statistical reports and studies

rapports et études statistiques

measurement of output of research and experimental development

unesco

PREVIOUS TITLES IN THIS SERIES

No. 1 - Film and cinema statistics.
(with resumé in French and Spanish)
Out of print.

No. 2 - Book production 1937-1954 and translations 1950-1954. (Paris, 1958)
Production de livres 1937-1954 et traductions 1950-1954. (Paris, 1958)

No. 3 - Statistics on libraries. (Paris, 1959)
Statistiques sur les bibliothèques. (Paris, 1959)

No. 4 - Statistics of newspapers and other periodicals. (Paris, 1959)
Statistiques sur les journaux et autres périodiques. (Paris, 1959)

No. 5 - Statistics on special education. (Paris, 1960)
Rapport statistique sur l'enseignement spécial. (Paris, 1960)

No. 6 - Requirements and resources of scientific and technical personnel in ten Asian countries. (Paris, 1960)
Besoins et ressources de dix pays d'Asie en personnel scientifique et technique. (Paris, 1961)

No. 7 - Pre-school education. (Paris, 1963)
L'éducation préscolaire. (Paris, 1963)

No. 8 - Statistics on radio and television 1950-1960. (Paris, 1963)
Statistiques de la radiodiffusion et de la télévision 1950-1960. (Paris, 1963)

No. 9 - Methods of estimating the demand for specialists and of planning specialized training within the USSR. (Paris, 1964)
Méthodes d'évaluation des besoins en spécialistes et de planification de la formation spécialisée en URSS. (Paris, 1964)

No. 10 - Estimating future school enrolment in developing countries, a manual of methodology. (Paris, 1966)
Estimation des effectifs scolaires futurs dans les pays en voie de développement - manuel de méthodologie. (Paris, 1967)

No. 11 - Methods of analysing educational outlay. (Paris, 1966)
Méthodes d'analyse des dépenses d'enseignement. (Paris, 1967)

No. 12 - Methods of long-term projection of requirements for and supply of qualified manpower. (Paris, 1967)
Méthodes de projection à long terme de l'offre et de la demande de main-d'œuvre qualifiée. (Paris, 1968)

No. 13 - Statistics of education in developing countries. An introduction to their collection and presentation. (Paris, 1968)
Les statistiques de l'éducation dans les pays en voie de développement - Comment les rassembler et les présenter. (Paris, 1968)

No. 14 - International developments of education expenditure 1950-1965. (Paris, 1969)
L'évolution internationale des dépenses d'éducation entre 1950-1965. (Paris, 1969)

No. 15 - The measurement of scientific and technological activities. (Paris, 1969)
La mesure des activités scientifiques et techniques. (Paris, 1969)

In the press

No. 17 - World Summary of Statistics on Science and Technology/ Statistiques de la science et de la technologie : Aperçu mondial. (Bilingual publication/ édition bilingue) (Paris, 1969)

measurement of output of research and experimental development

a review paper

by Christopher Freeman

Director, Science Policy Research Unit
University of Sussex

United Kingdom

unesco

Printed in 1969 in the Workshops of the
United Nations Educational, Scientific and Cultural Organization,
Place de Fontenoy, Paris 7^{e}, France

ST/S/16
COM.69/XVI-16 A
Printed in France

PREFACE

This study on the Measurement of output of research and experimental development was originally prepared as a working document for the first meeting of the Unesco/Economic Commission for Europe Working Group on Statistics of Science and Technology which met in Geneva in June 1969. The favourable reception given to this paper, prepared by Mr. C. Freeman, Director of the Science Policy Research Unit of the University of Sussex, United Kingdom, has led Unesco to make it available to a wider public by publishing it in the series Statistical Reports and Studies.

In the same series Unesco previously presented a document on The measurement of scientific and technological activities (Unesco, Statistical Reports and Studies, ST/S/15, Paris 1969) by the same author, designed to add to the literature on the classification and measurement of the inputs of human and financial resources into scientific and technological activities. The present study, a complement to the first, reviews the far more complicated subject of the measurement of the output of research and experimental development (R and D).

