PERFORMANCE AND OUTPUT MEASUREMENT: A JOINT MEETING OF EURO PUBLIC SECTOR AND HEALTH WORKING GROUPS 14-16 January, 1980

Peter Turner
Duncan Boldy
Philip Aspden
Editors

April 1980
CP-80-7

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

PREFACE

Most governmental and public service bodies have developed methods for measuring and controlling the inputs to major programmes, in terms of finance, manpower, and other resources. There has, however, been relatively little success in measuring the output of such programmes--except in rather special cases--and such measurement as takes place is usually on secondary rather than primary measures of output; it is easy to measure the number of patients handled in a clinic, but much more difficult to determine how far their health is improved. The issue of performance and output measurement is thus of major concern to operational research workers and applied systems analysts, much of whose work is directed towards public service programmes, and because it cuts across many fields of application. It therefore seemed particularly appropriate that two working groups of the European Association of Operational Research Societies should meet at the International Institute for Applied Systems Analysis in January 1980 to discuss this topic. As the discussion may be of wider interest than to those actually present it all seems appropriate to make these notes more widely available in the form of an IIASA Collaborative Paper.

The meeting was the prime responsibility of Peter Turner and Duncan Boldy, chairmen respectively of the Public Sector and Health Working Groups of EURO. At IIASA the coordination was undertaken by Philip Aspden of the Health Care Systems Modelling Task. These three are the joint editors of this report.

It was felt that the proceedings were worth recording but that, as they were really part of an ongoing debate, they should be made available as quickly as possible. To avoid further delay some of the papers are therefore presented in note form, and the discussion has been condensed to bring out the main topics of

interest. It is too seldom that those engaged in the analytical study of policy issues get together specifically to discuss their problems, rather than to parade their solutions. We hope that this collaborative paper may help to stimulate further such meetings.

Rolfe Tomlinson
Chairman, Management and Technology Area, IIASA
President Elect,
European Association of Operational Research Societies

April 1980

CONTENTS

On cost grounds it has not been possible to reproduce all of these papers in a standard format nor with consistent paragraphing. But the pages have been renumbered to provide a straight run through the whole report and we hope this will provide sufficient reference for readers to find their way in it.

SECTION 1

INTRODUCTION

1. On 14-16 January 1980 IIASA hosted a joint meeting of two EURO working groups, Public Sector and Health, on the topic "Performance and Output Measurement". The objective of the meeting was to bring together members of these two separate working groups and staff at IIASA working in similar areas and to discuss and make recommendations concerning the difficult yet common problem of "Performance and Output Measurement" as applied to their field(s) of interest.

2. The meeting was divided into four main sections:-

(i) Brief reports on some of IIASA's work relevant to the theme of the conference.

(ii) Invited discussion papers on Performance and Output Measurement.

(iii) Parallel discussion sessions; one Health syndicate and two separate Public Sector syndicates.

(iv) Reports from the syndicates, final address and general discussion.

All four invited papers are reproduced in this report, together with commentaries and the reports of the three syndicates.

3. There were a total of 42 participants representing 10 countries, of whom 19 were from the Public Sector Working Group, 13 from the Health Working Group and the remaining 10 either invited or from IIASA. A complete list of the participants is given in Appendix 1, and the agenda for the meeting is at Appendix 2.

4. No rigid structure was imposed on how the syndicates should function, except that each was required to elect a chairman and a raporteur. A common set of questions (discussion notes, reproduced as Appendix 3) were prepared beforehand, to be used by each syndicate in whatever way they wished. One of the syndicates chose to structure its report in terms of these questions whilst the other two used them as more of an aide memoire. Each syndicate met for two sessions lasting a total of about 5 hours.

SECTION 2

INTRODUCTION TO IIASA

THE WORK OF THE INSTITUTE

1. Rolfe Tomlinson said that the idea for the Institute was proposed in 1966, when the then-President of the United States, Lyndon Johnson, suggested that an Institute might be created to work on the common problems of the developed nations, and as such to serve as a bridge between East and West. In early 1967, he sent a representative to discuss this idea with officials in the Soviet Union. Agreement was quickly reached that the Soviet Union and United States would jointly propose the establishment of such an Institute. There were then five years of negotiations involving an increasing number of nations, until by 1972, 12 nations had agreed to participate in the establishment of this Institute.

2. An important principle was that although the funding came from national governments the Institute would be free from direct governmental intervention. For this reason its funds are channelled through, and its programme is managed by, independent scientific agencies like for example the Royal Society in London. There are seventeen national member countries: the USA and USSR each pay a major share of the costs and the remainder is provided by the other member countries. Last year the turnover of the Institute was $12 million.

3. The work programme of the Institute consists of two kinds of problems of international importance: Global issues, which inherently involve more than one nation and cut across national boundaries; these include preservation of the global climate, exploitation of the oceans, assuring that mankind (whose population will double over the next fifty years), will have sufficient energy, food, and basic services to survive and to live a satisfying life.

IIASA is one of the few places in the world, perhaps the only place, where scientists from many nations may gather in a non-political environment to share their understanding of these global issues. There are also universal issues, which lie within national boundaries; these include, the design and implementation of the Health Care System, and the preservation of the environmental quality of a city or region. These are national concerns, but they are international too because all nations share them; and IIASA, having on its staff, scientists from nations with quite different social, economic, and political characteristics, can play an unique role in the exchange of experience across these boundaries. The teams that work on these problems are mixed both in nationality and by discipline.

4. Although the title of the Institute emphasizes Systems Analysis, it is as <u>scientists</u> that its staff are primarily recruited. It is the intention that the work is <u>Applied</u>; the aim is to make an impact in the field.

5. Impacts might be in terms of increased understanding, or in information, or in working procedures, although in the latter case it is not the intention of IIASA to specify procedures for particular governments in particular cases. Rather, the aim is to find generalisable results. These results are promulgated by publication, by visits both long and short (eg secondments), and discussion (including conferences), and by building a network of contacts with interested parties.

6. The Institute tries to adopt a comprehensive or multi-disciplinary approach to its research. This has lead to a two-dimensional organisation of the Institute's work. One dimension is four Areas; each Area is concerned with a specific aspect of human experience or knowledge important for the studying of international problems. The Resources and Environment Area, is concerned with the earth's natural endowment, with its resources -

water, mineral resources, land, and with the environment - air, land and water environment, and with the global climate. The Human Settlements and Services Area, is concerned with the earth's human endowment - with its people, the way they are distributed on the globe, the services they need - such as health care and education, transportation and housing. The third Area is Management and Technology, and is concerned with man-made contributions to the global endowment. Its interest is in organisations and technologies; the Area comprises specialists in management science, in engineering, and so on. The fourth Area is System and Decision Sciences, containing methodologists, mathematicians, computer specialists, economists; all concerned with how one analyses complex systems and complex decisions.

7. The second dimension of the Institute's work are cross-cutting programmes: the Energy Programme, which is concerned with the development over the next fifty years of the global energy system; and the Food Programme, which is concerned with the development in the near decades of the global food production and distribution system. In addition, there is a General Research Programme. This acts as a seed bed for new programmes; a home for cross-cutting activities too small to be a programme, such as the Survey Project (currently producing a draft handbook of systems analysis); and as a place for miscellaneous activities, such as the Global Modelling conferences. Figure 1 lists in each Area and Programme the research Tasks that each is carrying out. The basic unit of research is a Task, with a leader and three or four scientists. The titles next to the dots in the figure are the names of the Tasks. There are about twenty-five Tasks. More details of these Tasks can be obtained from the Institute.

HEALTH CARE SYSTEMS MODELING AT IIASA

8. This was a combined address given by E Shigan, P Aspden, and P Kitsul, of the Health Care Systems Modelling Task within the Human Settlements and Services Area at IIASA. A paper summarising Health Care Systems Modelling at IIASA was given by Mr Kitsul at a meeting of the Health Working Group at Baden, Switzerland July 1978 (attached separately for members of the Working Group on OR in the Public Sector). The work of the Task is also summarized in "Health Care Systems Modelling at IIASA: A Status Report" by Shigan, Hughes and Kitsul, obtainable from the Institute.

9. Professor Shigan opened the talks: he said that in most countries there is a need to change the health care systems, which will depend on many other factors external to them. Whereas previously the custom had been to think narrowly about health only, and work had been concentrated on the hospital level, it was now realised that there are additional problems to be resolved at regional and national levels. It was also realised that the objectives of the health care system were not clear-cut and that some of them conflict. The problem breaks down into a multiplicity of sectors and conditions, and there is much uncertainty.

10. Within a framework of the functional structure of a public health system (Figure 2), the work at IIASA is concerned with a number of models (Figure 3) supported by data on demography, health status, and resource availabilities. The system of models such as illustrated in Figure 3 is most appropriate to countries with national health care systems.

11. IIASA's objective is not to prescribe health care systems but rather to look for common problem areas and develop approaches to them which may be of wide applicability. In each country they take care to work with local health managers who provide information and sometimes resources, but above all the local knowledge which

enables them to select from the work done to their own maximum advantage, employing parameter values appropriate to their own conditions. It was notable that in most countries there would be some aspect or other of the health care system for which it is difficult to collect information, but that information from other countries would normally be available to provide estimates to cover the gap. In this way the power of the work done in each country was enhanced.

12. Philip Aspden described the Disagreggated Resource Allocation Model (DRAM). He explained that in most countries there is a shortage of health care resources: demand always increases to meet any conceivable level of supply. In that case the problem is now to allocate these scarce resources. In practice in most countries resources are in fact allocated at the lowest level, eg by local or hospital doctors, and the purpose of DRAM is to simulate their preference function.

13. In DRAM it is assumed that the Health Care System allocates its resources in a way that appears to maximise a utility function whose parameters can be inferred from observation of past allocations. The model represents the actors in the HCS striving to attain some ideal pattern of care within resource conttraints. DRAM has been used to analyse data from UK, Canada and Czechoslovakia.

14. Pavel Kitsul briefly covered some work on Morbidity models. Some existing models are concerned with aggregative morbidity rates, infectious diseases and terminal degenerative diseases.

15. Where direct morbidity data is difficult to collect it is possible to use mortality rates as a proxy.

16. For infectious diseases, which have a short duration, dynamic models can look at the equalibrium states between the conditions of morbidity, mortality and recovery.

17. For terminal degenerative diseases, where the data can give age distributions for specific mortality and general mortality rates, and population age structures, it is possible to work with survival curves to examine the dynamics of the disease.

MONITORING OF HEALTH CARE SERVICES

18. Mark Cantley focussed on the opportunity we now have to look at the effect of new strategic planning policies within the UK Health Care System. He summarised the present system consisting of a first cycle concerned with the issuing of a consultative document, a Planning Manual and Planning Guidelines, and a second cycle at Area Health Authority level, interpreting national guidelines, augmenting the planning manuals, setting out information formats and so on. The question now is, how will the plans just made be monitored and controlled?

19. At the moment only the resources consumed and intermediate outputs (like numbers of beds) are measured. The discrepancies which arise, when the planned resources are applied to actual need, should be fed back into the planning cycle to adjust the plans made.

SYSTEMS ANALYSIS IN REGIONAL PLANNING

20. Mr Andersson gave examples of applications for Regional Planning:-

a. Long-term regional development.

b. Regional consequences of industrial (including agricultural) change.

c. Planning the location of indivisible production units under interdependency.

d. Inter-regional growth analysis.

e. Inter-relation between the private and public sectors in a spatial growth process (applied particularly in the Research and Development (R&D) field).

21. His talk would concentrate on a. and e. In each case the analysis would be "top-down" because each region is only very small indeed compared with the rest of the world and would be dominated by that.

22. On Regional Development, (see Figure 4) he talked about the inter-relationships between prices, investment and growth at the national level; population and space at the inter-regional level (using a model called "MIRROR"); land use within regions (using a model called "MALTOS"); and for public services, aspects of transportation network behaviour. The objective is to maximise the rate of growth, with other factors in balance, and determining the best trajectory from the current rate of growth to that best position.

23. Applications are being made in Bulgaria (on agriculture and industrialisation), in Sweden (environment and the economy) and in Italy (on labour markets and physical planning). The model is used not to produce specific plans but to provide guidelines for decision making, perhaps qualitative rules, and perhaps some quantitative recommendations of a more general kind.

24. For the inter-relation between the Public and Private Sectors, Mr Andersson focussed on the allocation of the savings ratio to R&D and material investment. For both Public and Private Sectors, along with Production (which is determined by capital stock), the savings ratio will determine the increase of R&D and

material capital stocks of the regions. Interaction between public and private sectors comes from the rate of taxation, which supports one at the expense of the other. The Public Sector R&D cannot be regarded as uniform across regions: the capability in any region is effected by its neighbours, the effect being attenuated as some function of spatial separation.

25. The model, which was not completely described, provides the opportunity to explore the effects of changing levels of production, taxation, savings ratio and so on, over the extended periods which were necessary for the consideration of activities on a regional scale. (See Andersson A.E. 1979. "Growth and Stagnation of Economies with Public Goods - A Neoclassical Analysis". WP-79-12. International Institute for Applied Systems Analsyis, Laxenburg, Austria.)

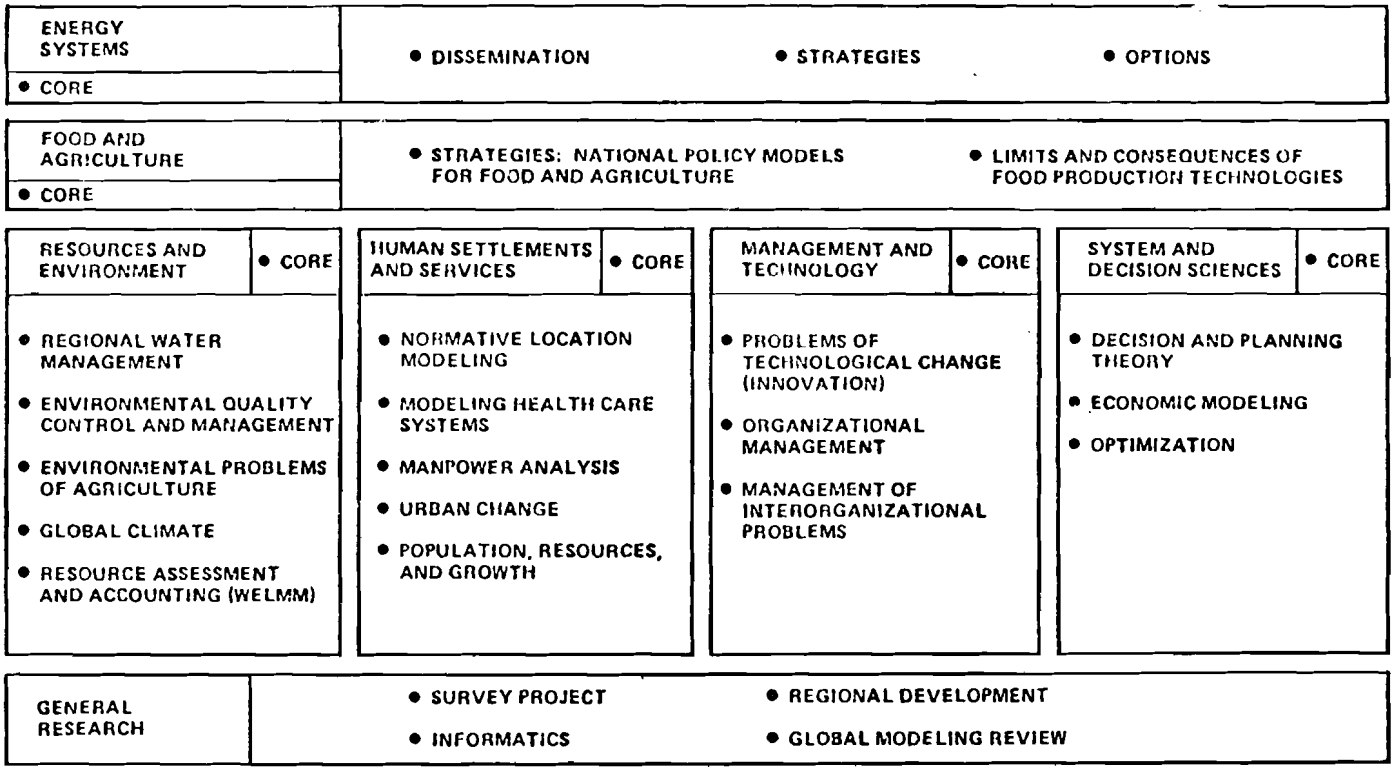

ENERGY SYSTEMS
• CORE
• DISSEMINATION
• STRATEGIES
• OPTIONS
FOOD AND AGRICULTURE
• CORE
• STRATEGIES: NATIONAL POLICY MODELS FOR FOOD AND AGRICULTURE
• LIMITS AND CONSEQUENCES OF FOOD PRODUCTION TECHNOLOGIES
RESOURCES AND ENVIRONMENT
• CORE
• REGIONAL WATER MANAGEMENT
• ENVIRONMENTAL QUALITY CONTROL AND MANAGEMENT
• ENVIRONMENTAL PROBLEMS OF AGRICULTURE
• GLOBAL CLIMATE
• RESOURCE ASSESSMENT AND ACCOUNTING (WELMM)
HUMAN SETTLEMENTS AND SERVICES
• CORE
• NORMATIVE LOCATION MODELING
• MODELING HEALTH CARE SYSTEMS
• MANPOWER ANALYSIS
• URBAN CHANGE
• POPULATION, RESOURCES, AND GROWTH
MANAGEMENT AND TECHNOLOGY
• CORE
• PROBLEMS OF TECHNOLOGICAL CHANGE (INNOVATION)
• ORGANIZATIONAL MANAGEMENT
• MANAGEMENT OF INTERORGANIZATIONAL PROBLEMS
SYSTEM AND DECISION SCIENCES
• CORE
• DECISION AND PLANNING THEORY
• ECONOMIC MODELING
• OPTIMIZATION
GENERAL RESEARCH
• SURVEY PROJECT
• INFORMATICS
• REGIONAL DEVELOPMENT
• GLOBAL MODELING REVIEW

