Systems Aspects of Health Planning, N.T.J. Bailey/M. Thompson, eds.
North-Holland Publishing Company, 1975

RP-75-7
Reprint

SYSTEMS ANALYSIS FOR THE EVALUATION OF BIO-MEDICAL RESEARCH

Mark Thompson
International Institute
for Applied Systems Analysis

INTRODUCTION

Systems analysis has long been attributed great potential for the guidance of research and development efforts. In the bio-medical sphere, systems analysis has been little tried and then without noteworthy success. In large part this is due to the huge, complex, and unmanageable uncertainties that plague the area. It owes also to the paucity of sound models developed for research evaluation.

Considerations that should be addressed by such models are presented in the second section. The rudimentary analytic framework described there discusses how the gains from morbidity reduction, from life extension, and from their consequent externalities might consistently be taken into account.

The third section describes a practical attempt to apply the concepts brought out in the second section. We enumerate a list of difficulties encountered when medical experts were asked to provide the quantitative estimates required by analytic models.

In the final section of the paper, we evaluate the evaluation described in the third section. While concluding that it failed in its main tasks, we find that gains were registered in tangential areas. The most important of these gains may have been an improved understanding of the potential for systems analytic evaluation of research. Ways are indicated for improving future efforts of a similar nature. It is argued that the misgivings of medical reviewers with the model used were largely based upon the failure to allow for and to incorporate uncertainties. The extent of these uncertainties seems so great as to prevent the model in present form from being used to direct the more basic research. We argue finally that systems analysis is essential to a more sensible allocation of research monies and therefore must work to correct the shortcomings here described.

IN THEORY

The Need for Models

To differentiate fairly and accurately between the relative values of competing research proposals, a model making clear the basis for those values is needed. The model should translate the foreseen results of bio-medical research--ultimate improvements in the delivery of health care--into benefits. It should thereby provide a comprehensive and systematic means for gauging the value--ex ante and ex post--of the research.

We will briefly discuss factors to be borne in mind in constructing such a model. Initial consideration will be given to the most basic situation of a contemplated research task with foreknown results and costs. Once an estimation of gross benefits has been obtained, costs may be subtracted to obtain net benefits. If the analysis is sound, projects with negative net benefits ought not to be funded and, of equally expensive competing projects, that with the highest net benefit should be preferred.

Benefit Estimations

Benefits of bio-medical research derive from two fundamental effects: the prolongation of life and the reduction of morbidity--which thereby enhance the quality of life. When these effects are consistently accounted for, distinction between advances in therapeutic care and in prophylactic praxis need not be distinguished.

At least two beneficiary groups should be borne in mind in gauging the effects both of extending lives and of reducing morbidity. The first is the population sub-group that is, will be, or would be afflicted by the disease, disorder, or condition in question. The second is the extensive union of individuals--in the family and the community--that become better off because members of the first group live longer or more healthy lives. This second set of benefits thus comprises a far-flung range of externality benefits--from the economic gains in reduction of absenteeism or the loss through death of highly trained personnel to the non-economic satisfaction that one's mother lives longer or with less pain. The difficulties in estimating non-economic externality benefits have led many analysts to exclude them from their models. Economically measurable benefits are most easily incorporated within quantitative analyses. Of these, the overall societal gain deriving from reduced absenteeism is that most often calculated. Other gains that should, at least conceptually, be included are the increased productivity achieved by workers with higher health status and the preservation of the socially sunk investment in training when the working life of a skilled craftsman is extended.

Cost Estimations

For most analyses, it suffices to estimate research costs--covering labor, equipment, and management--as a single monetary sum. In exceptional cases, modification is needed to reflect the opportunity cost of scientific resources. When research succeeds, it frequently occasions higher treatment costs to bring to the population the benefits of its advances. These should, in addition to the research costs, be subtracted from the gross benefits to obtain net benefits. When research leads to lower cost treatment, the total societal cost differential should be treated as a benefit.

