Scientometrics, Vol. 7. Nos 3–6 (1985) 281–308

EVALUATING BIG SCIENCE: CERN'S PAST PERFORMANCE AND FUTURE PROSPECTS

J. IRVINE, B. R. MARTIN*

Science Policy Research Unit, University of Sussex
Brighton BN1 9RF (UK)

(Received May 28, 1984)

After explaining the reasons why science policy-makers face a growing need for more rigorous forms of research evaluation, we outline an approach combining bibliometric and peer-evaluation data that has been developed at the Science Policy Research Unit in the course of a programme of studies of Big Science specialties. The paper describes the results obtained when this 'method of converging partial indicators' is applied to compare the past research performance of the accelerators at CERN – the joint European Laboratory for Particle Physics – with that of the world's other main accelerators. The paper concludes by demonstrating how, on the basis of an analysis of the factors that have structured research performance in the past, it is possible to arrive at a systematic set of conclusions about the future prospects for a major new research facility such as an accelerator.

Introduction

Among the great variety of research activities, it is perhaps the areas of Big Science that best exemplify the thesis of *Derek Price* that the pace and direction of scientific progress depend crucially on the instrumentation available to researchers.[1] It was the easy access to war-surplus radar equipment and the subsequent construction of the first large radio telescopes in the late 1940s and 1950s that opened up the exciting new field of radio astronomy.[2] Similarly, many of the more spectacular discoveries in high-energy physics have followed hard upon the heels of advances in

---

*No order of seniority implied (rotating first authorship). The authors are Fellows of the Science Policy Research Unit (SPRU), University of Sussex, where they work on a range of issues connected with policies for basic and applied research. They gratefully acknowledge the support of the British Economic and Social Research Council in carrying out this research, and that of the Leverhulme Trust in meeting the costs of additional analysis and the writing of this paper. The autors also wish to thank various colleagues at SPRU, especially Professors Linda *Wilson* and Keith *Pavitt*, for numerous useful comments and criticisms. An earlier version of the paper was presented at an Imperial College/Science Policy Research Unit seminar in May 1983, and at a Nordic Scientific Policy Council conference held in Helsinki during February 1984.

accelerator or detector technology. For example, the discoveries in 1983 of the W and Z particles[3] at CERN – the European Laboratory for Particle Physics at Geneva – were made possible by the new technique of stochastic cooling (required to produce an intense beam of antiprotons prior to colliding them with a proton beam).

Yet for each scientific problem that has been solved by a particular generation of instrumentation, several more have been generated.[4] These have then constituted the justification for scientists seeking support for the next, and usually more expensive, generation of research facility, and hence have given rise to an escalating obsolescence and succession process for scientific instrumentation. Research equipment that could previously be provided to individual university groups has become so expensive that it can only be made available on a regional, national, or even – in the most costly areas – international basis. Thus, as the Big Sciences have grown more capital-intensive, so research activity has become concentrated in ever fewer central laboratories. As we shall argue, this process of concentration poses difficulties for the peer-review process – the mechanism normally used by funding agencies in arriving at research priorities and constructing science policy.[5] It is partly for this reason that much of the work on research evaluation that has been carried out over the last seven years at the Science Policy Research Unit has focused on Big Science.[6]

In what follows, we describe the 'method of converging partial indicators' that has been developed for evaluating the scientific output from major central research facilities. To illustrate its use, results are presented from a recent study in which the scientific performance of the particle accelerators at the CERN laboratory is compared with that of other major accelerators around the world. In addition, we attempt to identify the factors explaining why some accelerators have been more successful than others in the past. Then, by analyzing which of those factors are likely to continue to structure success and failure in the field of high-energy physics, and which new factors are likely to begin exerting an influence on research performance in the future, we show how the approach enables one to arrive at conclusions about the likely prospects for major new research facilities like accelerators.

The need for research evaluation

There are perhaps three main reasons why there is a need for improved and more open methods of evaluating research performance in basic science, especially Big Science. The first relates to the fact that the rapid growth in the basic science budgets of industrialized countries between 1945 and the early 1970s has since given way to approximately level budgets or even cuts. This changed 'boundary condition' on science means that, at any one time, existing financial commitments must generally be reduced in order to free the funds to support promising new research areas and young scientists. This, we argue below, is not a task for which peer-review has proved particularly effective.[7]

Secondly, as has already been mentioned, resources have over the years become concentrated in a few central facilities. In Britain, for example, seven large centres accounted for some 65% of all expenditure in 1981/82 by the Science and Engineering Research Council on basic and applied science (but excluding engineering) – over three and a half times the total allocated in the form of peer-reviewed grants to university researchers.[8] With the annual budgets of individual research centres now running into tens or even hundreds of millions of dollars, there is, we could claim, a need not just for accountability to scientific peers, but for wider public accountability. This is only possible if systematic information on the activities and research performance of such centres is available in a form accessible to other scientists, government officials, politicians, and the public, and not just to the scientists in the specialty concerned. One solution to this problem lies in the greater use of output indicators in helping regulate the scientific system.[9]

Thirdly, there are reasons for believing that the traditional peer-review system is coming under increasing strain. One reason is that the previous pattern of scientific growth has led to the entrenchment of particular interests in decision-making bodies. For example, as fields like nuclear physics and astronomy grew rapidly in Britain during the 1950s and early 1960s, so did the level of representation of their practitioners on decision-making bodies. Since the strength of the case for additional funds made by a particular research area depends to some extent on how strongly it is represented, early established priorities have tended to become 'frozen' into the science-policy structure. In other words, because representation on decision-making bodies tends to reflect previous patterns of resource distribution, there has been a tendency towards the *reproduction* of that distribution among specialties and institutions.[10]

Another problem facing peer-review results from the concentration of research within a few centres. The successful operation of peer-review depends on the existence of a constituency of 'disinterested' peers able to provide independent expert judgements – that is, there must be sufficient scientists familiar with the research area where funds are being sought, but whose own material circumstances will be unaffected by the decision-outcome. When the number of distinct research groups working in a specialty is large, this condition is at least approximately met. However, in Big Science, when the allocation of resources to a central facility is being considered, nearly all peers will be users either of that centre (and so will benefit from a positive decision) or of a rival centre (whose own chances of obtaining funds may increase with a negative decision on the first centre). Instead of the 'free market' of scientific ideas, all competing for funding solely on the basis of scientific merit – the notion on which scientists have traditionally based their view of the peer-review procedure as a neutral disinterested process[11] – there has been a trend towards a situation of 'oligopoly' in which a few large centres and interest groups can exert a dominant influence or claim on resources.