Figure 1

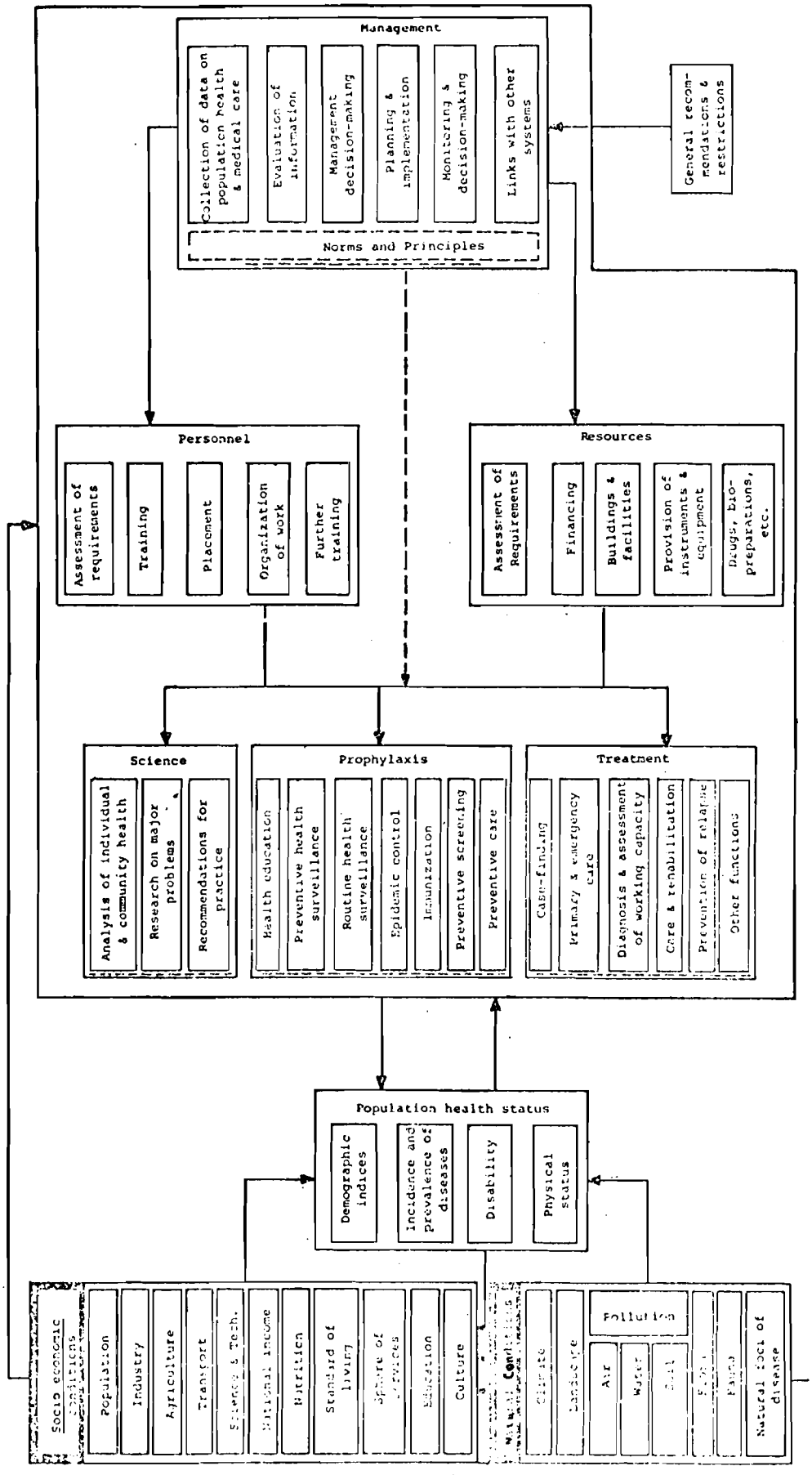


Figure 2 Functional chart of a public health system.

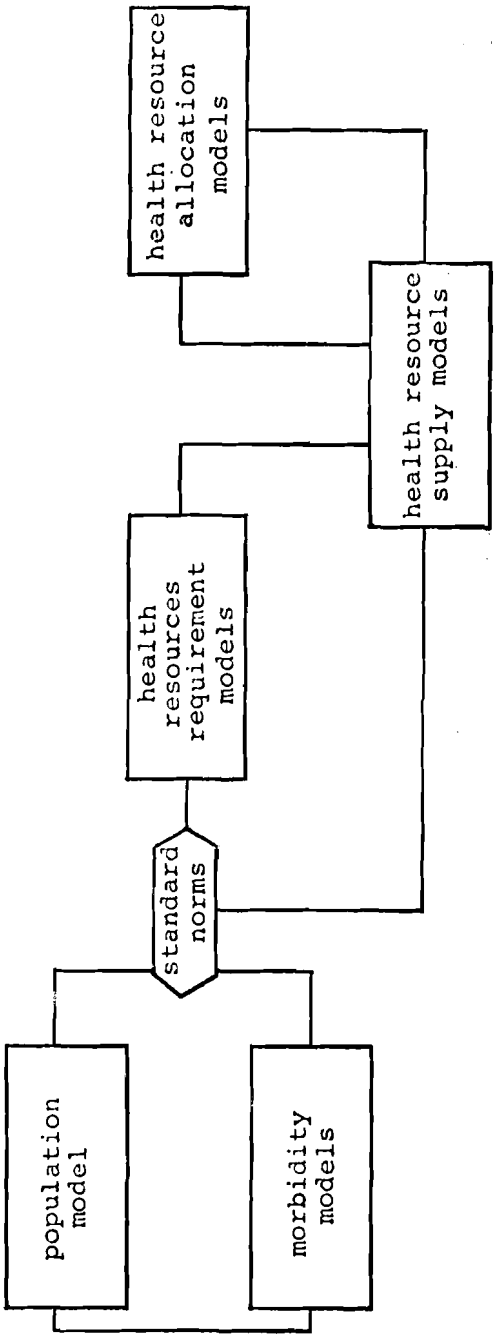


Figure 3 Composition of the IIASA National Health Care System Model (in 1978).

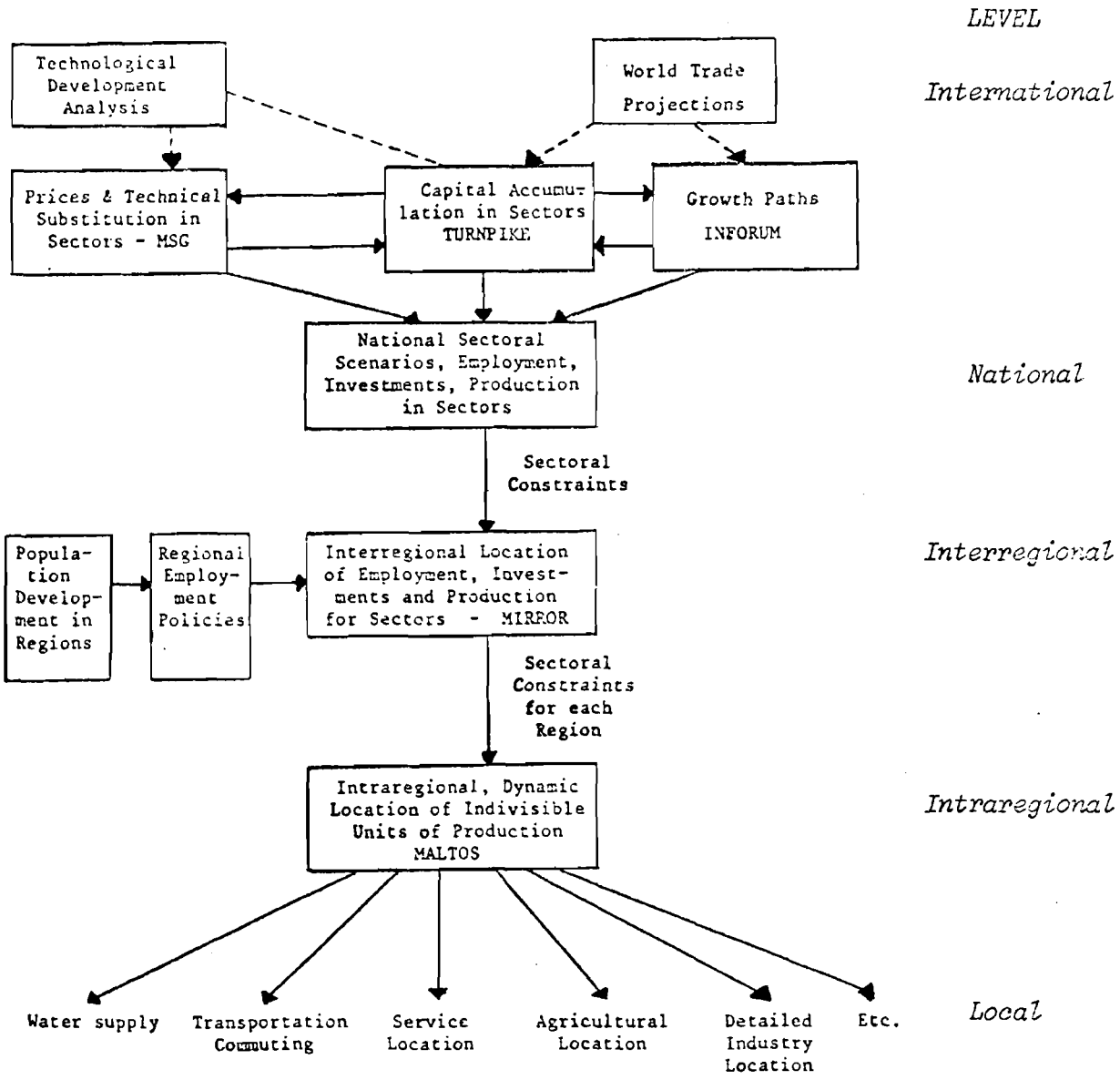


Figure 4 The Top-Down Approach to Regional Systems Analysis.

SECTION 3

Outcome Measurement: A U.S. Viewpoint

Roger E. Levien
Director, IIASA

I shall begin by talking about the history over the last few decades of the "analytical crafts" in the USA. I do not know whether "analytical crafts" is a common phrase. I will explain its relevance later. However you will recognise these three terms:

Operations Research
Systems Analysis
Policy Analysis.

There are obviously differences of opinion about the definitions of these terms. For today, I propose to define them historically. Operations Research was developed in the 1940's particularly out of the experience of the second World War. Here, the principal problem was how to design an operating procedure when the system, policies and goals were all fixed or specified, e.g., deployment of radar. In the 1950's, Systems Analysis evolved, particularly at The Rand Corporation, when it was recognised that for future choices it was not necessary to assume that the system was given. Rather, the purpose was to design the system, given a policy and a certain set of goals. In the 1960's, as we moved form the military questions, which gave rise to systems analysis, to public or social policy questions, we realised it was no longer the construction of physical hardware systems that was the key issue, but rather the specification of policies. This gave rise to Policy Analysis, in which the key choice was to determine the appropriate policy given a certain set of general goals. So the sequence in the US was from highly constrained optimization in Operations Research, to more flexibility in Systems Analysis, and still greater flexibility in Policy Analysis.

I am aware that the members of EURO might claim that this whole group of activities fits under the heading Operational Research. Those of us at IIASA tend to say that they fall under the rubric of Systems Analysis; and my home institution, The Rand Corporation, has for the last decade or so been using the phrase "Policy Analysis". I want to blur these distinctions, and just say that we should call these activities the "analytical crafts" and recognise that they lie along a continuum. In addition, we should recognise that the purpose of all these activities is to help decision makers make choices.

I would now like to consider the phrase "analytical crafts". Some people would like to think of us as having a discipline, some a profession, even some,a science. In my view, however, we deal rather with an applied activity whose purpose is to provide a useful product (think of a potter who is trying to produce a useful product), as distinct from an artist whose purpose is not use. In trying to produce a useful product we are drawing from many different fields of knowledge, many different sciences, as a potter might draw upon knowledge of ceramics, glazes, colours, etc. We transfer knowledge from one generation to another generation not only by text books, but through a system of apprenticeship. So it seems there is much in common with what we do and what a craftsman does. In this lecture, I want to talk about a specific aspect of craft knowledge, and that is, how one measures outcomes in these analytical crafts.

Before turning to this, let me begin a process I am going to follow in the rest of the talk. This is to make specific what I have said in general by turning to a particular set of examples in Health Care Systems. Since many of you are from the Health Care field, I do not think I will be telling you anything new, but I will perhaps make clearer what I am saying by giving examples in fields of your interest. What I mean by Operations Research, Systems Analysis and Policy Analysis in Health Care Systems are as follows:

Operations Research

-- scheduling of hospital facilities
-- operation of Blood Banks
-- ambulance dispatching;

Systems Analysis

-- design of a Health Care System for the elderly, combining hospitals, clinics, and home care.
-- design of an emergency care system;

Policy Analysis

-- determining the provisions of a National Health Insurance scheme,
-- the determination of the right mix of public and privately based care for the elderly.

Again I want to emphasize that these types of analysis fall along a continuum. I do not think it necessary to make sharp distinctions. My point is that they are all analytical crafts and they all face similar problems, most particularly the fundamental questions of outcome measurement. I would say there are two of them:

-- what are the outcomes of a specified policy, system, and mode of operation?
-- how well do they meet the goals of the interested groups?

I would like to emphasize that we are only rarely dealing with single decision makers. In most cases we are dealing with multiple decision makers and affected groups, with not all of the latter participating in the decision process.

I want now to turn to a series of classical issues in outcome measurement. For each issue, I will illustrate the general principles with reference to Health Care Systems.

1. DEFINITIONAL ISSUES

The first set of questions we all face in outcome measurement are what I might call definitional questions.

(a) Are the goals of the decision maker measurable at all? Frequently they are not, and the analyst must translate an abstract and vague goal into something that is measurable, usually through a proxy measure. However, this proxy measure, may introduce problems because

it is not a perfect measure of the goal. Therefore, at the end of the analysis we must make sure that we are not drawing some improper conclusion.

(b) Are the goals of the decision maker feasible? Often an analyst points out to a decision maker that his goals are infeasible. In such cases, the analyst may contribute by seeking a revised goal that is feasible to obtain.

(c) Is it desirable or appropriate to decompose our goals into subgoals, which can be quantified and lead to sub-analyses? This is another way of talking about sub-optimisation. This is sometimes essential in order to find quantifiable goals, but we are also aware of some of the deficiencies of sub-optimisation. For example, we may be moving towards objectives that are only partial, not taking into account the larger system's goals.

I will now illustrate these points by considering Health Care Systems:-

(a) Goal Measurability - Most Health decision makers would like to say the purpose of their system is to improve Health Status. But how is Health Status measured? This is still an open question and an adequate quantifiable measure has yet to be produced. We then move to the next stage and suggest a proxy goal such as "Reduce Mortality". This is measurable, but we must recognise this is only part of the whole question. We can also suggest an intermediate goal such as "Improve Health Care". We do not know the relationship between Health Care and Health Status, but care is something we can control. Sometimes it is even something we can measure, if we use some proxy measures such as physician hours per patient, beds per patient and so on. We must remember that these proxy goals may lead to sub-optimisation.

(b) Goal Feasibility - In Health Care, decision makers might like to provide the necessary care for all.

However, much analysis has indicated that the demand for health care exceeds the supply at least for the levels of supply that are feasible. Thus the analyst might encourage the decision maker to adopt alternative goals, for example, to provide equal access to care for all, or, to provide equal quality care for all.