Less often, successful research requires major reorganization or expenditure before its benefits are realized. These costs include:

1) the acquisition of equipment and other facilities, and

2) the training of personnel.

When significant reorganization of the health care delivery system is required, this should be treated as a cost, estimated, and included within the analysis. Another type of cost to be incurred before benefits are realized is the expense of campaigns--such as those to control disease vectors, to reduce environmental carcinogens, or to warn against habits dangerous to health.

Methodological Considerations

To estimate benefits and costs according to the framework described above, a number of hampering methodological difficulties must be overcome. These include:

1) Breaking down the population into sub-groups to clarify benefit incidence. Different levels of fineness or coarseness in this breakdown will be required for different studies and purposes. Although breakdown by age and sex cohorts often suffices, further disaggregation into groups of varying vulnerability may be required;

2) Attributing values to gains in life extension and in morbidity reduction as perceived by the potentially afflicted individuals. Analysts have had almost as much trouble themselves in describing a methodology for such imputations as they have in coaxing the quantifications

from others. Their arguments that many governmental decisions--necessarily if implicitly--place money values upon being alive or more healthy do not ease the estimations. Despite the difficulties incurred, these values ought not to be neglected because i) they are needed to differentiate between the work of an artificially prolonged but severely impaired life and a more healthy and natural existence, and ii) a major gain from morbidity reduction inheres precisely in the internally perceived benefits of a higher health status;

3) Discounting future benefits. For consistent and appropriate weighting of present and future priorities, time discounting is essential. In problems of bio-medical research valuation, large lag times are likely between completion of the research and its effected social benefits. The benefits may then be spread over decades. These factors magnify the importance of the methodology used for gauging intertemporal tradeoffs. The preeminent technique in such situations has been the simplistic application of a discount rate. Unfortunately, practitioners have frequently had difficulty in selecting the most suitable rates while, for some situations, the discounting methodology itself is inappropriate; and

4) Differentiating between ex ante and ex post benefits. Ex ante expected benefits are no more than the aggregation of possible ex post benefits weighted by their probabilities of occurring. Research funding decisions must be made on the basis of ex ante knowledge and, hence, the ex ante benefits. Evaluation of those decisions cannot be based wholly upon the ex post returns from the research but must take into account the uncertain set of possible results faced by the decision makers. The complex nature of these uncertainties we now consider in greater detail.

Additivity and Uncertainty

The methodological difficulties described above indicate that net value estimation--even for projects whose results are foreknown and require no further research before achieving their benefits--is not trivial. In fact, the direction of research monies must face such uncertainties as:

1) the broad spectrum of possible results that any project could lead to;

2) the broad choice of possible subsequent steps that might come under consideration for each of the many possible research results; and

3) a vast range of possible pathways by which a given research advance might eventually come to improve health care.[1]

The repercussions of basic laboratory research will be more difficult to estimate than those of the more applied clinical or epidemiological research. Thus, research on the mechanisms of cell division could only with great and perhaps prohibitive difficulty be evaluated on the basis of the specific diseases, disorders, or conditions which it could alleviate. Since such research can lay the foundation for so many diverse strands of subsequent research, it has frequently been termed "additive." On the other hand, the possible benefits of an epidemiological study for a specific disease can be imagined with much greater, if still far from perfect, precision. Those benefits, for instance, would most likely be upper bounded by all the good that could be done the population that has contracted or would contract

[1]The theoretically justifiable method for handling these factors, is set out in H. Raiffa (1968). Decision Analysis, Addison-Wesley, Reading, Massachusetts. Unfortunately, the extent and complexity of the uncertainties here render such techniques pragmatically inapplicable.

the disease.[2] It thus makes sense for grant administrators to prefer, ceteris paribus, epidemiological research into diseases with higher prevalence, incidence, and severity.

Having thus glanced at the theoretical framework that might be used to evaluate bio-medical research and at the lurking difficulties, we turn now to examine an instance in which a multidisciplinary analytic team sought conscientiously to apply that theory to actual research.