Statistical measures of the output of R and D are needed in order to reach a fuller understanding of the innovation process and its impact on the economy and to allow for a more rational allocation of limited resources to competing R and D activities. However, many theoretical and practical problems involved in the compilation of such statistics remain to be solved and it cannot be expected that solutions will be reached in the near future in view of the many difficulties still existing in the far more advanced area of statistics of R and D inputs. Although no immediate solution may be found, a review of the results achieved so far in measuring the output of R and D can nevertheless contribute to a better understanding of the problem and stimulate further discussion.

The ideas expressed in the present paper are those of the author and do not necessarily represent the views of Unesco.

The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of the Unesco Secretariat concerning the legal status of any country or territory, or of its authorities, or concerning the delimitations of the frontiers of any country or territory.

CONTENTS

		Page
Chapter I	INTRODUCTION	7
	The growth of research activities	7
	Professionalization and specialization of scientific research	7
	The measurement of inputs into research and experimental development activities	8
	Input measurement and output measurement	8
Chapter II	SOME PROBLEMS OF OUTPUT MEASUREMENT	10
	Theoretical objections to any scheme of output measurement	10
	R and D output as a flow of information and innovation	10
	"Accidental" factors affecting R and D output	11
	Use of input measures as a surrogate for output	12
Chapter III	THE USE OF SCIENTIFIC PAPERS IN MEASUREMENT OF RESEARCH OUTPUT	14
	Numbers of papers and other output measures	14
	The use of output measures in the sociology of science	15
	The need to relate "output" to "input" measures	16
	The need to relate output measures to institutional factors and social changes	16
Chapter IV	THE USE OF PATENT STATISTICS FOR MEASUREMENT	20
	Patents, inventive activity and "applied" research	20
	Limitations of patent statistics	21
	Applications of patent statistics	22
	"Science" and "technology" in relation to papers and patents	23
Chapter V	INNOVATIONS AND THE R AND D SYSTEM	25
	Invention and innovation	25
	Cost-benefit analysis of innovations	25
	Indirect measures of the output of innovations	27
	Output budgeting	28
Chapter VI	SUMMARY AND CONCLUSIONS	31

APPENDIX

Tables

1.	The flow of ideas through the stages of research, invention and development to application	34
2.	Schematic table showing numbers of authors of various degrees of productivity (in papers per lifetime) and numbers of papers so produced	35

Page

Tables

3. Origin of science abstracts 36

4. Percentage of women on the professional staffs of higher educational institutions in 1947 compared with percentage of articles in professional journals contributed by women in 1940 and 1960 37

5. Distribution of research publications and their origin 38

6. Institutions from which more than 100 papers were published during the year under analysis 39

7. Number of discoveries in the medical sciences by nations, 1800-1926 39

8. Patents granted to foreigners as a percentage of total patents, 1957-1961 40

9. Approved enterprise-to-enterprise agreements for the transfer of technology in Japan 41

10. Outline of a proposed output budget covering Home Office responsibilities 42

Chapter I

INTRODUCTION

THE GROWTH OF RESEARCH ACTIVITIES

One of the outstanding features of the Twentieth century has been the rapid growth of scientific and technological research and a range of associated activities. The growth of these activities and their impact on economic, military and social policy has been variously described as the "Research revolution", the "Scientific revolution" or the "Technological revolution". One result of this "revolution", however it may be defined or described, has been a demand for statistical measurement of the resources devoted to the generation of new science and technology and of the efficiency with which they are being used.

It may well be true, as Derek J. de Solla Price has suggested,(1) that scientific activities were growing very rapidly already in the Eighteenth and Nineteenth centuries. In this perspective, the Twentieth century growth may be regarded simply as the continuation of a long-term trend. But the absolute scale of the resources committed before 1900 was so small, even in the United States, that it amounted to much less than 0.1% of GNP. Today almost all European countries devote resources to research and experimental development which are the equivalent of between 1% and 3% of GNP,(2) while many developing countries are already spending the equivalent of 0.1% or more of their GNP.

In these circumstances it was inevitable that there should be increasing concern with the efficiency of the research-innovation system. In some sense, this implies measurement of inputs and outputs of the process. As long as governments or enterprises were spending only very small sums on scientific research, they could afford to regard this outlay in a very similar way to patronage of the arts, using "prestige" criteria rather than attempting to assess "efficiency". But it is one thing to endow an occasional eminent scientist; it is quite another to maintain laboratories regularly employing thousands of scientists and technicians on a continuous basis. The increased scale of scientific activities led inexorably to an increased concern with their effectiveness.