Yet another problem with peer-review concerns its ineffectiveness as a mechanism for restructuring scientific activity – a problem that has only become apparent since science budgets ceased growing. While peer-review may be relatively successful in deciding among promising new areas of research, it is far less satisfactory when it comes to identifying declining areas and groups. This is partly because, while there is greatest scope for savings in heavily funded research areas, these are precisely the areas where cuts are likely to be most strenuously resisted by senior scientists strategically situated on science-policy committees. Social and psychological factors also play an important role here. A scientist asked to judge whether the funds of a group in her or his specialty should be reduced is likely to know members of that group personally or professionally. For that scientist, a decision whether to recommend a cut is exceedingly difficult to make, jeopardizing as it may the future livelihood of colleagues. It is certainly far harder than (and qualitatively different from) deciding whether to give additional funds – often the main type of decision that had to be made in more affluent times, and one where a negative outcome merely meant that new equipment could not be purchased, or extra researchers not recruited.

To sum up, because of (1) the trend towards approximately level budgets; (2) the heavy concentration of resources in relatively few centres; and (3) the increasing strain on the peer-review system, there is a need for more systematic external[12] evaluation of research performance to complement *but not replace* existing policy-making mechanisms. It is important to stress that such data on past performance should be seen as constituting merely one input among several that policy-makers need to consider, and also to emphasize that research-evaluation data require careful interpretation. We would not advocate that policy-makers should come to rely solely on such information in some rigid 'mechanical' way, as certain critics of our work seem to imply.[13] Research-evaluation data would, in our view, improve the effectiveness of the peer-review process, but they cannot replace it without seriously diminishing the quality of scientific decision-making.

Evaluating basic research: the method of converging partial indicators

So how have we attempted to evaluate research performance in basic science? There are four main elements to the methodology.[14]

First, it is based on an input-output approach – that is, it involves identifying and evaluating the various inputs (such as funds, researchers, and technical-support staff) and outputs (for example, contributions to scientific knowledge, education and technology), and then relating the outputs to the inputs. For basic science, some simplification is possible since the primary output is contributions to scientific knowledge. Because of this, the paper will concentrate on the evaluation of these scientific contributions (although we have also undertaken assessments of educational and technological outputs – for example from radio astronomy[15].

Secondly, our approach is institutionally focused – the unit of analysis is not the individual scientist nor the specialty, but the research centre, facility, or group. This is because the major capital-investment decisions in basic science tend to focus on institutions rather than the individual or specialty, and it is here that the rigidities within the peer-review system, at least in Big Science, largely arise.

Thirdly, because no absolute quantification of research performance in basic science is possible, our approach is comparative, with the added condition that one can only legitimately compare 'like' with 'like'. One cannot, for example, compare directly the performance of a large optical telescope with that of a radio telescope, but one can compare it with the performance of similar-sized optical telescopes at other observatories.

Finally, the approach involves the combined use of several indicators (for example, numbers of publications in international refereed journals, numbers of times those publications are cited by other scientists, numbers of highly cited papers or 'discoveries', and peer-rankings). These indicators reflect different facets of research performance, although they are, of course, to some extent interrelated. For example, publication totals give some indication of the overall scientific production of a research group (this indicator clearly favours the larger groups), while numbers of papers per researcher or per dollar reveal something about the productivity of that group – that is, its output in relation to the inputs. The average number of citations per paper gives an indication of the impact those publications have on the scientific community, while peer-rankings (where peers are asked to rank in order the performance of similar research groups according to their relative scientific contributions over a given period) provide evidence on the perceived significance of the results from different groups. Lastly, data on the distribution of highly cited papers in a particular specialty reveal which groups have been responsible for the few key 'discoveries', while data on citation totals reflect the large number of small incremental additions to the sum of human knowledge.

It should,-however, be emphasized that all these indicators are imperfect or partial measures[16] that is, they reflect partly the relative magnitude of contributions to scientific knowledge, and partly a variety of social, institutional, psychological, and other factors. The various problems with the main indicators are summarized in Table 1.

One problem with publication counts, for example, is that each publication clearly does not constitute an equal contribution to scientific knowledge. However, data on citations per paper give some indication of the average impact of a group's publications, while data on highly cited papers enable one to identify those papers representing the most significant contributions. A second problem is that publication practices vary for different types of paper, among specialties, and so on. This is one reason why comparisons can only be drawn between 'matched' research groups using broadly equivalent facilities, producing similar types of papers, and publishing in essentially the same body of scientific journals.

Use of citation analysis presents a number of technical problems – for example, incomplete coverage of journals by the *Science Citation Index*, which provides the source data for citations; however, in the Big Science specialty described here, this is not a serious problem since virtually all the journals used by high-energy physicists are scanned. There are also several more substantive problems with citations; for instance, a paper containing results subsequently found to be 'mistaken' may be heavily cited, at least until its 'mistakeness' is firmly established. However, the high number of citations can be taken to reflect its impact at the time in terms of stimulating other (sometimes fruitful) research that might not otherwise have been carried out. Here, one must distinguish between the intrinsic 'quality' of a paper and its 'impact' on scientists; only for the latter does citation frequency provide a reasonable indicator.[17] Another problem with citations is the variation in citation rates among specialties, but, as with publication counts, it can be overcome by applying this indicator only to matched research groups within a single specialty.

Although peer-evaluation is the method for evaluating research performance most favoured by scientists, even this is not without its problems as can be seen from Table 1. However, these problems can be largely overcome by using a large representative sample of peers, by employing structured interviewing techniques, by assuring evaluators of confidentiality, and by checking and allowing for any significant variations between the assessments made by different groups of evaluators. In the study reported here, for example, we compared self-rankings with peer-rankings and ascertained the magnitude of any systematic differences.

In short, the 'method of converging partial indicators' is based on the application of a range of performance indicators to matched research groups using similar research facilities, publishing in the same body of international journals subject to comparable refereeing procedures, and so on. When the indicators all point in the same direction, we regard the results of the evaluation as being relatively reliable, and certainly as

Table 1

Main problems with the various partial indicators of scientific progress and details of how their effects may be minimized