IN PRACTICE

The Evaluation

In 1971 the evaluation arm of the US Department of Health, Education, and Welfare engaged a team of experts from Harvard University to evaluate bio-medical research performed in Yugoslavia under the cooperative bi-national Special Foreign Currency Program (SFCP).[3] The Program, begun in 1961, consisted of 110 projects which were to be individually and summatively assessed. The Harvard team included one biochemist, one clinical researcher, one economist, one public policy analyst, one sociologist, and four physicians. Simultaneously, a similarly constituted Yugoslav team was commissioned to perform a parallel evaluation of the Program. The Harvard team convened frequently during 1972, held two meetings with the Yugoslav team, and, in a splurge of eleventh hour activity, completed and submitted its report in January 1973.

The Focus of the Evaluation

The commissioners of the evaluation had as their goal the construction and implementation of an evaluation model much like that described above. They were not, however, so naive as to believe this an easy task or even one likely of achievement. In the end they received a well-written document that made enough intelligent points that they could be satisfied. In only the loosest way could it be claimed that the Harvard study had made progress toward the original goal of program evaluation--in the sense of ascribing monetary values to the whole program or comparative values to different segments of it.

That the team had consciously striven to apply a model of the type above could be seen in the questionnaire laboriously developed:

a) four separate questions inquired into the project cost magnitudes;

b) one question requested an approximate estimation of the ex ante net research value while another sought the ex post value;

c) four questions elicited the importance, prevalence, and incidence of the disease, disorder, or condition in question;

d) nineteen questions sought information about different types of potential and actual impact; and

e) two questions, by requesting the estimated time to impact in health practice, both enabled finer classification along the spectrum between basic and applied research and provided necessary information for the discounting of benefits.

---

[2] An exception would occur when research into one disease leads to insights about another.

[3] Harvard University, An Evaluation Study of the Special Foreign Currency Program in Yugoslavia, DHEW Report HEW-05-71-188.

Results of the Evaluation

The Harvard team could not be faulted for lack of effort when one considers the conscientiousness with which it developed its questionnaire, tested it in plenary sessions, then applied it to the program. The degree of interdisciplinary harmony and cooperation astounded this observer. And yet, though the team succeeded in many ways, it failed in its central task of evaluation. Before examining the reasons for this, we should look to what the evaluation did achieve.

The evaluation team obtained two pools of information from which it would draw in writing its report:

1) three questionnaire reviews--two US and one Yugoslav--for each of the 110 projects provided by a total of twenty experts; and

2) a mass of documentation and interview reports upon the administration of the Special Foreign Currency Program.

From these sources, the report author extracted near maximal information value to achieve a polished and professional account of the study.

The research report provided:

a) intricate classification of all projects--by granting agency, by budget size, by type of research, and by probable specificity of impact;

b) summative commentary of a Delphi nature upon the program though provision and rudimentary interpretation of reviewer response to each of the questionnaire items--concluding, for example, that the total impact would benefit Yugoslavia more than the US and the scientific establishments of both nations more than their health care delivery systems;

c) a series of inferences based upon the classifications and reviewer answers--as, for instance, the finding through correlation analysis that US non-financial contributions were significantly linked with actual project accomplishments;[4]

d) a set of comments upon the administration of the program--the proposal review process, the role of the project officer, the publication policy--and how it might have been improved.

Overwhelmingly as a result of the skills displayed by the writer of the report, it: 1) has been accepted with praise by the evaluation arm of DHEW which funded it; 2) has become a necessary addition to the bookshelves of all administrators remotely connected with the SFCP; and 3) has been published commercially.[5] Notwithstanding these tokens of success, the evaluation seems, to this writer, in part a failure for having neglected to provide information that could beneficially have guided the management of the program. We now examine why.