PROFESSIONALIZATION AND SPECIALIZATION OF SCIENTIFIC RESEARCH

The larger scale of scientific research was associated with its increased professionalization. A high proportion of scientific and inventive work in the Eighteenth and Nineteenth centuries was conducted on a part-time or amateur basis. This proportion is, of course, still significant today, particularly with respect to inventive activity, but more characteristic of the second half of the Twentieth century is the full-time professional research scientist or engineer and the specialized research institute. It is sometimes forgotten that, even in the Nineteenth century, the élite scientific organizations, such as the Royal Society, found it difficult to accept the idea of professional status for scientists. Ben-David has pointed out that: "Academic appointments ... were regarded as honours rather than careers and turning science into an occupation would have seemed something like a sacrilege."(3) The very word "scientist" only came into general use in the middle of the century. Even then the main awards to scientists were seldom on the professional basis of full costs, including instrumentation, overheads and supporting staff.(4) It was only towards the end of the century that a few industrial firms began to set up small professional research laboratories on a permanent footing. Thus, Whitehead was justified in describing the greatest invention of the Nineteenth century as the "method of invention itself",(5) in the sense of a network of full-time research organizations.

Although the professionalization of science proceeded rapidly in the industrialized countries in the early part of the Twentieth century, it was not until the 1930's that the number of patents taken out by corporations in the United States exceeded those taken out by individuals,(6) and it was not until 1953 that the first official government survey was made of the total resources employed in professional research and experimental development in the country.

It would indeed have been difficult to survey and measure scientific research activities before they reached this fairly advanced level of specialization and professionalization. Once it had been reached, however, it became possible to undertake reasonably accurate surveys of the numbers of professional research and experimental development scientists and engineers in industry and in government and of the expenditures necessary to finance their activities. But, even now, one of the biggest difficulties in surveys of research and experimental development inputs is the problem of part-time research workers, and no country has satisfactorily resolved this question with respect to university research.

THE MEASUREMENT OF INPUTS INTO RESEARCH AND EXPERIMENTAL DEVELOPMENT ACTIVITIES

The first official government statistics were those published by the Soviet Union since the early 1930's.(7) These related, however, to a range of scientific and technological activities somewhat wider than those now commonly defined as "research and experimental development".(8) Most other socialist countries now also publish annual statistics of scientific services, although without comparability in their coverage and definitions.

The first experimental attempt in a market economy to measure R and D inputs in all sectors (industry, government and universities) was made by J. D. Bernal in 1938-1939.(9) But Bernal had to make use of very poor data for industry and an unsatisfactory breakdown of government expenditure. Industrial organizations had begun to publish figures of expenditure on industrial R and D in the United States and in the United Kingdom in the 1930's,(10) but these were incomplete in their coverage and inconsistent in their definitions. It was not until the 1950's that the National Science Foundation (NSF) in the United States resolved these problems by systematic comprehensive surveys in industrial and government organizations on consistent definitions.

Since the NSF began their regular annual surveys of R and D expenditures and manpower in 1953, many other OECD countries have followed suit. Unfortunately, they often did so on the basis of varying national definitions and concepts, so that international comparability was difficult to attain. The Directorate for Scientific Affairs of OECD took the initiative in attempting to standardize definitions and systems of measurement. The first Frascati Conference in 1963 agreed on a standard system of measurement,(11) and as a result of this OECD was able to undertake an experimental international comparison of a few countries(12) and later a more systematic comparison - the first international statistical year for research and development.(13)

Within the United Nations, Unesco has taken the lead in encouraging and systematizing the measurement of R and D inputs.(14) It has recently initiated attempts to reconcile definitions and concepts as between East and West European countries.(15) As with other similar statistical series, institutional differences between socialist and capitalist economies make this especially difficult and progress so far has been small. Unesco has also stimulated measurement of scientific and technological services in many of the developing countries, but here too there are major problems of scope and comparability of national statistics.

Nevertheless, it is not unreasonable to suppose that as an increasing number of countries gain experience of regular statistics of R and D manpower and expenditure, and as international organizations become more familiar with the problems of comparison, it will be possible to use a fairly wide range of moderately accurate and comparable statistics of R and D inputs. Persistent efforts will be necessary to improve their accuracy and range.

