provision of preventive measures or treatment facilities) could then be decided upon according to the estimated effect of the programmes on the measure of effectiveness. Such an approach, although it involves a considerable effort in data collection and the development of individualised data schemes or linked records, would seem to be the most promising if health indicators are to be useful guides to policy.

References

Acheson, E. D., 'Some remarks on contemporary British medical statistics', *Journal of the Royal Statistical Society,* A, 131, 1, 1968, pp. 9–12.

Allen, D., 'Social Indicators', *Social Science Research Council Newsletter,* 4, 1968, pp. 13–15.

Anderson, Bent Rold, *Fysisk Handicappede i Danmark* (The Physically Handicapped in Denmark), Vol. 4, 'Abegde og Erhverv' (Work and Employment), Social-forskningsinstitutte (The Danish National Institute of Social Research) Publication No. 22, Teknisk Forlag, Copenhagen, 1964, p. 187.

Ashford, J. R., and Pearson, N.S., 'Who uses the health services and why?' *Journal of the Royal Statistical Society,* A, 133, 3, 1970, pp. 295–357.

Bailey, N. T. J., 'Calculating the scale of in-patient accommodation', in Nuffield Provincial Hospitals Trust, *Towards a Measure of Medical Care,* O.U.P., London, 1962.

Bauer, R. A. (ed), *Social Indicators,* M.I.T. Press, Cambridge, Mass., 1966.

Biderman, A. D., 'Social Indicators and Goals', in Bauer, 1966. op. cit.

Bransby, E. R., in Discussion on Kramer, 1969, op. cit. infra.

Butterfield, W. J. H., *Priorities in Medicine,* Nuffield Provincial Hospitals Trust, London, 1968.

Cohen, W. J., 'Social Indicators : statistics for public policy', *American Statistician,* 22, 4, 1968, pp. 14–16.

Dawson, R. F. F., *The Cost of Road Accidents in Great Britain,* Road Research Laboratory, Report LR 79, Crowthorne, 1967.

Department of Health and Social Security, *Digest of Health Statistics for England and Wales,* H.M.S.O., London, 1959, pp. 87–98.

Draper, G. J., 'A suggested method for constructing indices of morbidity', *Applied Statistics,* 3, 1963, pp. 26–37.

Dublin, L. I., and Lotka, A.J., *The Money Value of a Man,* The Ronald Press, New York, 1946.

Fletcher, C. M., and Oldham, P. D., 'Prevalence Surveys' in Witts, L. J. (ed), *Medical Surveys and Clinical Trials,* O.U.P., London, 1946.

Jones-Lee, M., 'Valuation of Reduction in Probability of Death by Road Accident', *Journal of Transport Economics and Policy,* Vol. III, No. 1, 1969, pp. 1–11.

Kramer, M., 'Statistics of Mental Disorders in the U.S.', *Journal of the Royal Statistical Society,* A, 132, 3, 1969, pp. 353–407.

Linder, Forrest E., 'National Health Interview Surveys', in W.H.O., *Trends in the Study of Morbidity and Mortality,* Public Health Papers No. 27, Geneva, 1965.

Logan, W. P. D, and Cushion, A. A., *Morbidity Statistics from General Practice,* G.R.O., London, 1958.

Magdeleine, M., Mizrahi, A. and A., and Rösch, G., 'Un Indicateur de la Morbidité Appliquée aux Données d'une Enquête sur la Consommation Médicale', *Consommation,* No. 2, 1967.

Moran, P. A. P., 'Statistical Methods in Psychiatric Research', *Journal of the Royal Statistical Society,* A, 132, 4, 1969, pp. 484–524.

Moser, C. A. (a), 'Measuring the Quality of Life', *New Society,* 10th December, 1970.

Moser, C. A. (b), 'Some General Developments in Social Statistics', *Social Trends,* No. 1, H.M.S.O., London, 1970, pp. 7–11.

Nuffield Provincial Hospitals Trust, *Casualty Services in their Setting,* O.U.P., London, 1960.

Packer, A. H., 'Applying Cost-Effectiveness Concepts to the Community Health System', *Operations Research,* 16, 2, 1968, pp. 227–253.

Parker, R. D., *Some Interesting Problems in Medical Operations Research,* John Hopkins University, Baltimore 1967.

Schelling, T. C., 'The Life you save may be your own', and Fromm, G., 'Comment on Schelling's paper', in Chase, S. B. (ed.), *Problems in Public Expenditure Analysis,* The Brookings Institution, Washington, D.C., 1965.

Sullivan, D. F., *Conceptual Problems in Developing an Index of Health,* Office of Health Statistics Analysis, U.S. Department of Health, Education and Welfare, Washington, D.C., 1966.

Taylor, P. J., and Fairrie, A. J., 'Chronic Disabilities and Capacity for Work', *British Journal of Social and Preventive Medicine,* 22, 1968.

U.S. Department of Health, Education and Welfare, *Towards a Social Report,* Washington, D.C., 1969.

World Health Organisation, *Measurement of Levels of Health,* Geneva, 1957.