(c) Goal Decomposibility - We are well familiar with the question of taking a goal like "improving health status" and reducing it to specific subgoals which are then sub-optimised, for example, "To reduce infant mortality", "To reduce the number of work-days lost". Each of these subgoals can then be worked on, and a policy defined to achieve the subgoal.

2. CONTEXTUAL ISSUES

Now let me turn to a second set of issues that always arise in outcome measurement, and these I have called contextual, by which I mean: what is the context within which the goals lie?

(a) Goals - In the early days of Operations Research we were happily working with single goals. However, in almost all situations, the real issue is that we have multiple goals. In the last decade this has become widely recognised, and analytical techniques are being developed to handle multiple goals. We must also consider whether these goals are complementary or conflicting. For instance, IIASA has three goals, to improve international co-operation, to advance science, and to apply systems analysis techniques to real international problems. If I tried to optimize a particular goal, then the other two goals might be jeopardized. So some balanced approach to these three goals is necessary. Further, these goals have some complementary aspects, for example, to advance science will help fostering international co-operation.

(b) Decision-makers - Again, in the early days of these crafts, we tended to have a simple model in mind

in which there was one decision maker. Now, increasingly we recognise that decision makers and, more importantly, the interested groups (not all of whom are necessarily decision makers) are multiple. We have to find ways of dealing with the collaborating and conflicting nature of their interactions. This is a feature of a study IIASA is doing on water resources in Sweden, where it is clear that the municipalities which draw water and which have to collectively provide the funds to make the water available , have both collaborative and conflicting relationships. Techniques derived from cooperative game theory are beginning to find their application in dealing with questions that ask not what the optimal decision is , but how to divide the benefits and costs among multiple decision makers.

(c) Time periods - Finally we recognise that we are no longer dealing with single time periods, but more generally with multiple periods, where there are trade-offs between the various periods. Sometimes this is handled by the use of market or social discount rates. Frequently, however, we face issues in which the discount rate does not seem to be an appropriate tool; for example, how do we deal with trade-offs across generations?

Our crafts must develop techniques for dealing with multiple conflicting goals, multiple conflicting decision makers, and conflicts over time. Let us now consider these difficulties within the context of Health Care Systems.

(a) Multiple, Conflicting Goals - For instance, improving care for in-patients, improving care for out-patients and reducing costs are conflicting goals.

(b) Multiple, Conflicting Decision Makers - Multiple conflicting goals are frequently associated with multiple conflicting decision makers. It may be that the hospital administrators want to improve care for in-patients, while the local community is interested in

improving care for out-patients, and finally the financial authorities may want to reduce cost.

(c) Multi-period Conflicts - In Health Care Systems, there are multi-period conflicts, which might take the form of the trade-offs between allocating resources to treatment of a particular disease or to prevention of this disease. Another trade-off would be between research and care for a particular disease. These are both inter-period trade-offs.

3. STRUCTURAL ISSUES

Under structural issues I propose to discuss the following "bread-and-butter" topics.

(a) Benefit Measures - How do we measure benefits? For example, we can do it by direct performance measures, e.g., the number of patients per hospital bed, or the number of lives saved. We could translate these direct performance measures into monetary values. Alternatively, we could take the decision analyst's advice and apply some utility function to the outcomes. Circumstances might suggest which one of these approaches is the most appropriate.

(b) Cost Measures - In some general sense costs can be considered as benefits foregone, either directly or indirectly. Alternatively, we could ascribe as we normally do, a monetary value to them. This is relatively easy when it is the cost of resources, but when it is a disbenefit like pollution or danger to health, then it is very difficult to assign a cost measure. While it is not frequently done, we could also assign a disutility or a utility measure to costs.

(c) Summary Measure - What will be the summary measure we are interested in? Will it be the difference between the benefits and costs, or the ratio of benefits to costs? A popular technique at The Rand Corporation is to maximize the benefits for a given

cost or alternatively, to minimize costs for a given benefit. The circumstances may indicate which is the appropriate approach. The problems associated with multiple benefits and costs are clearly going to be much greater than the simple approaches mentioned above.

I will now consider these structural issues for the Health Care Systems.

(a) Benefit Measures - These could be the reduction in deaths, the monetary value of the lives saved, or the utility of the lives saved.

(b) Cost Measures - These could be such general measures as the land, labor or capital allocated. They could even be the air and water pollution, or in a hospital setting, they could be the dangerous wastes or the risks from radiation exposure. Expenditures in the Health Care System are the usual cost measures. Theoretically, one could deal with the disutility of the benefits forgone.

(c) Summary Measures - These could be the
   -- value of lives saved less the expenditures,
   -- the ratio of value of lives saved less to expenditure,
   -- lives saved for m dollars, or expenditure to save n lives.

Let me now consider the styles of analysis within the analytical crafts currently popular in the U.S. I will characterize them in the following way:

(a) System Design/theoretical

Here we have to anticipate how the system will function, and we have to make theoretical estimates of outcome. This is classical systems analysis. In the US, a very popular form of this is environmental impact assessment, arising from congressional requirements to assess the impact on the environment of major construction projects. This has given rise to a big analytical business which has generated large amounts of information. Too much, in fact, and IIASA has made proposals on how to simplify environment impact assessment.

(b) Operational Design/empirical

Here the problem is how to improve the design in the case where the system is already functioning. I would call this Operations Research. We have access to empirical data and, therefore, the outcome measurement can draw upon actual outcomes, rather than the theoretical outcomes that result from modeling. Much work of this type in the US is done under the heading of "Program Evaluation", which is mandated by Congress in certain circumstances. This mandate gave rise to another group of analytical firms specializing in evaluation of public programs. Such analysis raises the question of how one measures outcomes empirically as distinct from theoretically.

(c) Policy Design/empirical

In the US, dissatisfaction with program evaluation as a way of learning about how to improve policies and programs gave rise to "social experimentation". In this case the attempt is to design policy and the technique is again empirical. In a burst of enthusiam in the 1960's, several new social policies (e.g., Medicare and Medicaid) were initiated. Post hoc evaluation indicated that these programs had serious, unanticipated effects. This lead to a desire to carry out small scale social experiments before inaugurating new policies, (e.g., experiments with income maintainance, direct housing allowances, and forms of national health insurance).

In the case of Health Care Systems all three of these styles of analysis could be employed.

(a) Environmental Impact Assessment - This style might be appropriate if we were dealing with the construction of a large hospital. It is interesting to note that such analyses tend to focus on the disbenefits. For example, in the case of a large hospital, we might be expected to measure the disbenefits to air, water, land quality, natural ecology and human habitability.

Less attention would be paid to the health care benefits of the hospital.

(b) Program Evaluation - Examples of this kind would be evaluating the effectiveness of cardiac intensive care units, or community mental health care centers.

(c) Social Experimentation - I have already mentioned the national health insurance experiment. Here various forms and types of insurance are being tested. The principal variables being considered are the co-insurance rate - the percentage paid directly by the patient, and the deductible amount - the amount not covered by insurance.

Let me draw to a conclusion with some general points about outcome measurement, which apply in all of the analytical crafts. In general, the questions I shall raise are difficult to answer.

(a) Virtues of Ambiguity - I mentioned earlier that the first issue we have to face is how to be specific about the goals of a decision maker. Unfortunately, we frequently have a conflict here with the decision maker, because most wise decision makers have learned early in their careers that there are great virtues in ambiguity. If the decision maker is trying to put together a coalition to support a particular action, the reason for that action may be seen quite differently by different groups and the decision maker wants each of these groups to see that action as serving their own goals. It may be necessary, in order to gain the support of all these groups, for the decision maker to be ambiguous about his goals and analysis may not be appreciated in an environment of that sort.

(b) Misincentive Effects of Proxy Measures - The second problem concerns the misincentive effects of proxy measures. I have already alluded to this during my presentation. If you decide you can measure health status by reductions in mortality, then you may force the system into a behavior that will optimize the response with respect to this particular measure, but not

to the overall goal of improving health status. This is a traditional problem, common to all economic systems. It is a central analytical issue of which we should all be aware. I do not have solutions for it, it is a part of the craft we must have in mind when we do any specific study.

(c) Undervaluation of Immeasurable Goals - The main concern that non-analysts have about analysts is the undervaluation of goals that are not apparently measurable, e.g., the well-being of society, the humaneness of life, privacy, etc. We appear inhumane to others because by focusing on those things we can quantify, we tend to forget the immeasurable goals. Of course, we all say that it is the job of the decision maker to bear those in mind. We, as analytical craftsman, must remind him when our analysis does not take into account aspects of his goal that are not measurable.

CONCLUSION

I apologize for repeating some things that I expect many of you already know. However, I hope it has been useful, in that I have tried to structure it in a way somewhat different from what you may have seen before. Let me conclude by talking about outcome measurement and the analytical crafts. It seems to me that outcome measurement is a central issue in all the analytical crafts. Further, the selection of outcome measures is often the most difficult and most influential craft choice in an analysis. We spend a lot of time deciding on the type of models we are going to use: Is a queueing model or a simulation model necessary? If we devoted as much attention to the decision about what outcome measures we should use and justify them to ourselves and to our clients, I think we would be making a major advance in the state-of-the-art of these crafts. In order to do this better, I believe that experience in outcome measurement should be gathered, criticised, distilled and shared among analytical craftsman. I am very pleased therefore that the EURO working groups here today have chosen to discuss outcome measurement. It is an important issue for all of us.

COMMENTARY

1. The discussion focussed on the role of the analyst in the resolution of conflicting interests and on the existence of "misincentives" ie the danger of some proxy or intermediate measures leading to the "wrong" goals being followed. Some examples of this last phenomenon were mentioned and it was suggested that a well-chosen set of proxies was necessary to reduce the danger of misincentive. Levien mentioned the related work of economists on the design of sets of incentives for people working in organisations. Experimentation with the use of alternative sets was said to be very worthwhile.

2. As to the role of the analyst in conflicts of interest Levien suggested two different approaches. On the one hand he argued that where there is conflict amongst decision makers then there should also be a conflict of analysis. He considers analysis to be a useful tool for argument. Since so many assumptions are made along the line, the analyst cannot stay outside of the conflict. He referred to citizens' groups in the US complaining that they do not have the same analytical support as the authorities and mentioned experiments taking place related to this point. On the other hand he gave IIASA's work on a Swedish water control problem as an example of a situation in which the analyst stays outside of the conflict. Several communities in a certain area have an interest in water use in a river basin in that area. Goals are conflicting as to where and how the water is used (irrigation, electricity, pollution, recreation, drinking water, etc). The analyst can guide the conflicting parties to a compromise in an interactive procedure. He even suggested that disagreement amongst decision makers might act as a stimulus to discussion and as a means to better understanding by the analyst. Clearly there's a paradox here which needs further discussion.

3. Also briefly discussed were (i) the use of social experimentation in other countries than the US and the behaviour of people in such experiments and (ii) conflicting interests in designing an emergency care system in Athens.

4. A point of particular interest worth further discussion is the suggestion that goals that turn out to be infeasible are a particular case where goals adjustment is necessary after the analysis has discovered what is possible. But bearing in mind his other warnings on misincentives and the danger of understating or ignoring "immeasurables", there is always likely to be some degree of infeasibility and perhaps re-definition after analysis should be the general practice.

SECTION 4

PERFORMANCE ANALYSIS - WHY AND HOW?*

Göran Arvidsson
The Swedish National Audit Bureau

The paper starts with a discussion of what performance analysis is and what problems analysts are faced with. Commonly used approaches are criticized especially for their concentration on effectiveness and productivity measurement. A multi-dimensional approach is suggested. Various purposes and uses of performance description and analysis are mentioned. Four vital questions are dealt with in some length: Who should make the analyses? What types of performance descriptions and analyses would be most useful in different situations? What demands should be put on performance descriptions? Under what conditions is it most probable that performance information comes to use? A final comment is that personal engagement of managers and other employees in performance description and analysis creates better conditions for responsiveness, flexibility, and initiative in public programs.

"If the schools are as good as they are expensive, they must be very good!"--This reflection by a person in a political cartoon illustrates a major dilemma in public administration. How should one go about evaluating the performance of public activities or--for instance--the performance of a public school:

1. By measuring resources used, i.e., costs, class hours, etc?
   -- To use costs or other measures of inputs as measures of performance does not seem to be very logical--although it is rather common to do so. When public programs are criticized for not being effective enough, politicians often reply by pointing out how much resources they have allocated to these programs. Underlying this type of reasoning must be an assumed--but unknown--positive relationship between inputs and performance or at least between inputs and outputs.
   (Research question: What do these production functions look like in different types of public activities?)

* The views expressed in this paper are not necessarily those of the Swedish Audit Bureau.

2. By measuring outputs such as student hours produced, students examined, etc?
   -- Measuring outputs would seem to be a better method than measuring inputs. But what do output measures really say about performance and about utility in a societal context, etc? What do examination statistics, lecture counts and so on say about the utility of the school? Well, they say something. At least, they give an idea of the level of activity.
   (Research question: What is the correlation between outputs and impacts in different types of activities?)

3. By examining quality assessments of the products such as grade statistics?
   -- Grade statistics might give some indication of the quality of the "products", but both validity and reliability can be severely questioned. We all know that schools awarding high grades are not necessarily good schools.
   (Research question; This means that we have some other source of information telling us the quality of a certain school. What are these sources? Where does "reputation" for good or bad performance stem from?)

4. By asking the stake-holders, i.e., the pupils? the teachers? the employers and other "users"?
   -- This seems to be a better way to acquire information of impacts and performance. But how should the questions be formulated? How should the answers be evaluated? How should one handle the different opinions that are liable to emerge? By ignoring them as being political and return to "objective statistics"? By reporting them as they are identified? By trying to structure and analyze the subjective evaluations according to some stated frame of reference?

My own experience is that output and performance measurement may be of just as little value as measuring female beauty by three measures of circumference. Such measures give one type of information, but not necessarily information on the really important features.

This does not imply that output and performance measurement is meaningless or superfluous. On the contrary. In times of stagnating economies and need for public sector reallocation of resources, these measures become increasingly

important. If decisions are to be made in a way recognizing the utility of different programs and various levels of activity, analysis is indispensable. There is no wonder that the need for follow-up and evaluation seems to be stronger than ever in the Governments of at least the OECD-countries.[1]

WHAT IS PERFORMANCE ANALYSIS?

Let me start with some brief definitions.

- o "Outputs" are the direct results of an activity, i.e., goods and services leaving the agency.
- o "Impacts" are the effects of the activities on individuals and organizations.
- o "Performance" means how well an activity or an agency fulfills its objectives. If the objectives are multidimensional, then performance is a multidimensional concept. (I will expand on this a little later.)

In public decision making, benefits are often considered only intuitively. Output information is needed, but such information may be both qualitative and quantitative. This means that it is more adequate to talk about output description than about output measurement. Also, different parties tend to have heterogeneous opinions on the objectives of public activities and their benefits. This means that the measures need be consistent with the frame of reference by which the activity is viewed. It also means that condensing output information into one or a few aggregate measures may be undesirable. A multidimensional description may be more relevant.

Performance analysis means to study past or ongoing activities in order to determine future activities. This, in turn, means that performance analysis necessarily contains assessments of outputs and their impacts as well as costs and noneconomic sacrifices. It also contains attempts to determine whether certain aspects are good or bad and to explain why.

Outputs can be viewed as "producing" impacts. These impacts can be realized step by step in a long chain of reactions. Therefore, a description of outputs is not a sufficient basis for performance analysis.

Having a chain of outputs and impacts means that the focus of output and impact description may be either closer to the activity or to the objectives. Cf Figure 1. Where to focus depends on the intended use of the description. If a choice in our school example concerns the mixture of lectures and

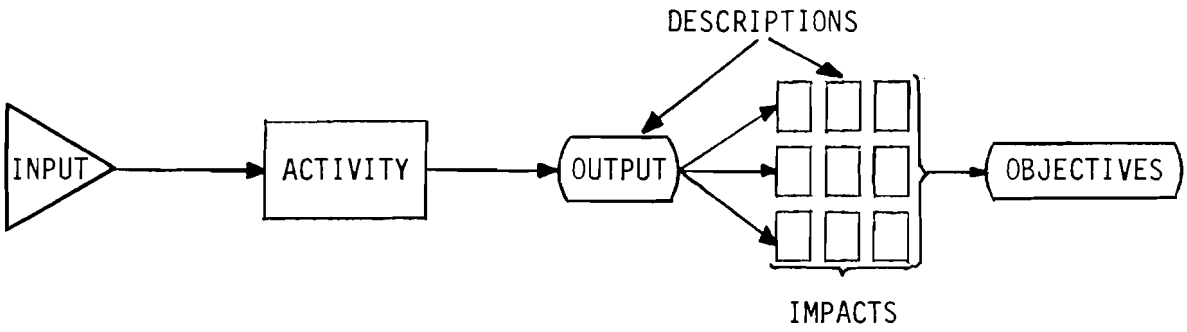


Figure 1 A traditional production model.

laboratory work within a specific line of vocational training, a description of classroom hours, exams passed, etc., may be most adequate. If the choice concerns the dimensioning of the enrollment of students one would have to search for more "ultimate" impact descriptions. Also, routine decisions require routine information. There is little time for specific studies. This means that for routine decisions one would have to be content with production and output statistics in addition to cost accounting information and intuitive impact judgements.