	Partial indicator based on	Problem	How effects may be minized	
(A)	Publication counts	(1) Each publication does not make an equal contribution to scientific knowledge	Use citations to indicate average impact of a group's publications, and to identify very highly cited papers	Use only indicators that yield convergent results
		(2) Variation of publication rates with specialty and institutional context	Choose matched groups producing similar types of papers within a single specialty	
(B)	Citation analysis	(1) Technical limitations with *Science Citation Index:*		
		(a) first-author only listed	Not a problem for research groups	
		(b) variations in names	Check manually	
		(c) authors with identical names		
		(d) clerical errors		
		(e) incomplete coverage of journals	Not a serious problem for "Big Science"	
		(2) Variation of citation rate during lifetime of a paper – unrecognised advances on the one hand, and integration of basic ideas on the other	Not a problem if citations are regarded as an indicator of impact, rather than quality or importance	
		(3) Critical citations		
		(4) "Halo effect" citations		
		(5) Variation of citation rate with type of paper and specialty	Choose matched groups producing similar types of papers within a single specialty	
		(6) Self-citation and "in-house" citation (SC and IHC)	Check empirically and adjust results if the incidence of SC or IHC varies between groups	
(C)	Peer evaluation	(1) Perceived implication of results for own centre and competitors may affect evaluation	(1) Use a complete sample, or a large representative sample	
		(2) Individuals evaluate scientific contributions in relation to their own (very different) cognitive and social locations.	(2) Use verbal rather than written survey so can press evaluator if a divergence between expressed opinions and actual views is suspected	
		(3) "Conformist" assessments (e.g. "halo effect") accentuated by lack of knowledge on contributions of different centres	(3) Assure evaluators of confidentiality (4) Check for systematic variations between different groups of evaluators	

Source: *Martin* and *Irvine* (1983)[6]

being more reliable than those based on a single indicator like peer-review (especially when derived in a less systematic manner, as appears to be the case at present with most research-funding agencies).[18]

The past performance of the CERN accelerators

Having examined the background to our work on research evaluation and the main features of the methodology, let us now consider some results from our recent study of the past performance and future prospects of the European Laboratory for Particle Physics. The main accelerators at CERN – which have all been proton machines – are shown in Table 2, together with the nearest equivalent facilities elsewhere in the world. In 1959, the CERN Proton Synchrotron (PS) took over from the Dubna accelerator (and before that the Berkeley Bevatron) as the world's highest-energy accelerator, although the very similar and slightly higher-energy Brookhaven Alternating Gradient Synchrotron (AGS) was competed a few months later in 1960. The Intersecting Storage Rings (ISR) – where two beams, each of about 30 GeV (i.e. 30 giga or billion electron volts) are collided head-on – was comp-

Table 2
The world's main proton accelerators (>5 GeV)

Accelerator	Began operating	Beam energy (GeV)
Berkeley Bevatron (U.S.)	1954	6
JINR Dubna (E. Europe)	1957	10
CERN PS (W. Europe)	1959	28
Brookhaven AGS (U.S.)	1960	33
ITEP Moscow (U.S.S.R.)	1961	7
Argonne ZGS (U.S.)	1963	12
Rutherford Nimrod (U.K.)	1963	7
Serpukhov (U.S.S.R.)	1967	76
CERN ISR (W. Europe)	1971	31
Fermilab (U.S.)	1972	400
CERN SPS (W. Europe)	1976	400
CERN $p\bar{p}$ (W. Europe)[1]	1981	270

[1] This is a colliding-beam facility, so its centre-of-mass energy is twice the beam energy. The same is true for the ISR in the case when the energies of the two colliding beams are identical (or, more precisely, when the moments of the colliding particles are equal and opposite).

leted in 1971, while the CERN Super Proton Synchrotron (SPS) began operating in 1976, four years after a similar accelerator at Fermilab. Finally, in 1981, the proton-antiproton ($p\bar{p}$) collider came into operation at CERN.

Bibliometric indicators

Limitations of space preclude us from presenting details of the inputs (funding, numbers of researchers, etc.) into these various accelerator centres, but they are given in full elsewhere.[19] In analyzing the outputs, we have divided the period from 1961 to 1982 into a number of four-year 'blocks'. One can see from Table 3 (covering the period 1961–64) that the CERN PS began to yield a large number of papers relatively quickly, accounting for nearly 30% of the world total of experimental high-energy physics papers in 1963–64, and overtaking the Berkeley Bevatron (which dropped from 38% to 23%). However, the CERN papers reported data from relatively simple experiments, so their overall impact was rather less than that of both the Brookhaven AGS and the Bevatron – the respective citation shares were 14.5%, 23%, and 35% in 1964. PS publications earned only 1.9 citations per paper, very low for a new accelerator – the figure for the AGS was 8.0. Moreover, in terms of highly cited papers, the Bevatron (with 33 papers cited 15 or more times in a year) and the AGS (with 24) both seem to have been responsible for a larger number of important advances than the PS which managed nine. Indeed, the three major discoveries that could have been made on a machine with the PS's energy – the identification of two types of neutrinos, the omega minus, and charge-parity violation – were all made on the AGS.

Over the next four years, the CERN PS began to benefit from a more advanced experimental programme and technical improvements to the accelerator's performance. Table 4 shows that it continued to yield more papers than any other accelerator, and, although the average number of citations per paper was still much less than for the AGS (3.1 compared with 4.6 in 1968), the total impact of the CERN papers seems to have been very similar to that of AGS papers, at least in terms of world citation-share – 26% compared with 26.5% in 1968. In terms of papers cited 15 or more times in a year, the PS had by then managed to catch up and even overtake the AGS (46 compared with 42), but the more important advances (cited 30 or more times in a year) still eluded the PS – it managed only one compared with 12 for the AGS.

For the period 1969–72, Table 5 reveals that the PS continued to produce 25% of the world total of experimental publications – much more than the AGS whose share dropped to 14.5% in 1972 following difficulties associated with a major

Table 3
Experimental high-energy physics, 1961–1964

	% of papers published in past two years		% of citations to work' of past four years	Average citations per paper	Highly cited papers: number cited n times[1]			
	1962	1964	1964	1964	n⩾15	n⩾30	n⩾50	n⩾100
Bevatron (6 GeV)	38.0	23.0	35.0	3.4	33	5	2	0
Dubna (10 GeV)	18.5	12.0	3.0	0.5	0	0	0	0
CERN PS (28 GeV)	10.0	29.5	14.5	1.9	9	1	0	0
Brookhaven AGS (33 GeV)	4.5	11.0	23.0	8.0	24	5	1	1
Moscow ITEP (7 GeV)	2.0	3.0	0.5	0.5	0	0	0	0
Rest of world	27.0	22.5	24.0	2.8	22	6	1	0
World total of papers	375 100%	535 100%	2590 100%	2.8	88	17	4	1

[1] Over the period 1961–80, an experimental paper required approximately 40 or more citations in any one year to be included in the top 1% most highly cited papers for the two decades, and 19 or more to be included in the top 5%. In this respect, n⩾15 corresponds to the top 7.8%, n⩾30 to the top 1.9%, n⩾50 to the top 0.6%, and n⩾100 to the top 0.14% most highly cited papers.