INPUT MEASUREMENT AND OUTPUT MEASUREMENT

The position with regard to measurement of R and D outputs is completely different. There is no nationally agreed system of output measurement, still less any international system. Nor does it seem likely that there will be any such system for some time to come. At the most, it may be hoped that more systematic statistics might become possible in a decade or two.

This paper, therefore, is not concerned to make proposals for international systems of R and D output measurement. It has the much more limited objective of reviewing briefly some of the experimental attempts at output measurement which have been made, of selecting those which appear to offer the best future prospects, and of indicating ways in which international organizations might stimulate further experimental developments to the point where regular national and international series become feasible.

First it will be necessary, in Chapter II of this paper, to deal with some theoretical objections to the whole idea of output measurement and with the present treatment of R and D inputs in systems of national accounts. The next three sections review various attempts to develop systematic output measures for part of the R and D spectrum of activities, and in the final section suggestions are made for further experimental work.

NOTES

(1) Derek J. de Solla Price, Little Science, Big Science, Columbia University Press, 1963

(2) OECD, International Statistical Year for Research and Development, Vols. 1 and 2, Paris 1967-1968

(3) J. Ben-David, "Scientific Productivity and Academic Organization in Nineteenth Century Medicine", American Sociological Review, December 1960, page 836

(4) R. MacLeod, The Institutionalization of Basic Research: the Government Grant Committee of the Royal Society, 1850-1914, Science Policy Research Unit, University of Sussex, to be published.

(5) A. N. Whitehead, Science and the Modern World, Pelican, 1937, page 120

(6) Jacob Schmookler, Invention and Economic Growth, Harvard University Press, 1966, page 26

(7) J-M Collette, "Recherche-developpement en URSS", Cahiers de l'ISEA, Institute de Science économique appliquée, August 1962

(8) Unesco, The Measurement of Scientific and Technological Activities, 1969

(9) J.D. Bernal, The Social Function of Science, Routledge, 1939

(10) N.E. Terleckj, Research and Development: its Growth and Composition, National Industrial Conference Board, (NICB), 1963
Federation of British Industries, Surveys of Industrial Research

(11) OECD, Proposed Standard Practice for Surveys of Research and Development: The Measurement of Scientific and Technical Activities, DAS/PD/62.47

(12) C. Freeman and A. Young, The Research and Development Effort in Western Europe, North America and the Soviet Union, OECD, Paris, 1965

(13) OECD, op. cit. (2)

(14) Unesco, Provisional Guide to the Collection of Science Statistics, COM/MD/3, Paris, 1968
Unesco, op. cit. (8)

(15) Unesco, op. cit. (8)
idem, Questionnaire on statistics of research and experimental development effort, 1967 (Unesco, STS/Q/681), Paris, July 1968

Chapter II

SOME PROBLEMS OF OUTPUT MEASUREMENT

THEORETICAL OBJECTIONS TO ANY SCHEME OF OUTPUT MEASUREMENT

The need for output measurement is seldom disputed by those actively engaged in the management of research and experimental development, whether in government, industry or universities. But, however desirable such measurement may appear to policy-makers, it is sometimes maintained that output measurements are either unattainable or useless.

It may be that the satisfactory measurement of part or all of R and D output will prove unattainable on purely practical grounds. This can only be established by attempting the measurements with skill, determination and ample resources over a considerable period of time. The measurement of Gross National Product or of R and D inputs at one time also appeared extremely difficult on purely practical grounds. However, there would be no point in making even the attempt to measure R and D output if it could be clearly demonstrated that the objective was in principle unattainable or unnecessary.

Two such arguments are therefore briefly considered here, as summarized by Machlup:(16)

> "One might take the position and defend it on good grounds that it is impossible even to define invention, let alone to identify, count and weight inventions, and if it is meaningless to quantify the output it must be meaningless to assert or posit the existence of a quantitative relationship between input and output . . . "

> "Or one might take a less negative position and grant the possibility of quantifying input and output at least roughly or for the purpose of constructional reasoning but at the same time hold that the incidence of accidents in making inventions is too great to legitimize even provisionally the assumption of a production function."