One common criticism of the usefulness of impact assessments is that the impacts of one activity cannot be isolated from the effects of other changes in society. Even though this is a serious problem, it should not be a reason for not undertaking otherwise desirable impact studies.

Now, let me turn to the multidimensional aspects of performance description and analysis. In order to avoid technicalities, I will restrict myself to some aspects of political and administrative relevance.

In Sweden, as in several other European countries, program budgeting ideas were introduced in the mid-sixties. One important feature of this new paradigm of public budgeting was the attention paid to economic rationality, to effectiveness and efficiency. This and other PB-ideas strongly influenced the concept of "good" agency planning and budgeting in Sweden during the first half of the seventies. However, on the Central Government level these ideas were felt to be too far removed from what was considered to be political rationality: "Public activities cannot--and should not be run by using control systems suitable for candy factories. Public activities have intrinsic values which cannot be expressed in terms of economic effectiveness or rationality."

Another reaction to the concept of effectiveness came from the unions of civil service employees and from central agencies for personnel administration and training. They felt that the stress placed on effectiveness and efficiency

was in conflict with their basic values of employee participation, personal development, working conditions, etc. The main proponents of increased effectiveness in Government administration was the National Audit Bureau in its Management (Effectiveness) Audits and the Agency for Administrative Development in its reorganization projects. These agencies soon realized that the concept of effectiveness had to be widened and take into account the aspects that the politicians and the civil servants felt were missing. The widened, presently used concept of effectiveness--or performance--states that a government agency performs effectively if it reaches its objectives

-- whilst husbanding with its resources,

-- with due regard to demands for public service, public disclosure and due process, and

-- with regard to the employees' need for job satisfaction, good work environment, job security and possibility to codetermination and personal development.

This definition has been officially established in an agreement between the National Agency for Government Employers and the central unions of civil service employees.

In comparison with the PB-concept of effectiveness this definition makes overall impact assessments more complicated. Husbanding with resources is not superior to the other goals. The latter may not be seen simply as restrictions. The attempts to make the goal functions more clear-cut by optimizing only effectivenenss in achieving production goals have been refuted. However, what may be lost in clarity for professional evaluators is probably out-balanced by gains in the possibilities of communication between politicians, administrators, analysts, the public, and other stake-holders.

This development does not mean going back to pre-PB-notions of public performance with obscure policies of desired qualities of public activities. Instead, we now have a new frame of reference for impact studies and performance evaluations. Nor is this development peculiar to Sweden. Similar approaches may be found in e.g., American literature. Fried (1976) has a very similar frame of reference.[2] He makes a strong effort to show that the traditional concept of performance is too narrow. He distinguishes three equally important aspects of "bureaucratic performance", namely

-- effectiveness, corresponding to the "work ethic"

-- responsiveness, corresponding to the "democratic ethic"
-- liberalism (or due process), corresponding to the "legal ethic".

Fried also discusses what further dimensions each of these incorporate. The personnel aspect belongs, for example, to the democratic ethic.

Elaborating on the production model in Figure 1 is one way of illustrating the widened concept of performance. See Figure 2. The production goals concern the production of goods and services to the customers, clients, beneficiaries, etc. These goals correspond to Fried's notion of "work ethic".

The organizational goals represent the interest of keeping and developing the agency's financial, material and immaterial resources, i.e., the basis for future action. Certain activities may be directed primarily to fulfilment of these goals. Others may affect them indirectly.

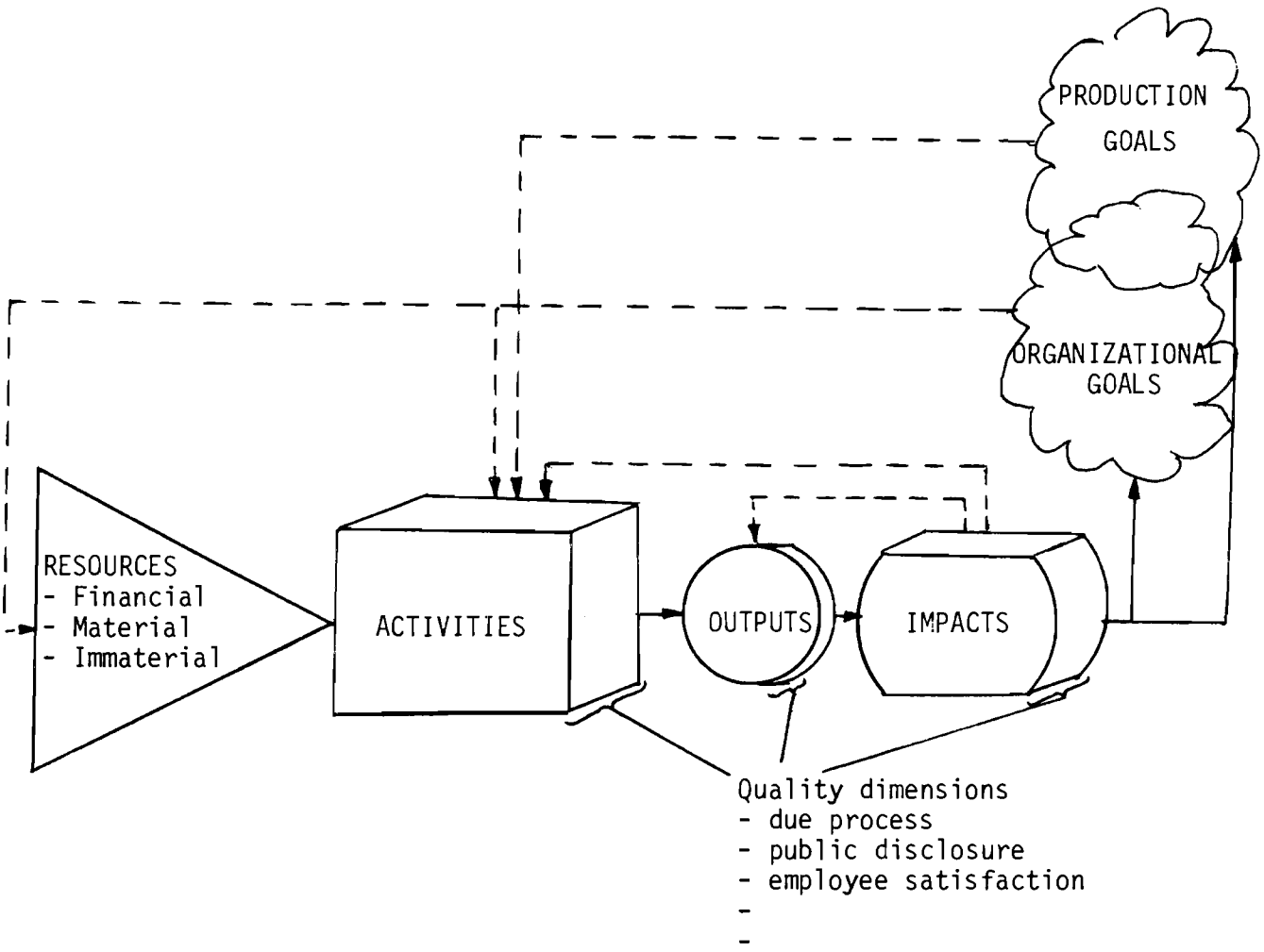


Figure 2 An expanded production model.

Instead of viewing an activity as a matter of producing outputs which "cause" impacts we can conceptually talk about producing "outputs" which are consumed by participation of different groups resulting in new outputs which in turn are consumed ad infinitum. Ultimately, the "quality of life" of different groups in society are involved as well as equity issues.

The point here is that we explicitly need decide where in the production/consumption chain the process should be evaluated and secondly that different consumption groups should be recognized. One major group, for example, is the employees of the organization. This group often value the organization's outputs differently from other stake-holders. To give an example: The inmates of a prison evaluate the organization differently from the keepers.

The quality dimension thus represents the demands from the public and from employees and other parties involved in or otherwise affected by the activities. Some of these demands may differ significantly from those directed to business activities. To use Fried's terminology they represent the "democratic" and "legal ethics". The degree to which quality demands are met can also strongly influence the conditions of future action.

In order to clarify what type of qualitative aspects should be regarded in performance analyses--and how they should be described or measured--a pilot project has been undertaken in the Swedish Central Government Administration. Its first phase was limited to four aspects of agency administration, namely planning, rationalization, personnel administration, and personnel training. Indicators were developed and tested with some success.[3] The study confirmed, however, that indicators are not meaningful if not developed for use in a specified context. It was, for example, quite evident that indicators developed for internal planning use, in but a few cases were considered usable on aggregate levels, e.g., as indicators in reports to superior authorities. Development of useful methods for description and measurement thus presupposes that a judgement has been made of what aspects different stake-holders should be informed about.

The conflict between relevance and measurability was a permanent dilemma. It was difficult to find measures that were convincingly related to policy goals. Another problem was the agencies' varying definitions of basic concepts. Without uniform concepts--and uniform formats for registration of basic data--there is no strong foundation on which to attempt to develop general indicators for use in different settings, e.g., for comparisons between different agencies.

What performance description and analysis is really about may be summarized as follows.

- o Performance description means to identify and express benefits and sacrifices of an activity in quantitative and qualitative terms. Based on the interests of legitimate stake-holders the relevant aspects--or dimensions--of performance should be identified and for each aspect one should try to find valid, reliable and unbiased measures or other modes of description.
- o Performance analysis is to relate performance description data to each other, to standards of performance, to the objectives of the activities, etc.

Performance analysis may be made on different levels of ambitions. In the Swedish Government Budget Manual three levels are distinguished. The lowest is the descriptive level, to describe and comment on actual performance in relevant dimensions. The next level is the follow-up level; i.e., to compare actual performance data (cost, output and impact data) to budgets or other standards. The most advanced level is evaluation, where cost, output and impact data are related to political and other objectives of the activity in question. This crude classification has been found to facilitate communication about performance analysis.

WHY PERFORMANCE ANALYSIS?

Performance analysis may be--and is in practice--used for various purposes. Some examples of what it is used for in Swedish Government agencies are:

- o to make judgements on outputs and impacts
- o to check if a plan has been fulfilled
- o to guide decisions on priorities and resource allocation
- o to find ways of cost reductions
- o to increase the quality of products and services
- o to give the employees a notion of what their work means in a wider context.

All these purposes contribute to one overall purpose, namely to provide better bases for future activities by drawing conclusions from past activities. Performance analysis can give answers to questions like

-- What has been accomplished and what did it cost? (Description)

-- Are the direct results of the activity in accord with plans and expectations? (Follow-up)

-- What effects have been achieved, who benefited and at what costs? How do impacts relate to needs? How do they relate to policy and other goals? (Evaluation).

Or, more widely formulated,

-- Are these activities adequate? Effective? Effective to whom? By what standards of measure? Are they up to expectations?

The underlying hypothesis of the usefulness of performance analysis is that knowing the answers to such questions contributes to better decisions. But, how well founded is this hypothesis? Several arguments can be raised against it. I will mention four.

Firstly, "performance lies in the eyes of the beholder" (Fried, 1976, p.V). Secondly, decisions differ widely in terms of scope, time-span, frequency, political and economic importance, influence on personnel involved, etc. Thirdly, performance information can be structured and presented in various ways, in quantitative or qualitative form, in more or less detail, etc. Fourthly, the analyst having found that performance is not up to expectations and suggestions for improvement being presented does not necessarily lead to change.

This raises four important questions:

1. Who should carry out performance descriptions and analyses?
2. What types of performance descriptions and analyses would be most useful in different situations?
3. What general demands should be put on performance descriptions?
4. Under what conditions is it most probable that performance information comes to use?

WHO SHOULD CARRY OUT PERFORMANCE DESCRIPTIONS AND ANALYSES?

There is a tendency to see performance description and analyses as a business for experts. And indeed it should be. The question is, what type of experts?

For the sake of simplicity I will distinguish between three types of experts.

The first kind are experts on "methods", i.e., statisticians, sociologists, economists, operations researchers, etc. In certain types of studies such expertise is indispensable. Also, when designing systems for production statistics, reporting, cost accounting and so on there is an obvious need for this type of expertise. Scarcity of such experts is often used as an argument for not undertaken performance studies and discussions, even in cases where they are not at all necessary for achieving good quality in the analyses. There is also the opposite problem: the experts play a too dominating role. Experts of this type often fail in the communication with the potential users of the analyses. The users may not comprehend what the analysts are saying or they may reject their findings and conclusions because these do not correspond with their own experience and conviction. In such situations the question is seldom one of right or wrong. Different aspects may be stressed. Expectations of the level of performance may differ. Validity may have different meanings to analysts and users.

The second type of experts are those engaged in the activity. Even if external evaluation may be necessary in order to induce major changes in public programs and even if methodological expertise may be needed in some stages of internal performance studies, my experience is that the main burden of performance description and analysis must be carried by the people in charge of the activity. They have the best insight into it and they will live with the activity even when the study is finished. Conservatism, unwillingness to open up one's own business to criticism and other barriers to reconsideration and change could be decreased by consciously involving the managers and other personnel in not only discussions on the results of performance analysis but also in the stages of formulating the aspects of performance to study, fact finding, analysis and formulating conclusions. Continuous change in response to new demands presupposes that the managers on all levels have the main responsibility for adjustments in their operations. This responsibility is hard to live up to if one does not actively engage oneself in performance description and analysis. I will return to the question of what type of performance aspects it would seem natural for managers to study.

The third category of expertise are those for whose benefit the activity is conducted. In some respects they are represented by the politicians. Political considerations are, of course, the best example. But there are many aspects of performance of interest to the beneficiaries which are not treated best in a political context. Detailed questions of the design of services to

different categories of "customers" are of this kind. This means that the "experts" of the demands and desires directed to the public activities also should be used in performance analysis. Their roles may vary from being asked by interviews and questionnaires to actively taking part in evaluation projects and similar activities.

I earlier noted that output and impact description could be made close to the activity or close to the objectives. To use this scale, it seems natural that performance analysis made close to the activity would be a main task for the managers while analysis close to the goals would be an important task for political and public scrutiny. Governmental, parliamentary and independent evaluation units can make those independent analyses which in many cases are necessary for identifying needs for major policy changes.

However, in both types of analyses methodological experts are needed. The design of data collection routines and of specific evaluations demands professional competence as to specific types of cost/benefit calculations and other types of analyses. A necessary but not sufficient condition regarding performance analyses is therefore that they be technically of high quality. If one believes in striving toward more rational decision making, the value of scientifically based analyses is obvious.

Presently, the main problem does not seem to be want of methodological experts but rather the opposite. The methodological experts tend to intrude into the other parties' domains. This means that they may "relieve" the managers from part of their responsibility, namely self-criticism and the task of suggesting better ways to fulfill political goals. It is easy to find examples of this. In some cases managers even hire evaluation experts as an alibi for not engaging themselves in follow-up and evaluation activities.

The methodological experts may also draw the politicians' and the public's attention to aspects of relatively subordinate interest. Experts on quantitative methods may concentrate upon the effectiveness aspect at the expense of democratic, legal and other qualitative aspects. Reporting low effectiveness may induce lower performance, since other wise useful programs, which do not yield tangible or immediate results may be crippled or abandoned. Also, there is a risk that focussing on outputs and identifiable results may render broader goals displaced by various indicators supposed to measure good performance.

Awareness of these dangers is necessary if methodological experts should be able to play a constructive and useful role together with managers and political and other principals.

WHAT TYPES OF PERFORMANCE DESCRIPTIONS AND ANALYSES WOULD BE MOST USEFUL?

Past experience--in Sweden and elsewhere--shows that general approaches to effectiveness measurement, program evaluation, etc., are liable to fail. The difficulty of defining "performance--and especially "good performance"--is one obvious reason.

In order to be concise, I will give some examples of questions which seem meaningful to ask with respect to different types of operations.[4] Most of these questions concern aspects which are close to the operations rather than to the ultimate objectives of the activity. This means a manager's perspective rather than an external evaluator's perspective.

Handling Incoming Cases, Applications, etc.

Performance analysis may concern the administration of the cases as well as the impacts of the decisions made on each group of cases. Relevant questions to ask include:

- How many cases have been brought up during the period in question? How many have been settled?
- Has the character of the cases changed? How? What tendencies for the future can be noted?
- What is the average handling time for a specific type of case? Has it changed?
- What are the resource demands of different types of cases? Have they changed?
- What are the quality demands to consider? Have they been properly considered?
- What are the clients' reactions? Do they understand the decisions? Do they comply with them?
- Do the resources consumed for different groups of cases correspond with their relative importance?