Source: *Irvine* and *Martin*[20]

Table 4
Experimental high-energy physics, 1965–1968

	% of papers published in past two years		% of citations to work of past four years		Average citations per paper		Highly cited papers: number cited n times[1]			
	1966	1968	1966	1968	1966	1968	n⩾15	n⩾30	n⩾50	n⩾100
Bevatron (6 GeV)	12.5	11.5	17.5	10.0	3.8	2.8	9	0	0	0
Dubna (10 GeV)	6.5	4.0	2.5	1.0	1.1	0.9	0	0	0	0
CERN PS (28 GeV)	32.5	25.5	28.5	26.0	3.5	3.1	46	1	0	0
Brookhaven AGS (33 GeV)	19.5	19.5	28.0	26.5	6.8	4.6	42	12	1	0
Argonne ZGS (12 GeV)	2.0	6.5	1.0	5.0	4.2	3.6	6	0	0	0
Rutherford Nimrod (7GeV)	2.5	2.5	2.5	3.5	6.0	4.6	9	0	0	0
Rest of world	24.0	30.5	20.0	28.0	3.3	3.4	38	5	0	0
World total of papers	645 100%	845 100%	4500 100%	5080 100%	3.8	3.4	150	18	1	0

[1] See note 1 to Table 3.
Source: *Irvine* and *Martin*[20]

Table 5

Experimental high-energy physics, 1969–1972

	% of papers published in past two years		% of citations to work of past four years		Average citations per paper		Highly cited papers: number cited n times[1]			
	1970	1972	1970	1972	1970	1972	n⩾15	n⩾30	n⩾50	n⩾100
Bevatron (6 GeV)	9.0	6.0	10.0	6.5	2.9	2.5	8	1	0	0
CERN PS (28 GeV)	25.0	25.0	22.5	21.5	2.6	2.3	19	1	0	0
Brookhaven AGS (33 GeV)	19.0	14.5	23.0	18.0	3.5	2.9	22	1	0	0
Argonne ZGS (12 GeV)	9.5	9.5	7.5	9.0	2.8	2.6	7	0	0	0
Rutherford Nimrod (7 GeV)	2.5	2.5	2.5	2.0	3.2	1.9	0	0	0	0
Serpukhov (70 GeV	3.0	6.0	2.5	5.0	4.8	3.0	12	4	2	1
CERN ISR (28+28 GeV)	–	2.0	–	5.0	–	12.9	13	4	2	0
SLAC (20 GeV)	5.5	8.0	8.5	12.5	6.1	4.9	21	4	2	0
Rest of world	27.0	26.0	23.0	19.5	2.4	2.0	21	6	3	0
World total of papers	950 100%	1020 100%	5270 100%	5310 100%	2.9	2.7	123	21	9	1

[1] See note 1 to Table 3.

Source: *Irvine* and *Martin*[20]

technical upgrade of the accelerator. As a result, although AGS papers still earned more citations per paper (2.9 compared with 2.3 in 1972), the PS had by then overtaken the AGS in terms of total citations (21.5% compared with 18%). However, the most highly cited papers during this period came from neither the PS nor the AGS but from newer accelerators. The PS and AGS each managed only one paper cited 30 or more times in a year, while Serpukhov, the new CERN ISR, and the Stanford Linear Accelerator (SLAC) yielded four each. There were two particularly important early results from the ISR (on the 'scaling' behaviour first seen at SLAC, and the discovery of the diffraction minimum), and its figure of 12.9 citations per paper was very high, even for a new machine. Overall, if the contributions from the ISR and PS are taken together, the figures in Table 5 suggest that by 1972 CERN had finally become the world's foremost experimental high-energy physics laboratory, a conclusion which, to judge from the *CERN Annual Report* for 1972, was evidently sensed at CERN at the time.[21]

Probably the most successful year in CERN's history (at least until 1983) came in 1973 with the discovery of neutral currents on the PS, and several other major advances made on the PS and ISR. However, the following three years were amongst the most tumultuous ever in high-energy physics, ushering in the revolutionary era of 'new physics'; and, despite CERN's promising start in 1973, the most important discoveries were still largely made elsewhere. It is certainly true that, according to the figures in Table 6, the PS and ISR continued to account for approximately 25 to 30% of the world's publications and citations, with the ISR earning an extremely high level of citations per paper (14.0 in 1974). However, in terms of papers cited 15 or more times in a year – that is, in terms of major advances – the PS and ISR with 18 and 28 were some way behind SLAC (37) and Fermilab (71). Furthermore, of the eight crucial discoveries (cited 100 or more times in a year), no less than five came from Stanford compared with one from the ISR (the confirmation of the rising total cross-section first seen at Serpukhov). Moreover, it was the Brookhaven AGS rather than the CERN PS which shared with Stanford the honour of making arguably the most important experimental advance of the 1970s – the discovery of the J/psi particle which paved the way for the 'new physics'.

In contrast, the four years that followed were more a period of consolidation, with 1977 witnessing the first experiments on the 400 GeV SPS, as well as the introduction of various second-generation detectors on the ISR. Table 7 shows that the three CERN accelerators together accounted for between 25 and 30% of papers and citations during the period 1977–80. While the impact of the PS and ISR declined somewhat from eralier years (although nowhere near as dramatically as the AGS which earned only 3% of citations in 1980 compared with 23% ten years earlier), this was compensated by the growing impact of the SPS which, with beams and detectors

Table 6
Experimental high-energy physics, 1973–1976

	% of papers published in past two years		% of citations to work of past four years		Average citations per paper		Highly cited papers: number cited n times[1]			
	1974	1976	1974	1976	1974	1976	n≥15	n≥30	n≥50	n≥100
CERN PS (28 GeV)	20.0	23.5	19.5	14.5	2.7	2.2	18	4	2	0
Brookhaven AGS (33 GeV)	13.5	8.0	11.0	10.0	2.5	3.2	15	5	2	1
Argonne ZGS (12 GeV)	9.0	6.0	7.0	4.5	2.3	2.1	5	0	0	0
Rutherford Nimrod (7GeV)	3.0	1.5	2.0	0.5	2.1	1.1	0	0	0	0
Serpukhov (70 GeV)	9.5	10.5	7.0	6.5	2.8	2.1	7	0	0	0
CERN ISR (28+28 GeV)	3.5	5.0	12.0	9.0	14.0	7.4	28	12	2	1
Fermi lab (400 GeV)	9.5	15.0	11.5	25.0	6.9	6.7	71	26	9	0
SLAC 20 GeV 4+4 GeV	7.0	9.0	10.0	17.0	4.2	7.1 (25.5)[2]	37	14	8	5
Rest of world	26.0	21.5	20.0	12.5	2.4	1.8	16	6	3	1
World total of papers	1150 100%	1180 100%	6780 100%	7740 100%	3.1	3.3	197	67	26	8

[1] See note 1 to Table 3.

[2] This is the figure for the 4+4 GeV collider alone.