---

[4] From this, the report drew the judgmental conclusion that non-financial contributions--such as the time of the project officer--effected better results. It lacked a sound statistical basis for ruling out the possibility that causation ran in the other direction: that the project officers tended to devote more time to projects that were developing well.

[5] Berry, R.E., Jr. et al. (1974). Evaluating Health Program Impact, D.C. Heath, Lexington, Massachusetts.

PROBLEMS OF ANALYSIS

Retreat from Quantification

The original version of the Harvard questionnaire attempted more precise quantification than was achieved in the final instrument. An example of the modifications made can be seen in two questions which originally asked for numerical estimates of the exact prevalence and incidence of the disease, disorder, or condition studied. The reviewers raised objections: they did not want to have to put down hard numbers that could perhaps be proven wrong. In concession to this sentiment, a new response scale was substituted enabling choice among such semi-quantitative[6] terms as "very high," "high," moderate," "low," and "very low." The change filed off one cutting edge of the questionnaire but the new softer and more rounded contours enhanced reviewer comfort.

The result of such modifications could be seen in the delicately balanced language of the final report. One question of highest importance was that inquiring into expected benefits in relation to project cost. The answers here clustered about the rating "great" with dispersion into the categories of "small" and "maximal." At first blush, this appears an important Delphi commendation of the program as the reviewers rejected the opportunity to be more negative. Unfortunately, the professional positions and outlooks of the reviewers made it extremely likely that their median assessment of any program with reasonably well-written reports and diligent researchers would impute a "great" ex ante net value. The program administrators should have expected such a rating.[7]

In this light, the utility of the evaluation is questionable. It becomes still more questionable when we consider a primary purpose of research evaluation: allocation of men and money to areas of greatest promise for societal benefit. On the assumption that our researchers will be diligent and will write able reports, we must seek to aim them in the most valuable direction. Any evaluation reflecting only their diligence and ability does not assist this purpose. If quantified estimates upon the net value of the research had instead been obtained, improved allocation of scientific resources would have been enabled.

A question gauging contribution toward US health objectives[8] provides another example of the consequences wrought by semi-quantitativeness. For each project, its best contribution toward at least one of the ten health objectives tended to lie between "some" and "maximal." The projects of one agency within the program had an advantage on this scale over those of another thus making a hard comparative judgment apparently possible. When, however, the greater average project cost within the former agency was considered, it was unclear which of the two had sponsored more beneficial research per unit cost. With quantitative estimations of benefit, this difficulty would have been avoided. Far from being an adventitious result of the semi-quantitativeness, it seems linked in a fundamental and causal way. The project reviewers may not have wished to quantify their answers precisely in order to hinder hard and unambiguous judgements that might have put researchers

---

[6]A term taken from the as yet unpublished work of H. Raiffa, who found substantial subjective disagreement upon the meaning of such expressions.

[7]In evaluative theory such findings might be valuable as a means by which top-level administrators could obtain information on program effects not biased by passing up through self-serving lower strata. With the SECP, the top administrator knew more before the evaluation started than Harvard was ever to learn. The report may, however, have been valuable in this sense to the Office of the Secretary of DHEW.

[8]The Yugoslav reviewers made parallel estimations on contributions toward the health objectives of their nation.

out of work and relegated their children to the streets.[9]

Disciplinary Roots

The majority of the Harvard reviewers were bio-medical professionals who could identify with personnel in the projects they examined. If the project leader seemed competent--from his record and his writing--the reviewer would be strongly guided by the motive of making the project look good. This inevitably inhibited wholly objective responses. The bulk of the reviewers were in the later stages of their careers, somewhat set in their beliefs, and not amenable to the idea that a new magic called systems analysis could benefit medicine. These factors reduced the utility of the questionnaire data.

The Concept of Value

The disciplinary rooting of the reviewers induced a number of conceptual difficulties. A typical example lay in their assessment of project worth. Though repeatedly instructed that research value should be perceived in terms of expected societal benefits, the reviewers resolutely rejected this guidance. If a project proposal was well written and assiduously executed, it was rated a good project regardless of its potential or actual consequences. When judging project worth against its budget, the reviewers could not be weaned away from their insistence on comparing the budget not to the probable social results but instead to alternative costs of obtaining the research knowledge.