Although "invention" is discussed here, essentially similar arguments may be applied to other types of R and D activity. Take first the argument that it is impossible to define and measure inventive output (or scientific output). It may be conceded that this is extremely difficult, but it cannot be denied that there is an output of some kind from all types of research and experimental development activity. An attempted representation of these outputs and the corresponding inputs for various stages of research and experimental development work is shown schematically in Table 1. (The differences between this scheme and the usual "input" classification of R and D are discussed in Chapter IV.) From this it can be seen that the output of all stages of R and D activity is a flow of information and the final output of the whole system is "innovations" - new products, processes and systems. This information is conveyed in various forms and through various media, with varying degrees of secrecy or freedom. Some of it is "intermediate" or "feedback" output. But there can be no doubt that such a flow of information exists and that it is valuable.

R AND D OUTPUT AS A FLOW OF INFORMATION AND INNOVATION

The problem is therefore one of defining and measuring the flow of certain types of information and the efficiency with which this flow is utilized by various organizations to make innovations. The main criterion to distinguish this flow from all other information flows is the criterion of novelty. It may be readily admitted that this is not an easy criterion, either in definitions or in practice. But it is not an impossible one. It is a criterion which, although difficult to apply, is constantly used. It has for centuries been the foundation of patent law in many countries and it is a criterion commonly applied for scientific publication. Whilst it is true that individual cases may be disputed, it would be hard to deny that there is an essential difference between repeating information which is already known and imparting something new, and that there are new products and processes. Therefore, the argument that the whole output of research and experimental development is in principle not definable is unacceptable. The problem is reduced to the practical one of separating the R and D information flow from other types of information flow, of trying to measure all or part of this flow, and of assessing efficiency in using new information generated by R and D activities.

If we cannot measure all of it because of a variety of practical difficulties, this does not mean that it may not be useful to measure part of it. The GNP does not measure the whole of the production activity of any country, largely because of the practical difficulties of measuring certain types of work. The measurements of R and D inputs omit important areas of research and inventive activity. But this does not mean that GNP or R and D input measures are useless.

Some parts of the information flow are captured and embodied in well-established, accessible forms. The best-known examples are published scientific papers and patents. It can scarcely be denied that these do represent a part of the output of research and experimental development activity, although it may be (as Machlup maintains)[16] that they do not represent the most important part of the output of fundamental research or of inventive work, or that they are not representative of the whole. Nevertheless, it must be conceded that if we are able to measure that part of the information flow which is embodied in scientific papers and in patents, then we would in principle be able to measure at least a part of the output of R and D activity.

It may still be argued that there is as yet no satisfactory way of reducing scientific papers or patents to comparable standards as yardsticks of measurement. This is a question on which there has been a certain amount of empirical research and on which there are some important findings. The third and fourth sections of this paper are, therefore, largely devoted to the use of scientific papers and patents, as a possible means of measuring part of the information flow of research and experimental development. It is argued that the results of empirical work already justify the use of scientific papers and patents for some output measurement purposes, despite the severe difficulties and limitations involved.

The measurement of that part of the output of research and experimental development work which is embodied neither in published papers nor in patents, presents greater difficulties and even less empirical work has been done. Nevertheless, there are some possibilities of measurement, although largely indirect. These are discussed in Chapter V of this paper.

For many policy purposes the information flow, which is generated during the R and D process, is only a means to an end. The ultimate aim is usually a flow of innovations, which may be considered as the final output of the system, while the information flow is an intermediate output. Chapter V therefore considers the measurement of the "final output" in this sense.

"ACCIDENTAL" FACTORS AFFECTING R AND D OUTPUT

We may now briefly consider the second main line of theoretical criticism which, if valid, might be sufficient reason to discontinue attempts at output measurement. This is the argument that the input/output relationship is too arbitrary and uncertain in research and experimental development activity to justify any attempts to improve efficiency or effectiveness. It rests largely on the view that unpredictable accidents are so characteristic of the process that rationality in management is impossible to attain. This argument need not detain us long.

It is evident immediately that the argument itself assumes some knowledge of both inputs and outputs in at least part of the range of R and D activities. Otherwise the lack of relationship could not be presumed. The view is usually largely based on some well-publicized anecdotes of supposedly accidental factors in scientific work, such as Fleming's penicillin mould.[17]

The straightforward answer to this type of argument was given by Cottrell when he said:[18]

> "If ... you accept an invitation by a pharmaceutical firm to investigate the medical effects of chemicals, you are distinctly more likely, to put it mildly, to turn up a new drug than a new alloy or a new radio-star."