Source: *Irvine* and *Martin*[20]

Table 7
Experimental high-energy physics, 1977–1980

	% of papers published in past two years		% of citations to work of past four years		Average citations per paper		Highly cited papers: number cited n times[1]			
	1978	1980	1978	1980	1978	1980	n⩾15	n⩾30	n⩾50	n⩾100
CERN PS (28 GeV)	22.0	11.5	14.5	12.5	2.2	2.2	13	2	1	0
Brookhaven AGS (33 GeV)	5.5	5.5	5.0	3.0	2.7	1.6	0	0	0	0
Serpukhov (70 GeV)	12.0	14.0	4.0	5.0	1.2	1.2	0	0	0	0
CERN ISR (31+31 GeV)	4.5	5.5	7.0	7.5	5.4	4.4	11	2	0	0
Fermi lab (400 GeV)	16.5	19.0	32.0	21.5	7.3	3.6	40	10	5	1
CERN SPS (400 GeV)	2.5	8.5	4.0	8.5	12.7	5.0	19	7	3	0
SLAC (i)20 GeV (ii)4+4 GeV	9.5	6.0	15.0	11.5	5.7 (12.5)[2]	4.4	26	6	1	1
DESY 7 GeV 5+5 GeV 19+19 GeV	4.0	6.5	5.5	15.5	5.7	8.8[3] (11.7)	36	16	4	0
Rest of world	23.5	24.0	13.0	15.0	2.0	1.9	19	5	0	0
World total of papers	1115 100%	930 100%	8190 100%	6090 100%	3.5	3.0	164	48	14	2

[1] See note 1 to Table 3.

[2] This is the figure for the 4+4 GeV collider alone

[3] This is the figure for the 19+19 GeV collider alone.

Source: *Irvine* and *Martin*[20]

that represented a significant improvement over those at the similar energy Fermilab accelerator, achieved a particularly high rate of citations per paper in 1978 (12.7). Taking the accelerators individually, the Fermilab machine seems to have contributed most over these four years with about a quarter of total world citations and 40 papers cited 15 or more times, although its performance did decline markedly over the latter part of the period – largely as a result of funding problems in the U.S. As in previous periods, crucial discoveries continued to elude CERN. The two most important discoveries (the upsilon, and parity violation) were made at Fermilab and SLAC. Even at a slightly lower level of advance, Fermilab yielded five papers cited 50 or more times and the rapidly improving DESY in West Germany four, compared with three on the SPS, one on the PS, and none on the ISR.

Finally, what has happened since 1980? In terms of experimental papers, although Table 8 shows that the world share gained by the SPS declined slightly to 15% in 1982 (reflecting its shut-down a year earlier in order to complete construction of the proton-antiproton collider), this was more than compensated by the emergence of the first results from the collider, and by an appreciable increase in ISR papers from 6% in 1981 to 9.5% in 1982. As a result, CERN's world share of experimental papers increased from 24.5% in 1980 to 33.5% in 1982. Similarly, in terms of citations, CERN users increased their fraction of the world total from 28.5% in 1980 to 33.5% in 1982. As for citations per paper, the figures for the SPS show a dramatic drop from over 12 in 1978 to under 4 in 1981 – comparatively low for an accelerator that had been operating only five years. This rapid obsolescence can perhaps best be explained by the fact that by 1980 the Fermilab accelerator (which had been completed in 1972) and the SPS had between them carried out most of the important work in the 400 GeV energy-range. If so, this illustrates the dangers of building a machine very similar in energy to one that has already been operating for several years, even if it did represent an appreciable technical improvement on that earlier accelerator. Significantly, it was the recognition of this problem of the premature obsolescence of the SPS which had been one of the factors that encouraged CERN to take the gamble involved in embarking on the proton-antiproton collider project – a gamble which by the end of 1982 had already begun to pay off with the publication of three of the world's most highly cited papers for that year (see the final column of Table 8). Overall, if the figures for the various accelerators at CERN and DESY are combined, they provide good grounds for suggesting that the onset of the 1980s heralded a European renaissance in high-energy physics, even before the dramatic discoveries of the W and Z particles on the CERN proton-antiproton collider in 1983.

Table 8

Experimental high-energy physics, post–1980

	% of papers published		% of citations to work of past four years		Average citations per paper		Highly cited papers: number cited n times[1]	
	1981	1982	1981	1982	1981	1982	n⩾15	n⩾30
CERN PS (28 GeV)	9.5	7.0	8.5	5.0	1.6	1.4	0	0
Brookhaven AGS (33 GeV)	4.5	3.0	3.0	2.5	1.7	1.4	0	0
Serpukhov (70 GeV)	15.0	11.5	6.5	6.0	1.4	1.3	0	0
CERN ISR (31+31 GeV)	6.0	9.5	8.0	10.5	3.8	4.7	1	0
Fermi lab (400 GeV)	15.5	13.0	20.5	16.0	3.1	2.7	0	0
CERN SPS (400 GeV)	16.5	15.0	12.0	16.0	3.6	3.8	2	0
CERN $p\bar{p}$ (270+270 GeV)	1.0	2.0	–	2.0	–	9.3	3	3
SLAC (1) 32 GeV (2) 4+4 GeV (3) 18+18 GeV	3.5	8.5	10.0	9.0	4.0	4.4	1	1
DESY (1) 7 GeV (2) 5+5 GeV (3) 20+20 GeV	7.0	9.0	16.5	19.5	7.5 (9.2)[2]	8.0 (9.3)[2]	6	0
Cornell CESR (8+8 GeV)	1.0	3.0	3.0	4.0	18.2	9.3	2	0
Rest of world	20.0	19.0	11.5	10.0	1.4	1.3	0	0
World total of papers	430 100%	390 100%	5190 100%	5070 100%	2.7	2.9	15	4

[1] See note 1 to Table 3.

[2] This is the figure for the 20+20 GeV collider alone.

Source: *Irvine* and *Martin*[20]

Peer-evaluation

How do all these results based on bibliometric indicators compare with the peer-review assessments made by high-energy physicists? A total of 182 researchers[22] from 11 countries were asked in the course of detailed interviews to assess the scientific performance of the world's six main proton accelerators on a 10-point scale (10=top) in terms of two criteria: (1) crucial experiments and discoveries; and (2) experiments involving more precise measurements of known particles and properties. The results are given in Table 9. The first point to note is that we found a high degree of consistency between the assessments of different groups of researchers, both across the Atlantic and between East and West.[23] Certainly, there was a 'self-ranking effect' – a tendency to rate one's own work more highly than do others – but this was not significant (except in the case of Serpukhov[24]). As can be seen, the AGS was judged considerably ahead of the PS in terms of discoveries – 9.2 compared with 6.9. This is in line with the earlier data on highly cited papers – the AGS yielded far more papers cited 100, 50, or 30 times in a year. Similarly, the Fermilab accelerator, which generated many more highly cited papers than the equivalent-energy SPS, was ranked ahead of it (at 7.2 compared with 5.7). As for the second criterion of 'precise measurement' experiments, a different pattern emerged, with the PS ranked 8.5, somewhat ahead of the AGS on 7.2. This is consistent with, for example, the data on total citations – in terms of this indicator, although the AGS was initially in the lead until 1964, by the late 1960s the PS had moved ahead, and the gap widened considerably during the 1970s.