Different types of statistics can be useful in this type of performance analysis as well as cost accounting information. Several examples of this can be found in, e.g., income tax control, courts of law, the issuing of drivers' licences and passports and so on.

Issuing Norms, Regulations, etc.

Performance analysis may concern the adequacy of the norms, the benefits and costs of them, the observance of them, the unintended effects of them and service aspects like ways of informing of the norms, preparedness to revise them when the prerequisites for them have changed and so on.

In these type of activities, "democratic" and "legal" ethics are obviously of far more importance than the work efficiency of the issuing agencies. This means that evaluations made by institutions external to the regulating bodies could be a valuable complement to the internally initiated studies. Also, external institutions would not be handicapped by their inferior knowledge of the internal processes of the issuing agencies.

Investigations, Research and Development

Typical features of this group of activities are that their outputs are unique and that they are undertaken in the form of programs or projects. Examples of questions in performance analysis:

- How do the results conform with the intentions? What amount of resources was used? How do results and costs correspond to the plans? Reasons for variances?
- What are the major quality aspects of the results? Are the results satisfactory in these respects?
- How did the project administration function?
- How have the results been used? Have they been presented to the potential users in adequate ways? Have they been properly understood and exploited?

Routinized measurement is seldom meaningful in these type of activities. This calls for careful planning of the performance analysis of each program or project.

Inspection and Control

Inspections are widely used in areas like fire prevention, environment protection, sanitary control, workers' protection, road safety, etc. Inspections are often performed according to certain routines and encompass certain groups of "clients". The following questions seem to be relevant to ask in a performance analysis:

- Are the inspections undertaken with due regard to the specific regulations as well as general values concerning fairness?
- How many "clients" have been inspected? What portion of the total population of "clients" has been covered?
- Do the observations and prescriptions lead to corrections?
- What are the chances that the most important deficiencies are identified?
- To what extent do the inspections lead to preventive actions?
- Is there sufficient and relevant information at hand for overall judgements of the status of the field covered by the inspections?
- Is there a preparedness for inspections with short or no notice in critical situations?
- Have inspection results been used in the issuing of laws and other regulations?

Output statistics and cost accounting data can be very useful as a basis for decisions concerning the dimensioning of the inspection staff, standard time per inspection, travelling routes, etc.

Information and Counselling

Performance analysis would cover aspects like target groups reached, relevance of information given, channels for communication, usefulness of the information to the receivers, etc.

Output statistics may prove to be useful in combination with qualitative considerations.

These examples of different categories of externally oriented activities could be refined and completed with different types of internal administrative functions to give a more complete picture, but that would lie outside the scope of this paper. The important point is that the relevant approaches of performance analysis vary from one type of activity to another.

WHAT GENERAL DEMANDS SHOULD BE PUT ON PERFORMANCE DESCRIPTIONS?

Even though performance descriptions should be made with a great deal of consideration to the current circumstances, it seems both possible and desirable to attempt to formulate guidelines. Six attributes of good descriptions are mentioned in a report[5] from the Swedish National Audit Bureau:

Availability

Routine activity statistics of outputs, including or excluding qualitative characteristics, and of various aspects of the production processes have the advantage of being available without elaborate investigations. Such data may be very useful for certain types of decisions, i.e., monitoring ongoing operations. However, available data do not always provide a basis of adequate information.

Comprehensiveness

The information should be easy to comprehend, which means a suitable compromise between detail and overview. Quantitative data are usually easier to grasp, interpret and communicate than qualitative information.

Unambiguity

Ideally, outputs should be clearly defined and homogeneous. This is not always possible. Impacts seldom meet this requirement. Nevertheless, it often seems possible to reach consensus on how to interpret output and performance data or indicators--or at least what changes in such data imply.

Completeness

This ideal can possibly be reached as far as outputs are concerned. Complete descriptions of impacts and qualitative characteristics of public activities can hardly be made. However, this attribute expresses an ambition rather than an absolute demand.

Relevance

As pointed out earlier, this is a difficult demand to live up to. Care should be taken to focus on what is essential to each situation. Also, it is important not to confuse real objectives with indicators or other standards of performance.

Acceptance

Performance descriptions can be expected to come to full use only if the decision makers and other parties involved understand and accept them. It is a common observation that managers of public organizations resent attempts by "outsiders" to interfere, to define and apply performance criteria to "their"

programs. Performance descriptions that do not comply with their own self-image and their desired public image tend to be rejected. Participation in making the description would promote acceptance. If participation is not feasible, care should be taken to explain what the descriptions tell--and what they do not tell.

UNDER WHAT CONDITIONS IS IT MOST PROBABLE THAT PERFORMANCE INFORMATION COMES TO USE?

There seem to be two main prerequisites for performance information to come to real use. One is the factor just mentioned, i.e., acceptance by the decision makers. The other is a willingness to change inadequate operations.

In the design and application of models for "rational decision making" it is usually presupposed that the problem is given and what it is all about is to make the right choice. What is needed is a goal function, relevant data on resources, restrictions, etc., analyses of alternatives and their consequences and finally an "optimal" solution. This is all very well, but often the crucial aspect is that a decision is made at all and, of course, that it is implemented.

If the willingness to make changes fails, it may seem as if performance analysis is meaningless. But this is a static way of looking at it. Seriously made performance descriptions and analyses may very well have their main value in that they influence the propensity to change in public administration. With facts to face, it is difficult for even the most conservative public officials to resist motivated changes. If responsiveness, flexibility, and initiatives are wanted from them, I am convinced that one of the best strategies is to demand that they engage in performance description and analysis. This we can do as concerned citizens. As experts, we can offer them our assistance.

REFERENCES

1) OECD's Technical Co-operation Committee is presently planning a joint activity on measurement of effectiveness, resource allocation and control in public administration.

2) Fried, Robert C, Performance in American Bureaucracy, Little, Brown and Company, Boston 1976.

3) Riksrevisionsverket (The Swedish National Audit Bureau) et al, Kvalitetsaspekter på myndigheternas internadministrativa verksamhet (Quality Aspects of the Internal Administration in Government Agencies), Stockholm 1979 (mimeographed).

4) The classification of activities and the questions are adapted from Riksrevisionsverket (The Swedish National Audit Bureau), Hur kan man få bättre resultatanalys? (How to Achieve Better Performance Analysis?), Stockholm 1979.

5) Riksrevisionsverket (The Swedish National Audit Bureau), Prestations- och Effektbeskrivning (Output and Performance Description), Stockholm 1977, pp 33-35.

COMMENTARY

1. A particularly important consideration within the context of the paper was the attempt to define more exactly the terms used. Indeed, it became apparent during the course of the discussions that not everyone agreed with the definitions suggested. There was, for example, a strongly expressed view that 'outcome' measurement was perhaps a more appropriate label than 'impact' measurement, particularly in the health field.

2. The paper presented a fairly detailed description of the Swedish experience with performance analysis and, more specifically, how this experience had influenced the form and style of analysis. Perhaps not surprisingly, most of the discussion in both the afternoon and the evening sessions was focussed upon this experimental aspect concentrating on the why, the who and the type of performance and the expectations that people had of the approach. The impression conveyed was that the guidelines suggested were the outcome of a considerable learning exercise. This impression was indeed reinforced in the subsequent discussion when a little more light was thrown on the reasons why the activities of the Audit Bureau have assumed a more "persuasive" and less "mechanistic" role than had been originally conceived.

3. A number of the participants had overlooked the fact that none of the Audit Bureau's energies are currently directed at health orientated activities, as a consequence of these services falling under the control of local authorities. This did not prevent the Health Group exploring the relevance of the concept within the Health Sector and there was unanimous agreement that it could prove to be a profitable route to follow. Indeed the impression gained was that elements of such a 'monitoring' system were already

emerging in a number of countries. In Socialist regimes a mass of data is collected which facilitates in depth routine medical auditing on the one hand, and the determination of input/output relationships within particular specialities, on the other. In West Germany it transpires that informal assessments of performance are emerging as, for example, the ranking of hospitals according to the behaviour patterns of physicians or those wealthy enough for their decisions to be influenced only by considerations of perceived quality. The British view was that monitoring should be more widely introduced but it should assume a 'bottom-up' approach which reflects, perhaps, the lessons learnt from the Swedish experience.

4. Other general views expressed were about the value of ordinal measures of output or performance when fully numerical values are not available, the need for clear definitions and the influence of both the style and the state of development of the organisation concerned. One delegate emphasized the importance of experience with low-level measures even for those aiming to develop higher-level systems, since the practical position must be kept in mind at all times. General agreement was felt that the political and social system in which performance/output measurement schemes are operated will strongly influence their design.

SECTION 5

APPLICATION OF OPERATIONAL RESEARCH METHODS TO THE MODELLING OF HEALTH CARE IN HUNGARY

I. Konya and G. Jeszensky

1. The health care of the population of Hungary has now entered a period of qualitative development. That is apparent - amongst other things - from the current changes in the organisation of public health eg. integration, the growing importance of labour organisation and the widening of research in connection with the efficiency of health care. At the same time the system of public health is getting more and more complex in accordance with the growing demands of the population and possibilities of their fulfilment. Thus, demand is constantly growing for the introduction of systems analysis and operational research at all levels of public health. Using these methods should lead to better decisions and to the avoidance of the waste of resources, high costs or the unnecessary use of labour.

2. Recently a growing number of studies have aimed to demonstrate the results of different possible alternatives of possible decisions rather than to find the "optimal" solution.

3. Examining public health at either national or institutional level, the application of mathematical and cybernetic methods and operational research cannot be easily separated. At the national (macro) level they are concerned with forecasting, long-term planning, the evaluation of the development of public health and the improvement of health care information systems using computers. Thus these studies are characterised by an inter-disciplinary and systems approach. In what follows , we present some examples of applications at the macro level in Hungary.

Distributing Medical Personnel

4. Important parts of health care planning in Hungary are the annual and the 5 year plans for the distribution of medical personnel, based on the long-term aims of the government. The plan is developed by comparing the demand for physicians and the expected supply,in all the components of the medical manpower.

5. The basic task is to ensure that the necessary number of physicians are located in the right professional and territorial distribution to fulfil the needs of the population in accordance with the existing and future capacities of health care.

6. With the help of a mathematical programming model we can search for the solution which minimises any territorial and/or professional discrepancy. The optimal solution to this model gives the necessary number of medical personnel for the different counties and within each county, for the various medical branches.

7. Let x_{ik} denote the additional number (development) of physicians required in the kth county of the ith institution-type and let b_{ik} denote the current number (base). According to professional and territorial planning we have the following conditions:

$$F_i^1 \leq \sum_k (b_{ik} + x_{ik}) \leq F_i^2$$

$$M_k^1 \leq \sum_i (b_{ik} + x_{ik}) \leq M_k^2$$

where F_i^1 and F_i^2 are the lower and upper limits of the development of the ith professional field and M_k and M_k are the lower and upper limits of the development of the kth county.

$X_{ik} \geq 0$, as we do not plan to decrease the base during the given period.

The possible solutions must satisfy the following norm-condition:

$$n_i^1 \leq \frac{b_{ik} + x_{ik}}{K_{ik}} \leq n_i^2$$

where n_i^1 and n_i^2 are the lower and upper levels of the norm; K_{ik} is the capacity of the ith institution-type in the kth county. The total increase in the number of physicians can be estimated relatively accurately. Hence, the following condition must be satisfied:

$$V^1 \leq \sum_{i,k} x_{ik} \leq V^2$$

8. A possible form of the target-function, in order to determine the disproportion distribution of physicians, is the following:

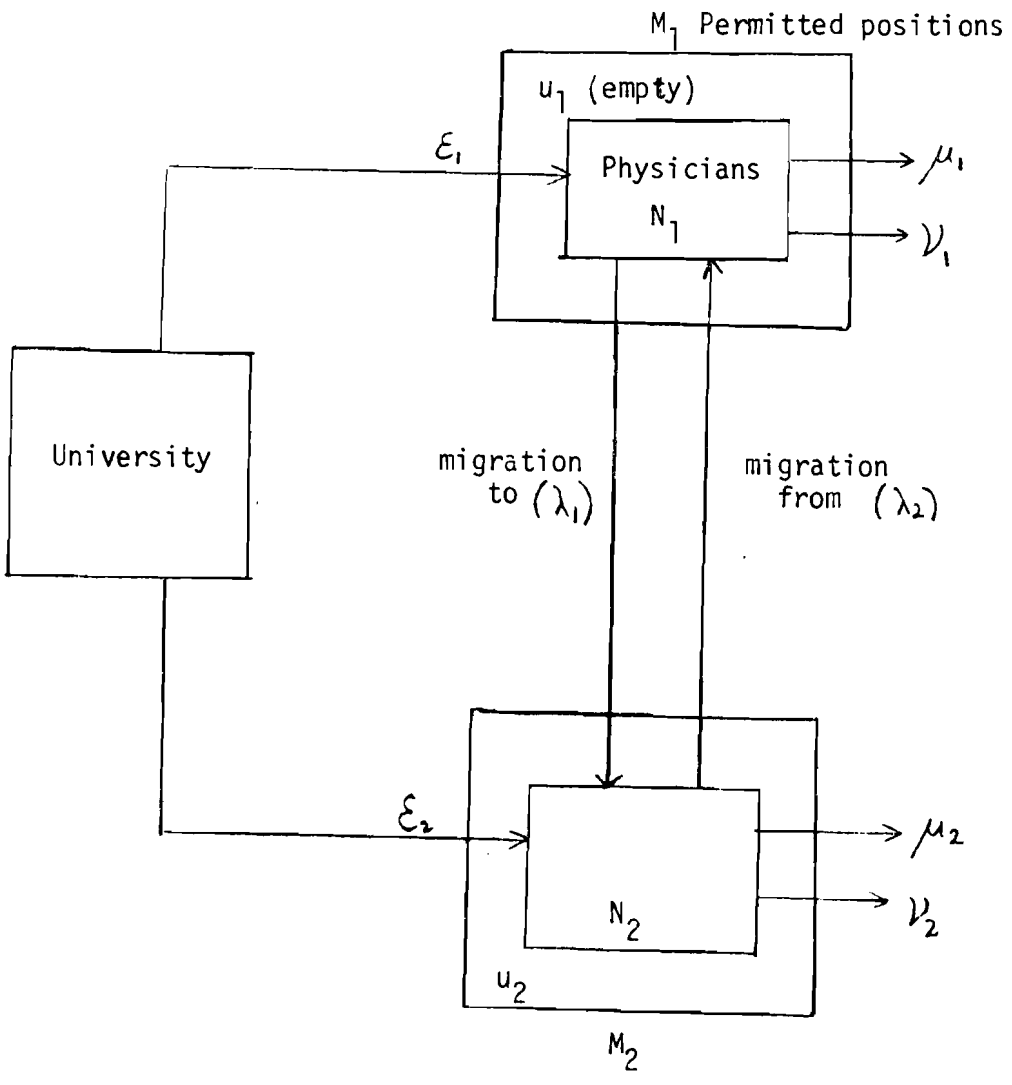
$$\text{Minimise} \sum_i Si \sum_k \left[\frac{b_{ik} + x_{ik}}{L_k} - \frac{Fi}{L} \right]^2 \cdot L_k$$

where $L = \sum_k L_k$ and L_k is the population of the kth county

and where $F_i = \sum_k (b_{ik} + x_{ik})$ and S_i is a weighting factor.

9. Theoretically, it is possible to carry out a full scale flow simulation of the movement of physicians, given the necessary information. Starting from the present state we can simulate the changes in the number of positions and physicians and examine the effect of planned measures. In order to simulate the changes in roughly 25,000 medical positions and 23,000 physicians in the system, we have to describe public health in relation to medical positions, with their appropriate characteristics. Unfortunately however, we do not have the data at present to enable us to use such a model to its full extent. We do have, however, the necessary data for a simplified simulation at a more aggregated level. The following diagram illustrates the model

SIMULATION



The task is to determine the transition probabilities $P_t(N_1,M_1,N_2,M_2)$ given the following conditions:

$$\mu_1 = \alpha_1 N_1 \Delta t \qquad \mu_2 = \alpha_1' N_2 \Delta t$$

$$\nu_1 = \alpha_2 N_1 \Delta t \qquad \nu_2 = \alpha_2' N_2 \Delta t$$

$$\lambda_1 = \alpha_3 N_1 u_2 \Delta t \qquad \lambda_2 = \alpha_3' N_2 u_2 \Delta t$$

We assume that we are dealing with a Poisson process.