Probabilities

Because the expected probabilities of research success varied across projects, we asked our reviewers to estimate the expected net value of the research. Thus a project with a success likelihood of 0.5 should, ceteris paribus, be funded before one with a likelihood of 0.1. Our reviewers insisted upon assuming project success and rating the research upon its maximum possible achievement with no consideration of the probabilities involved.

Two arguments were proffered for the neglect of success likelihoods:

1) that they were difficult to estimate; and

2) that the estimations would duplicate the efforts of the original scientific reviews prior to funding.

While the first objection cannot be gainsaid, the second seems but a specious cover for the first. Surely the original study section review was better qualified than Harvard to judge the probabilities of project success. Its approval may, however, have certified only that a project attained a necessary threshold level of success likelihood, say thirty per cent. In this case, no discrimination would be made between two projects with success likelihoods of thirty-five and ninety per cent. The discomfort of the Harvard team when faced with probabilities leads also to a more serious fear. It can be inferred that the study sections themselves--composed of men with similar backgrounds and outlook--were, most likely, similarly daunted by probabilistic concepts and sought to avoid considering their implications.

---

[9] As we shall argue below, quantified estimates--though attractive in theory--have problems of their own. The precise cost-benefit ratios of the US Corps of Engineers are examples of numbers whose precision is camouflage for unfathomable uncertainty.

Ex Ante and Ex Post Analysis

Through an error in the commissioning of the evaluation, the research program to be assessed turned out--to the surprise of both commissioners and evaluators--to consist overwhelmingly of ongoing projects. Thus a final impact evaluation from an ex post perspective was rendered impossible. Instead an ex ante estimation of potential impact--implying heavy duplication of the original study section reviews--was performed.[10] This seriously undermined the basic evaluative objectives.[11]

Shortcomings of Theory--Probabilities

The inability of the reviewers to incorporate probabilistic estimates into their project evaluations is matched by the inability of theoretical models explicitly to handle probabilities. There are important chance elements affecting the a priori value of research:

1) the probability that a research project will attain the goals it has set for itself;

2) the probability that the results of a successful research project will be adequately followed up by subsequent research to make possible practical benefits; and

3) the probability that research advances enabling practical benefits can be assimilated into the health care delivery system.

Potential gross benefits may be defined as the value of the research benefits to society assuming that its goals are attained, that adequate follow-up is forthcoming, and that its practical advances are not stymied by system resistance. Expected gross benefits are then the potential gross benefits multiplied in turn by each of the three probabilities above.

Of these probabilities, the most difficult to estimate is that of adequate follow-up. The pathways by which research begets research, in turn broadening knowledge and leading to societal benefits, are multifarious and unforeseeable. Prior predictions will never be able to foretell the serendipitous combination of men, moment, and the chance event that has led to many a research advance. Not only can inspiration and insight not be programmed, but they are likely to suffer when the attempt is made.

These problems proved too much for the Harvard team. The physician reviewers, sensing the magnitude of the problem, escaped it by treating likelihoods difficult to estimate as certainties. Those members of the team more comfortable in dealing with probabilities might have parsed the estimation problems into parts that the reviewers could more easily have handled. They might thus have obtained from the physicians a systematic Delphic estimation of research follow-up probabilities. Such an estimation might not consider all remotely plausible events but it would have enabled a more reliable judgment upon expected research value.

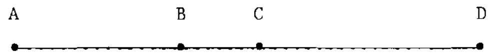
---

[10] This task required delicacy as the Harvard study was explicitly looking for mistakes of the previous reviewers.

[11] The ex ante evaluation from an ex post perspective required careful judgment by our reviewers. Decisions had to be rated not on how they turned out but on their reasonableness within the information framework of the moment. Our reviewers --perhaps because as physicians they have internalized an instinctive understanding of decision processes--had no trouble grasping this basic point in decision theory.