The logical fallacy lies in assuming that, because accidental features are present in individual cases, it is therefore impossible to make useful statistical generalizations about a class of phenomena, whether natural or social. Those concerned with an individual street accident are always impressed by the peculiar features of the occurrence - if X had not postponed his journey by 15 minutes, it would never have happened; if Y had not been worried by his wife's illness, if the street sign had been 10 yards further down etc., etc. All these factors are undoubtedly extremely important in determining the specific form of each accident, which individuals are involved in it, the nature of their injuries and so forth. But they in no way prevent the statistician from forecasting with a high degree of accuracy the number of street accidents which will occur in a given month to a given country, and from classifying many features of the "accidents". Similarly those involved in any individual scientific discovery or invention are always impressed by the number of apparently accidental features, and often they may be right to think that but for these accidents this particular discovery or invention would not have been made. But this need not prevent the social scientist from making useful generalizations about a class of discoveries or inventions.

This is not to deny the presence of accidental factors in research, as in many other human activities. Nor is it to deny the existence of very wide variations in the relationship between input and output. All industrial production functions involve a statistical distribution with fairly wide deviations from the norm.[19] In agriculture, for example, when the farmer uses certain inputs such as seed, fertiliser, land and labour, he knows very well that the actual output per acre, or per hour of labour, may vary enormously from season to season. These variations in output due to "accidental" factors, such as weather, or to factors over which he has limited control, such as pests, do not mean that he cannot take rational decisions about the use of inputs, or that he is rendered incapable of improving average yields.

By analogy, many R and D managers or scientists act "as if" they were farmers. They know that there are unpredictable and accidental factors present in their work. But they also know that, if they apply their labour with ingenuity and appropriate equipment over a sufficiently long period, they will probably achieve some useful results. This attitude has been justified in practice by the whole growth of science and technology over the past hundred years. The existence of commercial contract research institutes and the steady increase of company-financed R and D operations are evidence of the economic viability of a large range of R and D activities, which can be managed with some degree of rationality, despite the unpredictability of particular experiments.

USE OF INPUT MEASURES AS A SURROGATE FOR OUTPUT

The converse of the argument on the rôle of accidental factors in research and experimental development is the view that variations in output are so slight that they can be disregarded or that they average out over a large enough sample. Some such assumption is in fact involved in the use of input measures as a surrogate for output measures. We already know enough about output variations to know that for many purposes input measures are not enough although they are better than nothing. We cannot compare relative efficiency unless we have some direct or indirect measure of output as well as input.

In market economies the use of R and D input measures in national accounts may nevertheless be justified in the absence of any output measures, as in the case of many other service activities. But it must be remembered that the actual treatment of R and D in national income statistics today is complex, depending upon the type of economy, the sector of performance and method of finance.[20] In many capitalist economies, when a piece of research is both performed and financed by government, it will normally be treated as part of GNP - final output - and measured by its input cost. This would also be true if the work were performed extra-murally in industry but paid for by government. But if the work is "company-financed" it will normally be treated by the firm as a cost of production and will not be measured as a final product. If the research is financed by a private individual donor or a non-profit institute, it may be treated in national accounts as private consumption expenditure and included in GNP. Despite this variety of treatment by the national income statisticians, it is quite reasonable for the economist to treat all R and D expenditures as a form of social investment in both capitalist and socialist economies.

The treatment of the public sector in many social accounts systems involves frequent use of cost of input measures in lieu of output measures. As with R and D, efforts at output measurement are still at a very primitive stage in areas such as education, health services and so forth. Some direct indicators have been used, for example, numbers of patients in hospital or children in school, ratios such as patients per doctor or pupils per teachers, or indirect indicators such as mortality rates. But it is generally agreed that none of these measures yet provides any satisfactory general scheme of "output" or quality measurement for these services.

It may be that a general system of output measurement suitable for incorporation in national accounts will never be attained and that for this purpose we shall have to continue to use input measures. In the socialist countries measurements for some service activities, whether of "inputs" or "outputs" are in any case often excluded from the national accounts system. But this need not prevent the development of output and efficiency measures, which are specific to each activity and which can be used to compare the performance of organizations, of individuals and of countries in that activity, and with the financial outlays for each activity.

One of the greatest difficulties in representing research output in a form suitable for a national accounts scheme is that so much of it is:

(a) Feedback output to other parts of the system.

(b) Output which is used only after long and unpredictable time lags.

(c) Output which can be "consumed" an infinite number of times.