An Input-Output Model of Health Care

10. In dealing with hierarchical systems, a principal task of administration and planning (whether dealing with the present or the future) is to elaborate appropriately the capacity of the health care subsystems and the patient flow conditions among the subsystems in order to satisfy the population's demand for health care. The aim is to co-ordinate the care activities of the different subsystems. Without knowing the conditions of patient flow it is impossible to determine the optimal capacity of each of the health subsystems and to coordinate the care activities of the different subsystems. An input-output model of health care allows the examination of the capacity conditions of health care subsystems and for the consideration of patient flow among the subsystems. An input-output table may be used:

- to analyse past data
- to help with decisions concerning development and organisation.

At the different levels of health administration the necessary level of detail in the input-output model is determined by the tasks of the given level of administration.

1. Such input-output tables give the distribution of the treated cases detailed according to the level of administration and broken down by the sending and treating health subsystems. In addition it is desirable both for analysis and planning to make up separate input-output tables according to

- diseases (or main groups of diseases)
- the age of patients treated
- wage-earners and dependents

12. Using these input-output tables it is possible to determine in which age groups and in which diseases flow conditions are unfavourable. Having determined the possibilities of changing the conditions of flow, then with the help of the model it is possible to analyse the effect of the measures to be introduced. The input output table is illustrated in the diagram below.

INPUT-OUTPUT TABLE

	No. of cases treated	Sending health subsystem			No. of patients entering the care system
Treating Subsystem	X_1	e_{11}	e_{12}	....	Y_1
	X_2	e_{21}	e_{22}		Y_2
	. . .				
No. of cases leaving		Z_1	Z_2		
No. of cases treated		X_1	X_2		

In this input-output table, the following notations for the elements are used:

X_i = number of cases treated in the ith subsystem

e_{ik} = number of cases treated where the sending subsystem was the kth

Y_i = number of cases entering the care system at the ith subsystem

Z_i = number of cases leaving the care system at the ith subsystem

These variables may also be considered broken down by disease and age.

13. Using the above notation the following balance equations are valid for the elements of the input-output table (in disease-age combination):

$$X_i = \sum_k e_{ki} + Y_i$$

$$X_k = \sum_i e_{ki} + Z_k$$

These interrelations can also be used to determine the number of cases entering or leaving the care system.

Using such constructed input-output tables the distribution of the number of cases treated may be determined for each care subsystem broken down by the sending subsystem. The values obtained are called the relative numbers of flow. According to our notation they can be computed in the following manner:

$$a_{ik} = \frac{e_{ki}}{x_k}$$

Knowing the relative numbers of flow, the balance interconnections may be stated in the following form:

$$x_i = \sum_k x_k \, a_{ik} + Y_i \quad (i = 1,2 \ldots\ldots n)$$

Using matrix-vector notation:

$$X = AX + Y$$

or $$(I - A)X = Y$$

$$X = (I-A)^{-1}Y$$

where I is the unit-matrix.

Both the coefficients of flow and the above interconnections may be subdivided by any combination of sex, age, disease and wage earner/dependent.

14. According to the level of detail used in the input-output tables, the above interconnections allow the following analyses:

a. where the coefficients of flow remain unchanged
   - to what degree does a given capacity increase (expressed in terms of the number of cases) change the number of cases entering the care system? The analysis is carried out according to the following relation

$$Y = (I-A)\, X$$

- similarly, what is the change in the number of cases leaving the care system in the case of a given capacity increase? The analysis is carried out according to the following relation:

$$Z = (I-A)\ X$$

- to what extent does a change in the number of cases entering the care system alter the number of cases treated in the different care subsystems? Thus the planning of capacity increase becomes possible. The relation for the analysis is:

$$X = (I-A)^{-1}\ Y$$

b. where the coefficients of flow are changed, due to some direction

- what effect does the change of flow conditions have on the number of entering cases? The analysis is carried out according to the following relation:

$$Y = (I-A^{1})\ X$$

where A^{1} is the new matrix of flow coefficients.

- how does the number of cases leaving the different subsystems change by the effect of the change of the flow conditions? The relation for the analysis is:

$$Z = (I-A^{1})\ X$$

- how does the number of cases treated in the different subsystems change by the effect of the change of the flow conditions? The analysis is carried out according to the relation:

$$X = (I-A^{1})^{-1}Y$$

15. The input-output model of health systems outlined above is of a static type, in the sense that it relates to a certain period and is useful for determining the result of existing processes in the system. In order to examine these processes over time, the model must be dynamic. In the case of certain examinations the stochastical nature of health processes cannot be neglected. A stochastic and dynamic model would give the most reliable picture of the operation of health systems.

Commentary (Konya and Jeszenszky)

1. The ensuing discussion was largely concerned with the generalisability of the models described. It was eventually agreed that the models could reasonably be applied to health care provision in most countries, although structural limitations were recognised.

2. The point was made that where health systems were centrally planned and organised, then the setting of common standards or norms, linked to studies of morbidity, was a more likely possibility. Such situations more readily lead to the development and application of the type of models described in the paper.

3. The type of models described are aimed at gaining an insight into the technical, structural and operational parameters of a health system, rather than the behaviour parameters.

4. Normative planning and performance analysis should be a continuous adaptive process in which sensitivity analyses should play a big part.

SECTION 6

PERFORMANCE MEASUREMENT IN THE PUBLIC SECTOR - A DISCUSSION OF PROBLEMS, POSSIBILITIES AND LIMITATIONS

by Mårten Lagergren

1. Introduction: The public sector

A common feature of the post-war development of the Western world is the conspicuous growth of the public sector. Many kinds of explanations have been offered for this phenomenon. I will not go deeper into this in itself fascinating subject. It may, in this context, suffice to say that the growth of the public sector - which almost by definition is a service sector - commonly is described within the context of the transition to the so-called post-industrial society.

The growth of the public sector poses a host of problems concerned with financing and effectiveness. This is so because the public sector has to be financed by more or less willingly sacrified contributions from the citizens and because of the lack of simple criteria for how well these contributions are spent. Since O.R by definition is concerned with the efficient use of scarce resources and, in addition to that, has its roots in the domain of public affairs, more precisely defence, it is obvious that the growing problems of the public sector pose a tremendous challenge to the O.R profession.

The salient question then will be: In what ways can O.R contribute ? Before an answer is sought to that question, however, some more basic questions have to be adressed:

- What is actually the public sector ? What is its role and purpose ?
- Are there any common O.R. problems in this vast area of different organized activities ?
- Do these problems benefit from being treated within a generalized concept ?

The purpose of this paper is to make some comments upon these questions. I will do so drawing mainly from my personal experience working in two different areas of the public sector - defence, and health and welfare. Since the answer I propose to the second question concerns the absence of a monetary measure of output I will dwell mostly upon the question of performance measurements.

Beginning with the first question, however, I find that there seems to be no generally valid answer. The concept varies between different countries. For example, in Sweden and the U.K. most of health care belongs to the public sector in the sense of being publicly administrated. In Switzerland or Western Germany this is not the case. On the other hand in all developed countries health care is an area of public interest, and public money is in some way or another involved. This brings us to another problem: What is it that is public in the public sector ? Is it publicly financed ? owned ? or managed ? It is interesting to note that examples can be found of all the eight possible combinations in capitalist as well as socialist countries (1).

It is obvious then that it is not possible to arrive at a simple answer to our problem. All definitions are possible from the maximum one:
- an activity belongs to the public sector if it is either publicly financed, owned or managed;

to the minimum:
- an activity belongs to the public sector if it is financed and owned and managed by the public.

In the interest of the largest possible platform for our discussion I will subsequently adapt the "maximum" definition. Even with this broad definition, however, it is essential to distinguish between public decisions, i.e., decisions taken by the public for the benefit of society as a whole, and the actual public sector itself. It is the latter area that is the subject of our present discussions.

It is not uncommon, however, that we find that in order to achieve the objectives of that sector in the most effective way, we need to enlarge our analysis to the general area of public decisions. Ample examples of this are to be found both in the area of national security and that of health promotion and general disease prevention.

It is common to distinguish between three roles of the public sector:

- control
- transfer
- production.

The first role is the common role of governmental bureaucracy. It is

mainly concerned with public decisions and has a great impact on societal conditions but is not very important in terms of finance. It is also very dfficult to assess. Even if taxpayers often complain about inefficient bureaucracy it is obviously not a very profitable target for tax reductions. In Sweden it will account for only a few percents of the total public consumption.

The second role is much more important from a financing point of view. In Sweden about 35 % of the public consumption consists of transfers to the sick, the poor and the elderly. Even if the efficient distribution of that money is an important problem I feel it lies outside of the current issue of performance measurement. It must be noted, however, that transfers might be an interesting alternative to publicly organized activities - handing out money instead of services - and thus must enter a broad analysis of the effectiveness of the public sector.

I will however, in this paper concentrate on the public production sphere.

Public production can be of many different kinds: health care, education, air defense etc. The output of this production might be called public services. These are characterized by the following:

- the services are produced by publicly owned, financed or managed institutions
- the services are not distributed according to the rules of the market
- the services are measured in the national accounts by inputs instead of outputs

Public production accounts in Sweden for around 20 % of the total economy and employs around 25 % of the working population. The public production falls into the following main areas:

	% of economy	% of employment
Defense	2.5	1.2
Jurisdiction	1.1	1.1
Education	4.4	6.1
Health Care	5.6	8.3
Social Care	3.1	6.4
Streets & Roads	0.6	0.4
Others (church, culture, employment agencies...)	1.5	1.5
	20	25

This is obviously a very mixed lot of activities. What then has defense in common with health care ?, with education?, with the operations of an employment office ? I believe that the simplest way of describing this is to point out the common ABSENCE OF A MONETARY MEASURE OF OUTPUT. This means that in order to evaluate their operations one has to invent a substitute for that monetary measure. This substitute must in some way be connected with the purpose of the operations - what is one trying to achieve. So we find that to address the problem of measurements we must first deal with the problems of goals and objectives.

2. The general concepts of goals, objectives, and performance measures

Goals or objectives can loosely be described as what we want to achieve. This might be

- to put a man on the moon,
- to increase equality between men and women,
- to reduce poverty,
- to make all people enjoy a good health,
- to get a six-hour working day,
- to bring inflation rate under 5 % and unemployment rate under 3 %,
- to maximize occupancy at the surgical department
- to reduce personnel turnover at the X hospital below 10 % in 1980 etc.

Looking a little more closely on these different goals or objectives we find that they differ very much in character. Some are clearly operational in the sense that it is possible to tell if the goal is attained or not,

e.g., put a man on the moon, or reduce turnover. Others are much more diffuse, e.g., good health or equality. Some can at least in theory be satisfied, e.g., a six-hour working day, others merely point out a direction, e.g., reduced poverty, improved working environment, etc.

Some goals or objectives have a more permanent character, others are directly connected to a special situation. Objectives, that express the fundamental reasons for producing the services in question, might be termed ultimate objectives. These are generally high-level and characterized by not being dependent on the state of the system.

A fundamental problem in performance measurement arises when the ultimate goals are diffuse or even controversial. For example, how do we measure output of psychiatric care if we do not know whether this care is aimed at reducing the stress upon the healthy population of having lunatics running around loose, or restoring so-called "sick" people to "normal" psychic behaviour, or preventing dangerous ideas to float around and threaten society ?

Even if the ultimate objectives are not controversial they might be very diffuse, e.g.,
- what is a "state of complete physical, psychical and social well-being" ? (the WHO health definition).
- what is "national security" ?

The list can easily be augmented to involve most parts of the public sector. In such cases it is obviously not possible to introduce a precise measure that corresponds exactly to the stated ultimate objective. On the other hand there is for example no other really sound way of evaluating the result of health care than in terms of individual health and well-being. In the same way resources spent on defense must ultimately be evaluated in terms of national security.

The "solution" to this is in reality unsolvable problem is to introduce some kind of operational "proxy" objective. The crudest form of this is to measure inputs or some intermediary output instead of final outputs: to maximize bed-days and visits to doctors instead of health or airplanes and destroyers instead of national security. The rationale for this is that it is assumed that the inputs or intermediary outputs (beddays, destroyers etc...) lead to the desired - but non-measurable - final output. The corresponding input-related measures, dominate by far the public decision-making process. Input is mistaken for output and the only answer

to lack of fulfilment of the objectives is to increase resources spent. It is also important to note that even if inputs or intermediary outputs do approximate results fairly well, e.g., beddays in an efficient hospital lead to restored patients, this relationship is by no means stable enough to be used for effectiveness control purposes. If beddays at a hospital decrease as result of a lowered occupancy rate, the easiest way to improve this is to keep the patients longer at the hospital with no or marginal effects on the health result! Thus the assumed correspondance between measure and objective is in this case very easily manipulated. This example can be multiplied by thousands in the actual experience of attempts to control the output of the public sector!

Efforts to improve on this situation must involve some evaluation of how the inputs affect the final output. This means that the output must be described in terms, that are relevant with regard to the ultimate objective. For example, the result of the bed-day, or rather the entire episode of care, might be described in terms of improvement in different characteristics: pain, functional capacity, etc...

In the same way the result of teaching might be described in terms of capabilities and knowledge in different subjects etc...

Some remarks seem important to make in this context

- efforts to specify diffuse objectives generally result in the emergence of multiple objectives,
- specification of diffuse objectives is generally very complicated and involves a lot of data collection. For example, it takes a lot of variables and a lot of readings of these variables to describe the change in health status of a patient that has gone through some kind of surgery. In one production control study I took part in we needed 10-25 forms filled with different data for each patient in the study.
- no matter how well this is done the result at best is only a crude approximation of the real objective. There is always a profound risk that the measurable "proxy"objectives miss important features of the real ones.

Complication obviously is the most serious drawback. Money has the attractive property of being additive. A corporation's incomes from different sources can be added at will without regard to whether the money is

"green, black or yellow". For a publicly operated hospital things are not so easy. The effects accrue to a lot of different patients and there is no obvious way to add one patient's relieve from pain to another patient's improved ability to walk. Economists,in an effort to extend their "dismal science" to other areas, sometimes introduce future earnings of production as a measure of gain from medical care. This, as pointed out by Pole, is a serious mistake since increase of production is not the ultimate objective of the health services (2).

In general, so-called cost/benefit analyses have the serious weakness of not being able to distinguish between the different parties that benefit from the activity under study. Instead information is destroyed by aggregation in the effort of achieving a single-dimensional measure of outcome.

This effort is futile since multiple objectives in the public sector are the rule rather than the exception, and necessary to identify and distinguish . In education we want to improve knowledge and foster good behaviour, in the national economy we want to curb inflation rate and unemployment etc...

Not seldom are these objectives - or at least seem to be - in conflict. Attaining one objective then means sacrificing the other. This conflict could be more or less inherent, e.g., the conflict between a technically sophisticated medical care and a humane, dignified way of treating patients, or it could be the consequence of resource limitations: More money to health care means less money for defense.

In theory multiple objectives can be resolved by introducing an objective in a higher level: well-functioning children instead of just knowledgable and well-behaving. Unfortunately this procedure moves in the opposite direction of specification and thus will bring us more diffuse objectives.

As pointed out earlier this is not a feasible way to go.Thus we have to live with the multiple objectives and develop our methodology accordingly.

Closely connected to the formulation of objectives is the delimitation of the system under study. If we just consider the surgical unit the objective of that might be to perform as many successful operations as possible, (assumed we can agree on the precise meaning of successful), but if we take the whole hospital into account a more reasonable objective might be to restore as many patients as possible to health - regardless of medical or

surgical treatment. And for the people in the health care area it is clearly preferable not having to go to the hospital at all!

Thus, if objectives are considered on a too low level, i.e., for a too restricted system, we run into the risk of suboptimization - defeating an over-all objective by pursuing a minor one. This might result from two different causes:

1. Because the lower objective is inconsistent or only partly consistent with the higher one.
2. Because the smaller system does not contain all effective decision variables.

An example of the first kind is the coastal defense battery that has a very efficient artillery system but bad protection against attacks from the air. The lower objective is to sink enemy assault ships, the higher to protect against invasion. The enemy will however not pay the price of having its ship sunk if it is easier/cheaper to destroy the battery from the air. Thus increased artillery capability will not improve the effectiveness of the battery.

The second kind of suboptimization is illustrated by the case of a hospital department where fluctuations in nursing dependency are met exclusively by regulating the amount of nursing-hours without regard to the possibility of regulating the choice of patients with elective operations.

The crucial problem in the case of suboptimization is that an analysis in order to be relevant in the decision-making context must take into account the real domain of influence on part of the decision-maker. This will not seldom stand in conflict with an analytically sound delimitation of the system - a problem the analyst has small opportunities to do anything about if he does not want to change his job.

3. Formulating objectives and measuring performance - some comparisions between defense and health care

It has been pointed out earlier in this paper that a common feature in the public sector is the absence of a monetary measure of output. Some general problems regarding the formulation of objectives and performance measures that seem to be common to most parts of the public sector have also been mentioned. But is it possible to gain more insight into the realm

of objective formulation by drawing analogies between different parts of the public sector ? Since I have some experience from analytical work both in defence and the health services I shall try some thinking along this line as an illustration to what analysis in one area might learn us about an other.