Additivity

The difficulties in estimating chances of research follow-up are heightened by the tangled additive relationships of projects with related goals. To illustrate, we suppose that research leading from the status quo ante of knowledge, point A, to practical value, point D, would achieve ten units of benefit. The research is broken down into segments A to B, B to C, and C to D as shown below.



Suppose now that it is necessary to gauge the value of traversing segment AB. There are no conclusive a priori reasons for any given set of value assignments. Thus, AB could be valued at eight units and each of the next two segments at one apiece. But BC similarly could be accorded a value of eight and the other segments values of one. If the minimal costs of traversing the segments were given, the difficult problem of value attribution could be reduced but not eliminated. If, for example, these minimal costs are AB, two units; BC, one unit; and CD, four units; it would make sense to conclude that the net value of AB lies between two and five; of BC, between one and four; and of CD between four and seven.

One common sense resolution would be to assign the given value ten to AD and to declare that further segmentation of the values is without meaning. The research should be performed only if the cost of achieving point D from point A is no greater than ten.

But this does not solve the reviewer's problem of valuing project AB. Should there be four alternative ways of moving to C from B, five ways of moving from C to D, two additional ways to move from A to C without passing through B, and the possibility of performing the research simultaneously in parallel, the problem becomes still more intractable.

In additive situations, both simple and complex, reviewers cannot be blamed for failing to specify a precise research value that, even conceptually, may not exist.

OBSERVATIONS AND CONCLUSIONS

Rating the Results of the Harvard Study

We have argued that the final report produced by the Harvard team, for all the kudos granted it, did not fulfill the primary goals of the evaluation. This because:

1) the average rating of project net values as "great"--being unquantified and allowing vast subjectivity of judgment--did not provide a useful summative assessment of the program;

2) the comparisons between different segments of the program--again hampered by the lack of quantitativeness and by subjectivity--did not indicate which should be pruned and which expanded; and

3) the examination of internal program operations did not yield information not already known to its administrators.

Yet the results were not wholly negative. If Harvard's comments about program administration were far from novel, they at least were given stature by their mode of publication and may have effected operational improvements. In a related way, the mere act of the evaluation seems to have energized the program--heightening the motivation of its personnel through a Hawthorne Effect--as it was being scrutinized.

Lessons for the Team

In the opinion of this observer, the value of the Harvard study lies in the lessons it taught about the application of systems analysis to a difficult area. As is frequently the case with new instruments, first trials teach more about the instrument itself than about the phenomena to be examined.

We have seen in the study defects that can be remedied in future applications of systems analysis. Harvard learned the conceptual stumbling blocks of its non-analytic reviewers and just how they gave rein to their subjective feelings. Certain questionnaire items were too comprehensive and difficult and should have been broken down to more tractable parts. In areas where embracing quantification was not easy, the reviewers should have been provided a simplified series of questions to coax better their judgments. Ways should also have been found to permit expression of their subjective feelings lest they complete the questionnaires with misgivings that the wrong questions were posed.[12]

Team Composition

The composition of the Harvard team, though impugned perhaps by implication in the commentary above, could not easily have been improved. The inclusion of several distinguished physicians--leery of numbers and the tricks of analysis--was essential. While the medical expertise of these men gave a necessary ballast and depth to the study, their stature ensured that the final report would be read with care. Any report penned wholly by slick systems analysts could easily have been discredited and disregarded in the very decision forums it was designed to serve.

The team brought together talented individuals whose personalities meshed well and who were devoted to building bridges between disciplines. Even in failing, their efforts indicated possible strategies for linking the disciplines across the gulfs that separate them. No failure whose effort was mighty is altogether without gain. The shortfall of Harvard indicated both limits to the purview of systems analysis and obstacles that need now to be overcome before its full worth will be known.