This applies above all to basic research, whose function is to generate and maintain a "multi-purpose knowledge base".[21] It is clear that the results of basic research are, by definition, not intended to serve any specific practical aim, but to provide a flow of general scientific information which may be used in a great variety of applications. This "output" cannot be assessed in relation to the policy goal of any particular government department or industrial enterprise, but only in a much wider context. Even "applied research" which does have specific practical objectives may find very wide applications far outside the context of the original research. For this reason it is both more practical and more logical to attempt measurement of research output initially by the flow of published information, rather than indirectly through the ultimate applications. This becomes progressively less true as we move across the spectrum to experimental development.

(16) F. Machlup, paper contributed to The Rate and Direction of Inventive Activity, Princeton University Press, 1962

(17) R. Taton, Reason and Chance in Scientific Discovery, Trans. A. J. Pomerans, New York, 1957

(18) A. H. Cottrell, Science and Economic Growth in the United Kingdom, British Association, 1966

(19) F. Machlup, op. cit. (16)
C. Freeman, paper in Problems of Science Policy, OECD, Paris 1968

(20) F. Machlup, The Production and Distribution of Knowledge, Princeton, 1962, pages 184-7

J. P. Lamouche, Recherche Scientifique et Comptabilité Nationale, EEC, Brussels, 1968

(21) J. Schmookler, op. cit. (6)

Chapter III

THE USE OF SCIENTIFIC PAPERS IN MEASUREMENT OF RESEARCH OUTPUT

NUMBERS OF PAPERS AND OTHER OUTPUT MEASURES

Three main yardsticks have been used for the measurement of output of basic research: scientific publications (usually "papers"), "discoveries" or other major contributions to the advance of knowledge, and colleague evaluation or peer judgements. The last two methods often depend upon some qualitative evaluation of the first. Moreover, a count of papers, whether weighted or otherwise, is the only method which lends itself readily to large-scale statistical application. Whilst all these methods can easily be used on a small scale simultaneously or combined for ranking purposes in a field which is well-known to the investigators, it is difficult to extend such combined indices of performance across a wide range of disciplines and countries or over an extended time period.

The promotions board, or appointments committee for research or university teaching posts will normally try to take into account all three methods of evaluating the output of candidates. This is quite feasible, since they are usually familiar with the individuals concerned, as well as the subject. But despite the very widespread practical application of such rough and ready output assessment, little success has been attained in generalizing this experience across a wider frame of reference. Most historians of science and sociologists have also tended to work mainly on a "micro" scale, using the second and third techniques of output measurement. As in the case of the appointments board, this is quite reasonable procedure at the micro level. To everyone who is familiar with research, it is obvious that it may be dangerous to rely on a simple count of numbers of papers in assessing the output of any particular individual or small group. But it does not necessarily follow that such quantitative techniques cannot be applied to much larger aggregates. A great deal of statistical analysis is based upon the knowledge that in a sufficiently large population many individual variations can be ignored, even though they cannot be ignored at the level of the small group. For example, Pareto's law is generally valid for a country but not necessarily for the individual firm or for a village. It may be legitimate to use quantitative measures as a substitute for a qualitative assessment or a combined quantity-weighted-by-quality index, if it can be shown for any field of investigation that at the selected level of aggregation the quantity of scientific papers does not vary greatly from the combined quantity/quality index. Thus, if the national origin of a list of major medical discoveries in the twentieth century conformed almost exactly to the national origin of the "key papers" as assessed by experts, and the pattern of both in turn conformed closely to the national origin of the gross number of medical research papers appearing in a selected range of journals, it might be legitimate to use the third measure for some purposes as a proxy for the other two.

In practice the difficulty is that no one has yet established the range of applications, or the limits within which such surrogate quantitative measures may confidently be used. Some empirical work has shown a degree of correlation between the three types of measurement in a few areas of application. But the errors and difficulties associated with each type of measurement, as well as the restricted area of validation, do not yet give sufficient grounds for confidence in widespread application. Reasonable caution dictates that, wherever possible, several methods of assessment should continue to be used simultaneously as with the interviewing boards. Experimental work should be continued, as some of the results thrown up by straightforward quantitative analysis are of great interest for science policy.