First it must be pointed out that national defence and health care differ in one very important aspect: National defence is a collective good, i.e., if we defend one citizen we also defend all the others, whereas health care primarily is an individual good with general prevention and the treatment of epidemical diseases as exceptions.

The stated purpose of national defence is improvement of national security and the preservation of national independence. This might be described as its ultimate objective. Some might argue that the purpose is to defend the capitalists from the people and others that it is to reduce unemployment but we might leave that aside. In the same way the ultimate objective of health care must be the improvement and preservation of health and well-being, which can be interpreted as a form of individual independence and security. Defence in peace thus corresponds to prevention of disease and health care might be re-labelled "health defence" (if it were not for possible pejorative connotations!). By contrast the treatment of the sick corresponds to fighting the war.

In health care we distinguish between general health promotion, aiming at improving the general conditions of society, and specific health promoting measures directed towards specific diseases, agents or risk-groups.

In the same way national security measures can take the form of creating a generally favourable international climate, e.g., by supporting U.N. activities, or they can be more specific, e.g., making bilateral treaties with important countries.

Improving the strength of the body to fight the war against disease, e.g., by nutrition and immunization also prevents disease from breaking out. In the same way a strong defence will defer aggressors from waging war against the nation.

The out-break of a war is the equivalent of the out-break of disease. Once this has taken place it is necessary to restore peace/health with least

cost. Peace resp. health must in this case be defined, not only as the absence of war (so-called negative peace) resp. disease but as a state of complete security and independence resp. well-being.

In this effort national defence as well as health care not seldom fall painfully short!

The accomplishment of a certain unit of national defence resp. health care must be measured in terms of its contribution to the overall objective. The primary objective is prevention. Peace is to be preferred to war, health to disease. Prevention should be pursued to the point where it is less costly in a broad sense to take war/disease than to avoid it.

In this way an analogy can be made that will encompass - more or less far-fetched - almost all aspects of the two different systems. The fruitfulness of this might be arguable. The essential lesson might be that there exists an underlying theoretical framework of analytical methodology that can be used successfully applied in different parts of the public sector, part that superficially have very little in common.

4. Performance measurements and the decision-making environment

The theoretical framework of analysis thus seems to be generally applicable to the different parts of the public sector. This does not mean, however, that the conditions for implementing analysis and performance measures in particular are the same.

Above all, performance measurements must be adapted to the decision-making environment. The Measurements Research Working Party, a study group sponsered by the Institute for Municipal Treasurers and Accountants, London, summarized their result in the following statement: "The definition of output measures will only be useful insofar as the measures are designed with the needs of the decision-maker uppermost. Measures which do not ultimately enable better decision making are of academic interest only." (3).

I will conclude this paper by making some remarks on that statement and also comment somewhat on the differences between health care and defence from this aspect.

The full implications of the cited sentences are perhaps not immediately recognized. Let us recall the situation. We are concerned with some public organization producing or planning to produce a certain set of different services. We lack monetary measure of output of these services. The ultimate objectives are diffuse - and at least partly - controversial.

Several kinds of decisions might face the concerned decision-makers. The problem at hand might for example be if one should allocate resources to a proposed activity, e.g., building a new nursery, how to evaluate a proposal to extend an activity to new groups of beneficiaries, e.g., renal dialysis to older patients or if one should restrict a certain activity in order to save money, e.g., close a hospital or how one should improve the output of an activity at given cost or reduce costs with same output.

Since the public sector is subject to political control this decision-making will in general be made in a political context.

In economic theory all this is very simple. You just make a cost-benefit analysis comparing the outcome with all existing alternatives. In practise some problems might accrue. Let us state a few:

- how to summarize in a limited form all measurable aspects of the output that pertain to the ultimate objective of the services
- how to describe how these measurable aspects are distributed among the present or proposed beneficiaries (and how representative these are of the constituency!)
- how to compare the measurable aspects of the output to outputs of another kind from other public services and to other beneficiaries.

Making this analysis is difficult enough, but in addition to that the result must also be internalized by the decision-maker and - in order to fulfill the cited statement - have a significant impact on the decision-making in a political environment!

My purpose is not to discourage anyone from trying to develop or apply better performance measurements in the public sector. I merely want to point out that this is perhaps not as simple as one might believe and that one ought to be very modest in one's claims concerning the possibilities to improve public decision-making by performance measurements.

On the other hand decisions of the kind I have outlined are made all the time

and in almost all cases without the reliance upon meaningful output measurements. It is also an open secret that the public sector is not renowned for its effectiveness. What is then possible to do ? My personal belief is that analysis can improve decision-making considerably by pointing out alternatives, and by trying to describe outcomes and consequences. Performance measurements in the area of public services, where ultimate objectives at best are diffuse and not seldom controversial, serve a purpose only as long as they improve understanding and insight. They are directly contra-productive if they destroy or distort information or hinder useful, informal information to reach the decision-maker.

There are many ways of reviewing an activity - proposed or existing - and formal, quantitative methods might not always be the best!

It is also obvious that the methods chosen must be adjusted to nature of the decision-problem and the decision-making environment. Long-term high-level planning problems will obviously allow for much more sophisticated analytical support than low-level operational problems. Also there are very big differences between for example defence and health care as regards the capacity to digest analytical material and the willingness to accept analytical reasoning and quantified evaluations. These different attitudes between different parts of the public sector can be ascribed to differences in

- tradition
  Is quantified analysis part and parcel of current management ?
- awareness of need
  Are the problems suitable for a systematic, quantified approach ?
- level of knowledge
  Have decision-makers some training in theoretical thinking ?
- supply of meaningful data
  Do data describe what is really going on and can they be put into a meaningful framework ?
- access to technical equipment
  Even if computers are not necessary to make analysis or process data the presence of a computer certainly will encourage these activities!

The public sector undoubtedly is in very urgent need for improved management and effectiveness during the years to come. O.R certainly has a tribute to make in this effort. It is, however, essential that this contribution is made with a true perception of the difficulties and the inherent limitations of the methods. The Hippocratic oath - "Never to harm the patient" - must prevail

also within the realm of analytical advice. Performance measurement in the public sector is an area where it is very easy to make a false step - to misguide instead of lead the way to improvement. It is only meaningful if it improves real insight and understanding. Applied unwisely it only leads to distortion, hindrance and destroyal of information!

References

(1) F L Pryor, Public expenditures in communist and capitalist nations (London 1968), p. 437

(2) J D Pole, The use of outcome measures in health service planning, Intemat. Journal of Epidemiology, Vol. D. No: 1, pp. 23-30

(3) The Measurements Research Working Party, Output measurement discussion papers, The Institute of Municipal Treasurers and Accountants, Dec. 1972

Commentary (Lagergren)

1. In the discussion several pleas were made for some degree of sub-optimisation as this involves the people working on a project to a greater extent compared with some "optimal" solutions at a high level. It was thought important to take psychological and sociological phenonmena into account and the speaker also stressed the importance of being prepared to work with qualitative measures instead of quantitative ones, and of a flexible attitude. It is better to use subjective measures which are strongly related to the system under study than so-called objective measures with no or little relationship with this sytem.

2. Another point mentioned was the fact that health care decisions (as all others in the public sector) are strongly political, which introduces further complications.

SECTION 7

SYNDICATE REPORT A

HEALTH WORKING GROUP

a. General Points

1. The discussion of the seminar has generally dealt with measurement of output and measurement of input but the group emphasised that a true measurement of performance must also attempt to assess the relationships between them ie behavioural, technological etc.

2. Health activities fall into three main categories, ie cure, care and prevention. Much more of the total expenditure on "health" is on care rather than cure, ie doctors less frequently cure a person from an illness than they improve the quality of his life, stop pain, etc without removing the basic cause. Thus, instead of speaking of some kind of Health Status Index we should consider instead a "Health and Wellbeing Index".

3. It was agreed that need in the health field cannot be measured - demand can be measured but this will always expand to meet whatever supply is provided.

b. Specific Points

4. Assessing the performance of any health programme was seen as a multi-dimensional problem - any single figure quoted as a measure of efficiency could be very misleading. A variety of data is required which must be analysed to correct for local variations. Thus it was felt that the term performance analysis was a better term than performance measurement since it more accurately describes the nature of the process.

It was also pointed out that we have no absolute yardstick of efficiency in the health services available to us; we can only compare facility with facility, area with area, or facility with itself over time.

An example of these points quoted was that of two hospitals A and B; A has a high reputation for treating a certain illness so that over time it tends to receive the more complicated cases of this illness. Thus a simple measure such as length of stay might indicate that hospital B was the more efficient of the two. A proper performance analysis would allow for the fact that hospital A had the more difficult cases but even this would not allow its performance to be measured on any absolute scale, only relative to B or other establishments.

5. There appeared to be little dissent in the group when a member stated the basic objective of a health service as the "maximisation of the utility of the patient". A number of problems raised by this definition were identified:-

a. The customer is part of the process - the result achieved is not neutral of <u>how</u> it it was achieved. If a patient is "cured" medically but feels that in the process his dignity as a human being was not respected then that process must be regarded as only partly successful.

b. All output of a health system has a large qualitative element. Even in situations where it appears that money can be used as a measure of value and that therefore such processes are susceptible to approaches such as cost benefit analysis (ie screening for cancer, etc) it was agreed that there are also non-tangible costs in terms of human anxiety.

c. It was not suggested that quantitative data should be neglected - the contrary. However, this data must include the results of consulting professional opinion - the definition of professional being broadened to include such people as the customers of the service ie the patients.

6. It was suggested that monitoring functions (and hence data collection) should begin at the lowest level where people feel that they have some decision making power. At this level not only will they be more aware of what information is really needed but also how it is going to be used. The process can then move further up the organisation structure until it reaches the top. It was felt that the success of monitoring processes established in this way was more assured than one imposed from above. This process would also mean that the information structure established, which is essential for any monitoring process, would be more realistic.

7. A general conclusion appeared to be that countries with a centrally controlled health system had an advantage in the relative ease with which they could collect data. However, it did not automatically follow that such countries were better at achieving an equitable or logical distribution of resources. Centralisation gives the power to achieve this end but not necessarily the will to do so, which is also essential. Political points of view and judgements will always have a very important part to play in such decisions.

8. OR workers, while helping to eliminate the current inequalities in the provision of service should take advantage of these while they exist, to try and measure their effects.

9. There was general agreement that management audit in the health service was a good idea. Again this should start from the bottom of the organisation and move towards the top rather than the reverse.

SECTION 8

SYNDICATE REPORT B

PUBLIC SECTOR WORKING GROUP

Preamble

a. What is the Public Sector?

1. The group adopted as a general definition, the proposition that any activity which is publicly financed or managed or owned belongs to the Public Sector. This corresponds to the "maximum" definition used by Lagergren in his paper (see Section 4). Within this broad definition we distinguish:

A. Intervention in the market

i. controls

ii. income redistribution

iii. investment grants

iv. public enterprises (selling goods in the market)

B. Public production

i. quasi-market goods (not marketed)

ii. public goods (eg defence)

b. What do the Key Terms Mean?

2. The definitions of Annex A, extracted from Göran Arvidsson's paper, were agreed.

c. <u>What are the Pre-requisites of Output Measurement</u>?

3. One of us proposed the following list:-

i. local circumstances must be taken into account

ii. there must be a well-organised and effective public administration

iii. certain political and social choices must be accepted

iv. the analyst should have micro-experience.

4. The group generally agreed to this: on ii. some of us argued that it should in principle be possible to say something about the impacts of public decisions (perhaps if only in qualitative terms) even when public administration is weak.

Question 1: Why, and to what extent, is output measurement needed in the Public Sector?

5. The group agreed that output measures are needed:-

i. for management control at all levels

ii. for planning and selecting between alternatives

iii. for motivating staff at all levels

iv. for clarifying objectives and communicating within large organisations

v. for resolving conflicts within organisations

vi. for information to external stake-holders

vii. for informing public debate over policy questions.

They were especially valuable whenever resources were scarce, both for planning ex ante and for monitoring ex post.

6. Output measures should be carried out to the extent:-

i. that marginal benefits continue to exceed marginal costs. In particular, that considerations of equity demand

ii. that users are able and ready to cope with the level of detail

iii. that is appropriate to the particular level in the administrative hierarchy.

7. We agreed that these conclusions applied throughout the Public Sector as we had defined it, including interventions in the market sector. Examples here might include the extent to which investment grants actually increase investment, instead of merely subsidising investment that would have occurred anyway, and the costs imposed on industry by government controls.

Question 2: Over what types of Public Sector activities, and to what extent, are output and performance measurable. Are there practicable upper limits?

8. It was generally felt that output and performance measures were relatively easier for Public Sector production organisations like nationalised industries than for the production of 'quasi-market' or public goods. We tend only to measure intermediate outputs, such as number of crime squads or student contact-hours, and ignore impacts (final outputs).

9. However, most of the group were fairly optimistic of discovering relations between levels of activities, intermediate outputs and impacts. At the least this was an area where we should do more research, rather than accept from the start that it was impossible.

10. Some even thought that there were no effective upper limits to the extent and precision with which output and performance could be measured. Others thought that upper limits were set by:-

i. the readiness or ability of decision makers to use them

ii. the relative impact of the Public Sector and uncontrollable variables (eg schooling and family influences or educational attainment)

iii. costs.

11. Finally we distinguished regular monitoring of output/performance from special ad hoc research; the latter could study more subtle impacts (or to a greater precision) than was worthwhile for the former. But we noted that there could be a continuous spectrum between these two extremes.

Question 4: Is it easier to measure output and performance for some types of organisation (and hence in some countries) than for others? If so, are the differences inevitable or negotiable?

12. We agreed that ease of measurement would vary widely. Certain sources of this variability identified were:-

i. the degree of centralisation

ii. traditional relationships between levels in the hierarchy

iii. the number of sub-units in the hierarchy

iv. political and social factors - eg the willingness of workers or trade unions to accept work measurement; or the disguising of unemployment as under-employment, which may be useful in some countries as long as it is not made explicit.

13. Differences are largely inevitable - we must not assume that institutions will change overnight to suit the analyst! On the other hand, output measures may in time encourage central organisations to seek different patterns of relationship and devolved responsibility to sub-units.

Question 3: How much effort goes into measuring output and performance? Is this consistent with the effort that goes into measuring input?

14. Members of the group were reluctant to generalise. Some thought that a lot of effort was already going in their countries into output measurement, eg measures of air pollution. But they felt that more work was needed to give a reasonable balance.

15. The UK members thought that a lot of output information, both quantitative and qualitative, was available at lower levels in the hierarchy, but that very little was passed up to the upper levels. Thus many decisions at the centre were taken purely in terms of inputs. Other members agreed with this distinction. One member explained that he was particularly interested in the regional distribution of outputs (this was attempted in Sweden for road construction, but not in general).

16. Another member distinguished between the level of effort devoted to output and performance measures. In his experience, far more work had been done on output measurement. Possible reasons were:-

i. the difficulty of defining objectives

ii. the lack of a 1:1 relation between outputs and objectives

iii. the difficulty of looking at the impact of programmes from the client's view.

Question 5: How do we compare with what should be achievable? Where and how can improvements be made?

17. Although considerable efforts were already being made, for instance in Holland and Sweden, everyone thought that more needed to be done. The UK members said that they did not even know yet where the most obvious improvements could be made: the first need was for a survey of current practice, from which gaps could be identified. Swedish members reported that such an examination had been in progress for about a year in their country, and rather more than another year was required to complete it.

18. One member suggested that improvements in telecommunications might vastly reduce the cost and difficulty of carrying out detailed surveys of the satisfaction or experience of the population with different aspects of public service, like the health status of the nation. For instance, with a viewdata system the respondent could be led through a branched programme: if this was well designed it would be much more powerful (and cheaper) than current survey techniques. Thus our vision of the achievable should not be limited by current technology.

Question 6: How far has a new technology for measuring outputs been developed: what are, or should be its characteristics, and what is needed to develop it further?

19. Despite the work carried out so far, the group did not feel that a consistent set of methods had yet been developed. Such methods might need:-

i. a clarification of key terms

ii. a search for practical intermediate outputs, in the belief that these would help in the ultimate search for measuring impacts and performance

iii. a way of relating intermediate outputs to impacts

iv. an investigation of the properties of measures - eg timeliness, relevance, aggregability

v. development of structure for reporting and using outputs, like the highly developed structures for inputs.

20. The group also discussed the problem of aggregation. Some felt that it was inherently impossible: that aggregating information meant destroying it. Others argued that it was necessary to make the attempt, if output measures were to be exploited to their full potential in allocating resources. We agreed that one might need to go into greater detail for particular decisions (eg to decide where cuts could be made in programmes) than for day-to-day management.