Although the picture yielded by the various bibliometric indicators is complex and its interpretation is by no means easy (allowance has to be made, for instance, for a number of highly cited papers that were subsequently shown to be 'mistaken'), it is clear that there is a certain consistency between the results yielded by the bibliometric indicators and peer-evaluation. It is this consistency, together with the comments of a large number of high-energy physicists who read early drafts of our research papers arising from the CERN study, which leads us to conclude, as with our earlier studies of optical and radio astronomy[25], that outsiders can carry out evaluations of past research performance within individual basic-science specialties. When the work began six years ago, the general view was that this was an impossible task[26], yet most scientists now seem to accept as valid the figures our approach yields, even if they sometimes differ in their interpretation of them.

Table 9
Assessments (on a 10-point scale[1]) of main proton accelerators in terms of (a) 'discoveries' (b) experiments providing more precise measurements

		Self-rankings	Peer-rankings	Overall rankings (sample size = 169)
Discoveries	Brookhaven AGS	9.5(±0.1)	9.0(±0.1)	9.2(±0.1)
	CERN PS	7.1(±0.2)	6.7(±0.2)	6.9(±0.1)
	CERN ISR	6.8(±0.3)	5.9(±0.2)	6.1(±0.2)
	CERN SPS	5.9(±0.3)	5.6(±0.2)	5.7(±0.1)
	Fermi lab	7.4(±0.3)	7.1(±0.1)	7.2(±0.1)
	Serpukhov	3.8(±0.5)	2.6(±0.1)	2.7(±0.1)
More precise measurements	Brookhaven AGS	7.1(±0.2)	7.2(±0.2)	7.2(±0.1)
	CERN PS	8.5(±0.1)	8.5(±0.1)	8.5(±0.1)
	CERN ISR	7.3(±0.3)	6.9(±0.2)	7.0(±0.1)
	CERN SPS	8.2(±0.2)	8.2(±0.2)	8.2(±0.1)
	Fermi lab	6.3(±0.2)	6.0(±0.2)	6.1(±0.1)
	Serpukhov	4.3(±0.5)	3.5(±0.2)	3.6(±0.2)

[1] 10=top. The assessments are based on the relative outputs from the accelerators over their entire operational careers up to the time of the interviews with high-energy physicists in late 1981/early 1982.

Source: *Irvine* and *Martin*[20]

Assessing future prospects

At the end of the first SPRU work on research evaluation, another criticism came to the fore – that while our approach may reveal interesting information about the *past* performance of research facilities, it does not address the more central concern of policy-makers with the *future* prospects of rival facilities competing for funds. Consequently, addressing this criticism was one of the main tasks of the CERN project described here. This involved adding two further stages to the assessment approach used.

First, having arrived at conclusions on the comparative research performance of different accelerators, we attempted to identify in the course of the interviews with high-energy physicists the factors that had structured success and failure in the past. The most important factors reported as accounting for the differing performance of the CERN PS and Brookhaven AGS are listed in Table 10. For example, 51% of those questioned on this issue attributed the greater success of the AGS in terms of major discoveries to the bolder and more speculative approach of U.S. physicists during the 1960s, contrasting this with the more conservative and less risky approach of their European counterparts. Another equally important and closely related problem facing CERN stemmed from its international character and the need to ensure that all 12 Member States were fully represented on decision-making bodies. According to over half those interviewed, this resulted in slow and often over-conservative decisions as regards the selection of experiments for the PS. As for the respective fortunes of the CERN SPS and Fermilab accelerators, a somewhat different set of factors was identified, as can be seen from Table 11. The fact that the Fermilab accelerator has been associated with significantly more discoveries was attributed by 58% of those interviewed to its four-year lead over the SPS. Conversely, the superior record of the SPS in relation to experiments involving more precise measurements and better statistics was, in the view of 50% of interviewees, due in large part to the greater resources and higher level of technical support available at CERN.

The second additional stage incorporated in the CERN project then involved, on the one hand, analyzing further the various factors to ascertain which were likely to continue to exert an influence on research performance in the future (and which seem set to disappear), and, on the other, identifying any new factors that are likely to emerge over the coming years. On the basis of this analysis, a set of thirteen criteria was drawn up and used to assess the prospects for the various new accelerators planned to come into operation over the next decade, including LEP, the large electron-positron collider due to be completed at CERN in 1988. The criteria relate mainly to future scientific potential, but also take into account factors such as relative resource requirements and potential for the development of future research facilities. Table 12 evaluates each accelerator in terms of the criteria. It provides a convenient way of summarizing the patterns of comparative advantage (indicated by + signs) and weakness (– signs) that exist among the various projects, patterns that can then be used in helping arrive at an overall assessment of their future prospects.

While limitations of space again prevent us from going into details, some brief examples will help illustrate the approach. From Table 12, it can be seen that for the proton-antiproton collider at CERN one of the main problems that had to be overcome was technical – ensuring that enough particles collided for the event-rate to be sufficiently large to generate adequate statistics. At the time our study was

Table 10
Factors explaining the relative scientific performance of the CERN PS and Brookhaven AGS
(% of interviewees believing each factor to have been important[1])

	All interviewees (n=125)	Users of AGS (n=42)	Users of PS (n=69)	Users of AGS and PS (n=17)
1. More bold, speculative ethos of U.S. physicists – Europeans more conservative, less risky approach.	51	55	54	65
2. Greater experience of U.S. physicists Europeans had to learn to use a large accelerator	36	31	42	47
3. U.S. had better experimentalists in 1960s	30	29	29	29
4. Higher work ethic and more competitive attitude of U.S. physicists	20	29	20	35
5. Europe's 'missing generation' of physicists – emigration to U.S. of many of the best in the 1930s and '40s	11	19	14	29
6. Luck of Brookhaven in choosing right experiments	17	21	13	24
7. Better scientific management at Brookhaven-e.g. quick to respond to the unexpected	16	21	14	24
8. CERN's tendency to 'over-engineer' – e.g. detectors built bigger but later	10	10	14	12
9. Problems of CERN being multinational – e.g. slow committees, over-conservative choice of experiments	51	62	46	59
10. Social structure of European groups more hierarchical – non-scientific ('political') factors introduced	10	17	4	6

[1] These figures represent minimum values only since they are based on a content analysis of answers to a general question concerning the factors structuring the relative scientific performance of the two accelerators.