For example, Rangarao has estimated that the average output of papers by Indian researchers is approximately equivalent to one paper every 10-12 scientist-years.[22] This may be compared with Price's rough estimate for world science of an output of one paper every two scientist-years.[23] Careful examination of both estimates would be necessary to ascertain the degree of comparability in definition of input (numbers of full-time equivalent research scientists) and output (range and procedures of abstracting services) and dates of measurement (which differ slightly). At first sight the difference in output is very great and at variance with Price's own hypothesis of a roughly similar input/output ratio for world basic research in general. But if valid it provides important supporting evidence for the views of Dedijer(24) and others on the research environment in Indian research institutes and universities. Kapitza has also estimated very roughly that the output of Russian research scientists in terms of papers is only half that of their United States colleagues.(25) He emphasizes the severe statistical difficulties in making such estimates and all these comparisons need very considerable care in their interpretation. It would be unwise to jump to policy conclusions without serious critical analysis of the data and methods. It seems likely that the Indian definition is much wider than Price's, including scientists who do not publish at all.

THE USE OF OUTPUT MEASURES IN THE SOCIOLOGY OF SCIENCE

Several important contributions to the sociology of science have already been made by studies based on the use of scientific papers as a method of measurement. The broad scope of these contributions has covered such questions as the following: the age at which scientists are most productive in various disciplines; the relative contribution to science of industrial, government and university scientists; the long-term rate of growth of the output of various scientific disciplines and sub-disciplines; the relative contribution of various countries to world science in particular disciplines; the relative contribution of male and female scientists to research output; the growth of multiple authorship of scientific papers and its implications; the relative contribution of outstanding and lesser scientists to research; the institutional environment most conducive to high research productivity. Whilst most of these results must be regarded as in need of further validation and testing, they already constitute an important body of knowledge. Here it is only possible to indicate very briefly some of the most important findings and some of the hypotheses which these have generated or tested.

Alfred J. Lotka, in a pioneering article in 1926 on "The frequency distribution of scientific productivity",(26) demonstrated for some branches of natural science that for every 100 authors who produce only one paper in a particular period, the number of people producing "n" papers is approximately "$1/n^2$" (Table 2). Derek Price has provided additional evidence supporting Lotka's observations, has reformulated the "law" governing distribution of productivity and has pointed to some of its implications for the long-term growth of the scientific community.(27) Price emphasizes the importance of the findings of Wayne Dennis(28) and others on the output of the most eminent men of science. These show that the most outstanding scientists have usually been prolific in the volume of their output. Price is, of course, well aware that one paper by Einstein cannot be compared with one or even 100 papers by "John Doe";(29) and that there is "no guarantee that the small producer is a nonentity and the big producer a distinguished scientist". He argues, nevertheless, that in spite of obvious exceptions and variations, "on the whole there is, whether we like it or not, a reasonably good correlation between the eminence of a scientist and his productivity of papers. It takes persistence and perseverence to be a good scientist and these are frequently reflected in a sustained production of scholarly writing".(30)

Price has used the output of scientific papers and the number of scientific journals to generalize about the long-term growth rate of the scientific community,(31) in the United States and elsewhere. In broad terms he has deduced an input measure from an output measure (the opposite procedure from some national income statistics), arguing that the number of scientists, the number of papers and the number of journals have all been increasing at an exponential rate of 5% to 7% per annum (i.e. doubling every 10 to 15 years).

These observations provide extremely interesting hypotheses on the long-term trends in science and likely future trends, but it should be noted that they are not entirely consistent with Price's modified version of "Lotka's Law" and his generalizations about the much slower rate of increase in numbers of "good scientists". He argues that the "total number of scientists goes up as the square, more or less, of the number of good ones".(32) But if "good" scientists are much more prolific than lesser ones, then not only would there be diminishing returns, in terms of average quality of output per scientist, as Price postulates, but there would also be a continuous slowing of the rate of increase in quantity of papers, by comparison with number of scientists. The point is by no means academic, since the period under consideration is three centuries and "Lotka's Law" suggests that the "good" scientists account for a high proportion of total output.(33) The tendency might, of course, be offset by other long-run changes, such as varying pressures to publish, the changing pattern of scientific careers, "professionalization" of research, the growth of post-graduate scientific research degrees and so forth. The variations which apparently exist between Indian, Soviet and United States' output of papers per "scientist" provide grounds for considerable caution in making generalizations about "world science" over long periods. Not only are publication practices different but the amount of secondary material in journals varies significantly.