21. Finally we discussed the possibility of exchanging fruitful techniques between countries. We would like to propose the establishment of an international index, classified by service. which would list references to:-

i. people working in the field

ii. published work

iii. unpublished reports.

We invite one member of the seminar for each country to attempt a survey and to provide material for such an index within, say, 6 months. The index could then be used:-

a. to aid informal contacts between people with similar interests

b. to suggest suitable topics for future meetings of EURO (or similar international gatherings).

[Chairman's Note:

This proposal was not given sufficient backing by the conference to justify immediate action. It was decided to explore it tentatively between the two most enthusiastic countries, Sweden and the UK.]

ANNEX A

to Report by
Public Sector Working Group
Syndicate B

D E F I N I T I O N S

1) "Inputs" are the resources needed for a special activity

2) "Outputs" are the direct results of an activity, ie goods and services leaving the agency

3) "Impacts" are the effects of the activities on individuals and organisations

4) "Performance" means how well an activity or an agency fulfills its objectives

ANNEX B

to Report by
Public Sector Working Group
Syndicate B

Supplementary Questions not Tackled by the Group

a) Can we define a common framework for measuring output and performance?

b) Are all goals measurable?

c) What is the role of the analyst in conflicts?

d) What is the client's view of output?

SECTION 9

SYNDICATE REPORT C

OUTPUT AND PERFORMANCE MEASUREMENT IN THE PUBLIC SECTOR

1.1 The need for output and performance measurement (OP and PM) in the Public Sector is related to six uses:-

(a) internal program management. The staff in charge of programs need guidance on their aims, which will be sharpened by having well-described and measured indicators.

(b) external allocation of resources to programs. There was disagreement about whether OP and PM can, in practice, actually help politicians to make decisions about the allocation of resources between programs as different as roads and hospitals. This disagreement focussed on the willingness of political figures to make explicit value judgements.

(c) external audit of programs. OP and PM are needed to provide benchmarks to assist external auditors to examine the effectiveness of a program.

(d) public examination of programs. A wider aspect of external audit is the need for the public to be able better to assess the impacts of programs, to provide an informal input to political choice.

(e) to help motivate staff. Not only do managers need guidance, but all levels of staff may be better motivated, if OP and PM can make more visible the final outputs of their specialised activities.

(f) to provoke, and structure, self-examination. All stable organisations need to be encouraged to continually re-examine their procedures and aims. OP and PM can provide a healthy influence for this process. However, if this process is to be helpful, it is essential to avoid static, rigid systems of OP and PM, which would tend to reduce flexibility and responsiveness.

1.2 However, OP and PM is not an end in itself, but must be justified by its effects - which are often difficult to measure! In principle, the resources consumed by OP and PM should be monitored and its usefullness judged in relation to its effects.

2.1 The Syndicate took an optimistic view of the measurability of outputs. It was felt that some level of measurement could be found for the outputs of most Public Sector activities. However, these "levels" will vary greatly from one activity to another. Even so, some measures will be sufficiently complex to cause a problem of communication, especially to non-specialists. In forming this view about outputs, it was emphasized that it is not necessarily also true for "impacts".

2.2 Where outputs are genuinely many-sided, a convincing description or measurement must necessarily be many-dimensional. This seems to be unavoidable, and produces both a practical problem of data collection and handling and a problem for easy communication.

2.3 There seem to be fewer problems of measurement, and communication of information, when we concentrate on the performance-measurement aspects close to operations, compared with the more politically-orientated output-measurement approach.

This leads to the suggestion that if OP and PM is to succeed in becoming established in unwelcoming bureaucracies, the approach should be "bottom-up", rather than "top-down". A "top-down" strategy was felt to be more likely to suffer the kind of collapse which overtook PPBS.

3.1 The Group felt that little effort is expended in most parts of the Public Sector on the measurement of output, but that performance measurement was much more prevalent. However, the view was expressed that many managers only assess performance in an informal, impressionistic way, and that when quantified measures are made their existence is not revealed to superior management - they are for the personal use of the manager, who may fear a loss of independence if he discloses his assessments.

3.2 The information usually used to estimate inputs was basically designed for a different purpose - financial control. It is primarily intended to prevent fraud, allocate costs, and control expenditure within budgets. These purposes demand great accuracy and detail, which makes them expensive systems to maintain. The information on inputs is essentially a spin-off from the prime financial control purposes, so it is difficult to compare the direct efforts put into OP measurement and input measurement.

4.1 We chose to examine whether "organisational structure" affects measurability of OP and performance. We concluded that in principle it is the nature of the outputs, rather than the type of organisation structure, which mainly determines measurability.

4.2 However, organisation will have a strong influence on practical aspects of measurement, such as the ability to collect various types of information.

4.3 Another relevant practical aspect will be the differences in attitudes which often exist between public and private sectors, and may make impractical the monitoring of performance in both sectors, for example because of Union pressures.

4.4 The closeness of political influence may also have its effects. One view expressed was that politicians prefer ambiguity of aims - although they may feel that there is political advantage in subjecting bureaucracies to performance measurement.

5.1 Four needs for further development were noted:-

- the need to move forward from defining output to measuring its impacts

- to orient managers' focus on output rather than the traditional stress on inputs

- to draw attention to the need for co-operation between agencies, in identifying how the work of one had an impact on the objectives of another

- the use of output measurement to stimulate the creativity of the individual managers.

6.1 Finally the Syndicate felt that the question of a "new technology" for measuring outputs and performance probably did not arise. The necessary condition was the approach and attitudes of individual managers at all levels to the tasks of management. If the manager is open minded and questioning, and trained in assembling relevant information as a basis for decision, he will find sufficient techniques already available in statistical and survey methods.

SECTION 10

DISCUSSION ON THE SYNDICATE REPORTS AND FINAL ADDRESS

1. The discussion focussed mainly on the next steps. The suggestion made by Syndicate B for the construction of an international index, classified by service, which would list references to the people working in the field, published work and unpublished reports, and that this should be used to aid informal contacts and to suggest suitable topics for future meetings, was not opposed but was supported only by the UK and Sweden. Action on that can therefore be handled bilaterally at the moment.

2. Nevertheless, delegates felt it is desirable to keep each other in touch with what is going on, and there is little doubt that will be accomplished in the normal course of events, within the Health Working Group. This was in tune with some people's views that the subject of performance and output measurement was next best taken forward within specific fields of interest.

3. Other more general approaches thought worthwhile included:-

a. More work on the methodology of performance and output measurement.

b. The need to be able to deal with complex, inter-related and conflicting measures.

c. To ensure that measures are produced for working-level management as well as for planning.

d. To note the increasing tide of criticism from taxpayers and to provide measures that they will regard as relevant.

e. To investigate the wider applicability of methods of social accounting.

4. In his final address Rolfe Tomlinson augmented these points by reiterating some general principles of Systems Analysis and the work of the working groups.

5. First, that all measurements must be related to a purpose, and the purpose is usually associated with the future; there is little point in analysing the past except to learn from our mistakes and for adjusting future plans.

6. Second, we must be realistic about the kind of uses to which measurements can be put: we must not raise unreal expectations in minds of the people who are going to receive the results.

7. Third, we should be concerned to make sure we are giving good advice. The only possible way we can be sure of that is if we receive adequate feedback. We must act responsibly, and be involved with and concerned about the decision making process.

8. Fourth, a complex organisation may need a complex control system: we must match the variety of measures to the variety of those needs. There will always be a need to keep things as simple as possible: nevertheless there is danger in single or too-simple performance measures which can act as misincentives to people whose performance is being assessed.

9. Systems change, objectives change, and needs change over a period of years. Any performance measurement system must be able to adapt to new circumstances and new objectives.

10. In performance and output measurements above all else, the analyst is not dealing with a "one-off" problem, but with a need for a system of measurement which will provide a continuing management and planning tool.

11. Finally, as a future President of the EURO Federation, Mr Tomlinson urged delegates to continue to make sure that their working groups acted as a positive stimulus to the subject. He hoped more people would take a purposeful attitude to the working groups and EURO itself.

SECTION 11

FINAL COMMENTS

1. The form of the meeting, in which syndicates discussed a common set of questions in the light of the papers read, was designed to bring out points of agreement and focus on questions requiring examination. As editors, we do not feel it is useful for us to attempt a further synthesis of the points made by each speaker and syndicate. There is much commonality of views between them and some differences, and we urge readers to examine the papers, particularly the syndicate reports if time is really not available for more, to arrive at their own conclusions on the implications for their own area.

2. Meetings of the Health Working Group commonly number about 30 participants out of a total group membership of about 50 - 60, and meetings of the Public Sector Working Group commonly number 15 out of a working group membership of 135. The success of our first objective, to bring together members of these two separate groups, and the staff at IIASA, was therefore met rather better for the Public Sector Group than for Health. Perhaps this reflects one view of our subject matter, "performance and output measurement", as providing a coherent theme for people interested in a wide range of Public Sector work, which can capture interest more widely than a conference focussing on some particular part of the sector, while on the other hand the Health Group already have a sufficiently common focus to their subject. It is consistent too with the consensus view amongst the Health Group, who felt that the conference had provided a useful widening of horizons but the next step would be to focus more on specific examples within their own interests, and a wider feeling in the Public Sector Group that the subject of performance measurement is the general interest, and that while specific cases would be useful for illustrative purposes we should not focus too narrowly at this stage.

3. The second and main aim of the meeting was to attempt to bring out common threads of thought and methods of working, if possible to develop and discuss new ideas at the meeting, and to identify specific ways in which the subject of performance and output could be taken forward. It seems to us that there was insufficient time to develop that programme: although some common threads emerged and common problems were identified, and some interesting ideas were suggested, discussion did not proceed long enough to clarify the choice or to fire sufficient enthusiasm to overcome the barriers of distance and expense, in arranging common programmes of work or frequent meetings. Thus the outcome was in terms of individual contacts, and bilateral discussion of ideas and identification of common interests, rather than a common programme of interest to all.

4. Apart from that, the joint meeting provided an encouraging view for those in the Public Sector Group, of the progress that can be made when discussion focusses on one particular sector, such as Health, not only in developing ideas but also in coming to terms with realistic measures rather than theoretical ones. Whether such encouragement would be derived from other sectors like education, transport, or urban planning, we do not know. The picture we obtained of the work at IIASA was also of great interest to members of the Public Sector Group, although it was not so new to members of the Health Group who are already in touch with the Health programme there.

5. There will be meetings of both Groups at EURO IV at Cambridge in July 1980, with an opportunity to consider further progress both on the subjects of our meeting and the value of joint meetings with other groups.

APPENDIX 1

PERFORMANCE AND OUTPUT MEASURES

A JOINT MEETING OF EURO PUBLIC SECTOR AND HEALTH WORKING GROUPS

January 14-16, 1980

Schloss Laxenburg, Austria

LIST OF PARTICIPANTS

Agrell P	FAO 1, S-104 50 Stockholm	Sweden
Arvidsson G	Revisions Directeur Swedish National Audit Bureau	Sweden
Boldy D	University of Exeter	UK
Bos I R J	Ministry of Agriculture	The Netherlands
Broekhof C	Economisch Institut, Tilburg	The Netherlands
Brolin B	Televerkets Centralforvaltening	Sweden
Brown P (Miss)	HM Treasury, London	UK
Cutolo I	Instituto Universitario Orientale	Italy
De Vries G	Eindhoven University of Technology	The Netherlands
Di Carlo G S	Regioneria Generale Dello Stato, Roma	Italy
Dietrich C	IABG, Munich	West Germany
Dobson C	HM Treasury, London	UK
Eriksson B	Statskonsult	Sweden
Firnkorn	Robert-Bosch-Stiftung Stuttgart	West Germany
Hafkamp W A	University of Amsterdam	The Netherlands
Himatsingani C	Department of Health, London	UK
Howard K	University of Bradford	UK
Jeszenszky G (Mrs)	Ministry of Health, Budapest	Hungary
Konya K I	Ministry of Health, Budapest	Hungary

Lagergren M	Secretariat for Future Studies Stockholm	Sweden
Matson E	University of Trondheim	Norway
Montesarchio E	Regione Lombardia, Seveso	Italy
Moores B	University of Manchester	UK
Morgan P (Miss)	Home Office, London	UK
Muller-Sloos A (Mrs)	Erasmus University, Rotterdam	The Netherlands
Nousiainen I	National Board of Education Helsinki	Finland
O'Kane C	Queens University, Belfast	UK
Ovcharov V K	Institute of Social Hygiene and Public Health, Moscow	USSR
Papoulias D B	National Technical University Athens	Greece
Ramsin H	Statskonsult, Stockholm	Sweden
Renieri A	SAGO, Florence	Italy
Sissouras A	University of Patras	Greece
Turner T P	Civil Service Department London	UK
Turtiainen K	Planning Secretariat Ministry of Finance, Helsinki	Finland
Whitworth B	Local Government OR Unit Reading	UK

IIASA PARTICIPANTS

Andersson A

Aspden P

Cantley M (now Commission of the European Communities)

Kitsul P

Levien R

Shigan E

Tomlinson R

APPENDIX 2

PERFORMANCE AND OUTPUT MEASURES
A JOINT MEETING OF EURO PUBLIC SECTOR AND HEALTH WORKING GROUPS

January 14-16, 1980

Schloss Laxenburg, Austria

AGENDA

Monday January 14, 1980 (Wodak Room)

12.00	Registration at IIASA	
12.30	Lunch in the Schloss Restaurant	
	IIASA'S WORK - Chairman: R Tomlinson (IIASA)	
14.00	Introduction to the Institute	R Tomlinson Management and Techology Area
14.30	Health Care Systems Modelling at IIASA	E Shigan, P Aspden, P Kitsul Human Settlements and Services
15.00	Discussion	
15.30	Break	
16.00	Monitoring of Health Services A UK Care Study	M Cantley Commission of the European Communities
16.30	Discussion	
17.00	Systems Analysis in Regional Planning	A Andersson Integrated Regional Development
17.30	Discussion	
18.15	Heuriger	

Tuesday January 15, 1980 (Wodak Room)

DISCUSSION PAPERS ON PERFORMANCE AND OUTPUT MEASURES (1)

Chairman: D Boldy (EURO Health Working Group)

9.00	A US Viewpoint	R Levien, Director of IIASA
10.00	Discussion	
10.30	Break	
11.00	A Swedish Viewpoint	M Lagergren Secretariat for Future Studies Stockholm
12.00	Discussion	
12.30	Lunch in the Schloss Restaurant	

DISCUSSION PAPERS ON PERFORMANCE AND OUTPUT MEASURES (2)

Chairman: T P Turner (EURO Public Sector Working Group)

14.00	A Hungarian Viewpoint - Application of Operational Research Methods for Modelling Health Care Delivery	I Konya, Mrs G Jeszenszky Ministry of Health Budapest
15.00	Discussion	
15.30	Break	
16.00	Another Swedish Viewpoint	G Arvidsson Revisions Director of the Swedish National Audit Bureau
17.00	Discussion	
18.00	Buffet	

PARALLEL DISCUSSION SESSION (1)

19.00 Separate meetings of Public Sector and Health Working Groups to discuss the issues raised by presentations, and to produce a short paper setting out recommendations for future work.

21.00 End of discussion

Wednesday January 16, 1980 (Seminar Room)

PARALLEL DISCUSSION SESSION (2)

9.00	Separate meetings of Public Sector and Health Working Groups (continued).	
12.30	Lunch in the Schloss Restaurant	
	REVIEW OF THE CONFERENCE Chairman: R. Tomlinson (IIASA)	
14.00	Report from the Health Sector Working Group (Syndicate A)	
14.20	Report from the Public Sector Working Group (Syndicate B)	
14.40	Report from the Public Sector Working Group (Syndicate C)	
15.00	Break	
15.15	Final Discussion Summary of the Conference	R Tomlinson President-elect of EURO
16.15	End of Conference	

APPENDIX 3

PERFORMANCE AND OUTPUT MEASUREMENT IN THE PUBLIC SECTOR AND HEALTH CARE

Discussion Notes for Meeting at Schloss Laxenburg January 14-16, 1980

1. Why and to what extent is output and performance analysis needed in the Public Sector and Health Care?

2. Over what types of Public Sector and Health Care activities and to what extent, are output and performance measurable? Are there practicable upper limits?

3. How much effort goes into measuring output and performance within a given Public Sector, or Health Care, and is this consistent with, for example, the effort that goes into measuring input?

4. Is it easier to measure output and performance for some types of organisation (and hence in some countries) than for others? If so, are the differences inevitable or negotiable?

5. How do we currently compare with what should be achievable as regards to output and performance measurement? Where and how can improvements be made?

6. How far has a new technology for measuring outputs and performance been developed for the Public Sector and Health Care: what are, or should be, its characteristics and what is needed to develop it further?