Source: *Irvine* and *Martin*[20]

Table 11
Factors explaining the relative scientific performance of Fermilab and the CERN SPS
(% of interviewees believing each factor to have been important[1])

	All interviewees (n=136)	Users of Fermilab (n=39)	Users of SPS (n=56)	Users of Fermilab and SPS (n=7)[2]
1. Fermilab had 4-year lead – SPS too late	58	56	54	29
2. Fermilab more bold, speculative experiments – CERN solid, precise 2nd-generation experiments	39	49	36	43
3. Problems of CERN being multinational – e.g. slow committees, over-conservative choice of experiments	11	10	9	14
4. More resources and technical support for CERN experiments – sometimes inadequate at Fermilab	50	49	55	43
5. SPS a much better accelerator – e.g. better beams	26	15	32	29
6. Fermilab cut too many corners building accelerator and detectors – unreliable, 'under-designed'	31	31	32	29
7. Fermilab philosophy wrong – too great a proportion of running costs channelled into new accelerator projects	14	15	13	29
8. Fermilab spread themselves too thinly on the ground with experiments, given the resources	13	21	14	14
9. Energy range of both accelerators turned out to be relatively unexciting	10	13	5	0

[1] Minimum estimate only (see note 1 to Table 10).
[2] This is a very small sample size, and the % figures may not therefore be statistically very significant.
Sources: *Irvine* and *Martin*[20]

completed, it was by no means certain that a sufficiently high luminosity would be obtained. However, the technical track-record of CERN (one of the factors determining the performance of earlier CERN machines) suggested that they would ultimately be successful, as indeed was the case a year later in 1983. In contrast, one of the main problems experienced by Brookhaven in its attempt to build the ISABELLE collider was again technical (it centred on mass-producing super-conducting magnets), but the laboratory's uneven technical track-record – in particular in upgrading the AGS accelerator – led one to be less sanguine about their prospects of solving this problem. (At the end of 1983, the ISABELLE project was finally terminated after expenditure of $200 million had already been incurred.) Another example concerns the new Stanford Linear Collider (SLC). This is likely to have only about a year to exploit its position of world-leadership (before LEP starts operating at CERN), and its users will face the disadvantages of a low event-rate, a single experimental area, and a limited variety of experiments possible on the machine. Nevertheless, their scientific track-record in exploiting previous Stanford accelerators leaves one optimistic that the SLC will succeed in generating interesting physics results, as well as providing an experimental test-bed for a radically innovative research technique.

Policy implications

Let us conclude by looking briefly at the wider significance of the above approach to evaluating basic scientific research. First, the method yields information on research performance in a form accessible not just to researchers in the specialty concerned but also to other scientists, science policy-makers, politicians, and the public. It is, for example, no longer necessary for government officials to rely exclusively on the opinions of high-energy physicists for information on how well a particular accelerator is performing, although these must, of course, continue to be given great weight.

Secondly, the method raises the possibility of tracking the performance of any major research facility, whether it be a telescope, particle accelerator, research reactor, laser facility, or oceanographic vessel. This could be undertaken relatively quickly and cheaply once the initial data had been obtained, although it would clearly be necessary to check the continued validity of the bibliometric indicators perhaps every five years or so through renewed peerevalution.[28] In case of high-energy physics, for example, it would cost some $5–10 000 per annum to keep track of all the world's main high-energy physics accelerators – a small sum compared with the total annual worldwide expenditure of well over billion dollars on particle physics. If applied to other specialties, the resulting data could help senior scientists and policy-makers spot a significant decline in the performance of

Table 12

Comparison of major accelerator projects

Summary of main comparative advantages (+) and disadvantages (–)[1]

			CERN	Fermilab	Fermilab	KEK	SLAC	Cornell	CERN
			$p\bar{p}$	1 TeV	$p\bar{p}$	TRISTAN	SLC-50	CESR-II[2]	LEP-50
		Estimated date of first experiments	1981	1985	1986	1986	1987	1988?	1988
Construction requirements	Financial	Cheapness	++	++	+	–	++	–	––
		Accessibility of resources	+++	+	+	++	+	–––	+
	Technical	Technology required relatively simple/undemanding	––	–	–	+	––	–	++
		Technical track-record of laboratory	++	–	–	•	+	•	++
Scientific potential	World lead in a new energy region	Relative increase in centre-of-mass energy	+++	•	++	•	+	–	–
		Period of world leadership	++	++	+++	•	•	–	+
	Ability of accelerator to generate physics results	Event rate	––	++	––	•	–	•	•
		Number of experimental areas	–	++	–	•	––	•	•
		Variety of experiments possible	•	++	•	–	––	–	–
		'Cleanness' (ease of interpretation) of data	–	•	–	++	++	++	++
	Ability of users to exploit accelators	Scientific track record of laboratory users	+	•	•	–	++	•	+
Future potential		Potential spin-off to accelerator physics	+	+	•	•	++	+	•
		Flexibility/potential for future development of accelerator	–	+	–	+	•	•	++

			Brookhaven	Serpukhov	DESY	KEK TRISTAN	CERN	U.S.
			ISABELLE[2]	UNK	HERA	ep	LEP-90/130	SSC
		Estimated date of first experiments	1987/8	1990?	1990?	1990?	early 1990s?	mid 1990s?
Construction requirements	Financial	Cheapness	–	+	•	•	–	–––
		Accessibility of resources	–––	++	++	•	+	•
	Technical	Technology required relatively simple/undemanding	––	•	•	•	+	–
		Technical track-record of laboratory	––	––	•	•	++	?
Scientific potential	World lead in a new energy region	Relative increase in centre-of-mass energy	––	+	+++	––	+	+++[3]
		Period of world leadership	––	++	++	––	++	++
	Ability of accelerator to generate physics results	Event rate	+	++	•	•	–	+
		Number of experimental areas	•	++	•	•	+	?
		Variety of experiments possible	+	++	+	+	–	+
		'Cleanness' (ease of interpretation) of data	–	•	+	+	++	–
	Ability of users to exploit accelerator	Scientific track record of laboratory users	++	––	•	–	+	?
Future potential		Potential spin-off to accelerator physics	+	•	•	•	+	+
		Flexibility/potential for future development of accelerator	•	+	•	•	+	?

[1] Presented in terms of a 7-point scale ranging from (+++) to (•) to (–––).
[2] Project discontinued in 1983.
[3] This assumes that a superconducting proton collider is not completed at CERN ahead of this U.S. machine.
Source: *Martin* and *Irvine*[27]

a major research centre or group (for example, because of insrumental obsolescence), and perhaps suggest to them where cash infusions were most needed, or where they should look most closely in their search for areas where commitments could be reduced in order to free the funds to support new areas and people.