The Neglect of Variance

The resistance of the medical reviewers to the matrix forced upon them by the questionnaire may have derived from a fear that the numbers they were asked for could not have the importance or the infallibility attributed them by the analysts. In this, their reluctance may have had grounding. Too many mistakes of policy have been committed because the arguments for them could be expressed in numbers.

---

[12] Four questionnaire items designed to allow such ventilation of misgivings were insufficient.

A shortcoming common to both theoretical and practical exercises in evaluation is the neglect of variance. Few quantities called for by models for evaluating research can be estimated with confident precision. The judgments upon the unit values of life extension and of health status enhancement are especially subject to uncertainty. With the subsequent introduction of probabilities and of time discounts, the imprecision of the primary value measure--the expected net value of the research--mounted appreciably.

Any single expert called upon to consult in evaluating future bio-medical research will harbor his own uncertainties that can be roughly expressed as variances of the values estimated. Still another source of variance lies in subjective differences of opinion between experts. The absolute magnitude of the combined variance is so great that special systems methods for dealing with it are badly needed.

Dealing with Uncertainty in Guiding Research

In many areas of endeavor, the application of systems analysis at the managemental level has the effect of so ordering knowledge and expert opinion that perceived variance is minimized and that a small number of action alternatives is thereby shown to be distinctly superior. In the more applied areas of bio-medical research, this may be possible. Thus, systems thinking may indicate which epidemiological investigations and which clinical trials promise greatest potential benefit and should therefore be given priority.

In the area of more basic or additive research, this approach is judged impossible. The terrain is so dominated by the unknown and the uncertain that to attempt to draw a consensus from the disparate opinions of one's expert consultants is to misconceive the problem. Instead the entire range of uncertainty and disagreement should be examined to find appropriate opportunities for research. If ninety-five per cent of the experts judge a research path without promise in opposition to five per cent who believe it potentially important, it should not be automatically dismissed. Instead all research themes accorded a minimal probability of reward should be covered. How extensive the coverage would be would depend upon the magnitude of the potential reward and upon the probabilities estimated for success.[13]

These arguments imply that precise models incorporating the considerations of the second section cannot be applied to basic research. The predominant mechanism for passing upon the funding of research in many countries has been that of scientific peer review. This mechanism is not perfect: scientists may be motivated by the desire for intellectual stimulation or be guided by the inertia of their training and past research to the neglect of potential social benefits. Nevertheless, peer reviews do ensure that the work maintains high professional standards and that the knowledge sought is, by scientific standards, important. The greatest potential for improvement in basic research guidance thus seems to lie in modification of the scientific peer review groups rather than in the

---

[13]Problems of this nature arise in connection with the guidance of national cancer programs. See, for instance:

The National Cancer Program--Report of the Director, January 1973, DHEW Publication Number (NIH) 74-472;

The Strategic Plan, DHEW Publication Number (NIH) 74-569; and

Multilevel Analysis of NCI Research Grants by Scientific Category--Fiscal Year 1972, prepared by J.H. Schneider of the National Cancer Institute.

imposition of a comprehensive model that attempts to trace knowledge advances through long and tortuous paths to ultimate social benefit.[14]

The Need for Systems Analysis

One might take the Harvard study as cause to despair of using systems analysis for guiding bio-medical research. This would be premature and wrong. Governmental and foundation sponsorship of research will continue and the donors will be besieged with more requests for funding than they can meet. Every time one request is favored over another, external direction of research takes place. We may permit this direction to be inconsistent and random or we may invoke analysis to systematize its workings. Conscientious application of systems analysis--which requires recognition of its deficiencies--still holds promise of better guiding bio-medical research toward societal benefits. In the world of the unknown, such guidance cannot be precise and without turnings, but it can reduce duplication, fill lacunae of effort, and cut short unpromising endeavor. For these reasons, systems analysis should be strengthened and modified to deal better with the idiosyncratic problems posed by bio-medical research. The critique presented by this paper has sought to aid that adaptation.

---

[14] Similarly, the scientists themselves should be allowed to select the tactical approaches toward the problems they face. In the UK and the US, the trend in the last decade has been to fund grants primarily on the basis of scientific merit. Overall strategic redirection of research monies occurs in the cases of major campaigns--such as those against cancer or sickle-cell anemia--and in the selection of projects considered marginal in terms of scientific merit.

Observations of the Discussant, Mr. Spencer

Mr. Spencer recently had participated in an exercise to gauge the value of bio-medical research that resembled closely the project described by Mr. Thompson. Their group too had developed a model to quantify the research benefits that was remarkably similar to the model exhibited in the paper. The major point of difference was that Spencer's group found it desirable to include an additional term for benefits that result solely from the funding of the research. An example is the political gain in allocating research funds to any area salient in the popular mind.

From his experience, Mr. Spencer drew a number of conclusions:

1) that systems analysis provides a set of quantified relationships that may be aggregated into a model;

2) that such models can be aids to decision making but can never be replacements;

3) that a primary result of cost-benefit modelling is its improved statements of program goals; and

4) that a danger in such modelling is the comparison of non-comparables. In their quest to reduce all important factors to numbers, cost-benefit practitioners sometimes are excessively zealous. Comparisons of different benefits often depend critically upon value assignments. Since these will vary widely among social groups, analytic reduction of the benefits to numbers inevitably entails subjective judgments.

Hard Models and Soft Phenomena

Mr. Koch-Weser was a team member of the project described by Mr. Thompson. He concurred in most of the judgments made in the Thompson paper but felt that the criticism there was perhaps too gentle. He would formulate the vital failing of the project as its attempt to apply hard analytic methods to data--subjective judgments of achievement and progress--that were intrinsically soft.

The Communications Gap

Mr. Venediktov favored the use of systems analysis to evaluate bio-medical research. He, however, found the model as laid out difficult to understand. He urged that such models be carefully presented so that doctors and other non-analysts would more readily comprehend them. As did Mr. Spencer, he felt that a chief problem in applying such a model lay in quantifying the non-quantifiable.

Advice for Model Application

A participant with experience in related research efforts offered his observations. He felt that formulas provided in Mr. Thompson's paper could be clearly understood by non-technical experts if a series of carefully structured discussion sessions were held. Such sessions should be held before the analysis itself is initiated. He felt that models of this type were more effectively applied to lines of research activity composed in individual projects rather than to the projects separately. He was pleased that the model distinguished between gains in reducing morbidity and in extending life. Recent data show that the number of morbidity episodes has no effect upon life expectancy.

Mr. Spencer agreed wholeheartedly that great efforts should be made to achieve an initial meeting of minds. He related that one program evaluated was attributed 350 different effects. Extensive discussion was required to coalesce these into forty-six effects which could more easily be understood and evaluated. The effort required to achieve interdisciplinary harmony and mutual understanding can take years.

Scientific Merit Evaluation

One speaker described an evaluation of bio-medical research that compared the original scientific merit rating with an ex post merit rating, and with the numbers of citations relating to papers arising from the project. Correlations across all three indices were high. Mr. Thompson felt that this should be expected since all three measures were based on scientific judgment of scientific merit. The study he described was more difficult in that it sought to link ascribed scientific merit to foreseen social benefits.

Further Remarks upon Cost-Benefit Analysis

Additional observations upon cost-benefit analysis were proffered:

a) The importance of defining benefits at the initiation of the study was stressed.

b) An area of great confusion is the application of time discounting. In bio-medical research this is critical as there may be a substantial lag until benefits are realized whereupon they may be spread over decades.

c) Benefits can be broken down into the categories of direct, indirect, and intangible. Both of the former two classes can generally be calculated, but the greatest difficulty arises in estimation of intangible benefits which often are highly significant.

d) The limitations of cost-benefit analysis are indicated by frequent selection of projects with poor ratios over those with higher ones.