Lastly, as we have seen, the method makes possible a formal, systematic and comprehensive appraisal of the future prospects for a major new research facility compared with those of rival facilities around the world. As the industrialized world has moved into an era of approximately level science budgets, so the task of deciding which of the competing claims of researchers on the limited resources available for new projects should be given priority has become ever more difficult. If policy-makers were to apply the approach outlined here to the range of proposed new research facilities for which scientists were currently seeking support, it would, we would argue, provide them with information of direct relevance to this crucial task of establishing scientific priorities.

Notes and references

1. D. de SOLLA PRICE, Philosophical Mechanism and Mechanical Philosophy: Some Notes Towards a Philosophy of Scientific Instruments, *Annali del'Institutio e Museo di Storia della Scienza di Firenze,* V (1980) 75–85. D. de SOLLA PRICE, Selaing Wax and String: A Philosophy of the Experimenter's Craft and its Role in the Genesis of High Technology, Sarton Lecture at the AAAS, May 1983. D. de SOLLA PRICE, The Science/Technology Relationship; The Craft' of Experimental Science, and Policy for the Improvement of High Technology Innovation, *Research Policy* 13 (1984) 3–20.
2. See D. O. EDGE, M. J. MULKAY, *Astronomy Transformed,* Wiley Interscience, New York, 1976.
3. See A. L. ROBINSON, CERN Vector Boson Hunt Successful, *Science,* 221 (26 August 1983) 840–42.
4. For evidence on the exponential growth of science, see e.g. N. RESCHER, *Scientific Progress,* Blackwell, Oxford, 1978.
5. cf. J. IRVINE, B. R. MARTIN, C. H. G. OLDHAM, *Research Evaluation in British Science: A SPRU Review,* a report prepared for the French Ministry of Research and Industry by the Science Policy Research Unit, University of Sussex, Brighton BN1 9RF, U. K.
6. The programme of research began in 1978 with a two-year study focusing on five British Big Science laboratories working in the areas of radio astronomy (see B. R. MARTIN and J. IRVINE, Assessing Basic Research: Some Partial Indicators of Scientific Progress in Radio Astronomy, *Research Policy,* 12 (1983) 61–90), optical astronomy (see J. IRVINE and B. R. MARTIN, Assessing Basic Research: The Case of the Isaac Newton Telescope, *Social Studies of Science,* 13 (1983) 49–86), and electron high-energy physics (see B. R. MARTIN and J. IRVINE, Internal Criteria for Scientific Choice: An Evaluation of Research in High-Energy Physics Using Electron Accelerators, *Minerva,* XIX (1981) 408–32). This was followed in 1981 by an evaluation of applied research (mechanical engineering and electronics) for a Norwegian

Royal Commission (see M. SCHWARZ, J. IRVINE and B. R. MARTIN, with K. PAVITT and R. ROTHWELL, Government Support for Industrial Research: Lessons from a Study of Norway, *R&D Management,* 12 (1982) 155–67). The CERN project described here was carried out over a 15-month period between 1981 and 1982, this paper constituting a summary of the three articles listed in references 19, 20 and 27, below. Since then, the authors have undertaken evaluations of: (1) the Norwegian Institute for Energy Technology; (2). the steel-research programme of the European Coal and Steel Community (ECSC); the mechanisms used by policy-makers in Britain for evaluating past research performance (see reference 5, above); and (4) Britain's international standing in the fields of ocean currents and protein crystallography. We should like to record the fact that the late Professor de SOLLA PRICE played a central role in helping establish our work as a legitimate area of research, and we shall always be indebted to him for his crucial early interest and enthusiasm.

7. See also J. IRVINE, B. R. MARTIN, What Direction for Basic Scientific Research? in: M. GIBBONS, P. GUMMET, B. M. UDGAONKAR (Eds), *Science and Technology Policy in the 1980s and Beyond,* Longman, Harlow, 1984, 67–98.
8. cf. ibid.
9. For an analysis of the very limited extent to which output indicators are currently used by science-policy makers in Britain, see J. IRVINE, B. R. MARTIN, C. H. G. OLDHAM, op. cit., note 5.
10. For example, the situation in Britain is now such that, within the Science and Engineering Research Council (the organization responsible for funding research in the natural sciences), one subsection of physics – nuclear physics – has its own Board, as does astronomy and space research, while all other areas of science – physics, chemistry, biology, mathematics and computing – have only one Board betwen them. The result is that Big Science consumes almost double the resources of all the remaining natural sciences together, whilst constituting only a small proportion of total scientific activity, at least in terms of numbers of researchers (see J. IRVINE, B. R. MARTIN, op. cit., note 7, for further discussion).
11. As with the concept of the 'free market' beloved by classical economists, this notion was probably always more a convenient myth than an accurate reflection of 'reality'.
12. I. e. 'external' in the sense of being carried out by outsiders to the research area under review.
13. See, for example, H. COLLINS, Scientific Knowledge and Science Policy: Some Foreseeable Implications, paper presented at Council for Science and Society Conference on Growing Points in Science Studies, Imperial College, 25 June 1983.
14. The methodology is described in detail in B. R. MARTIN and J. IRVINE (1983) op. cit., note 6.
15. J. IRVINE, B. R. MARTIN, The Economic Effects of Big Science: The Case of Radio Astronomy, *Proceedings of the International Colloquium on Economic Effects of Space and Other Advanced Technologies, Strasbourg, 28–30 April 1980,* European Space Agency, Paris (Ref. ESA SP–151, September 1980).
16. See B. R. MARTIN, J. IRVINE (1983) op. cit., note 6, pp. 66–74, for a fuller discussion of what is meant by the term 'partial indicator'.
17. The distinction between the 'quality', 'importance', and 'impact' of a publication is discussed further in ibid., pp. 69–70.
18. Clearly, the indicators do not always converge perfectly. Where there is a marked lack of convergence, the expert assistance of the scientific community is obviously required to ascertain *why* this is the case. This is another reason why research-evaluation data (in particular, bibliometric data) cannot replace peer-review but only complement it.
19. See B. R. MARTIN, J. IRVINE, CERN: Past Performance and Future Prospects – I – CERN's Position in World High-Energy Physics, *Research Policy,* 13 (1984) 183–210.
20. See J. IRVINE, B. R. MARTIN, CERN: Past Performance and Future Prospects – II – The Scientific Performance of the CERN Accelerators, *Reseach Policy,* 13 (1984) 247–84.

21. *CERN Annual Report 1972,* CERN, Geneva, 1972, pp. 11 and 26.
22. Of the 182 physicists, 169 were able to carry out this ranking. Further details of the interviews are given in J. IRVINE, B. R. MARTIN, op. cit., note 20.
23. The degree of consistency is analyzed in detail in B. R. MARTIN, J. IRVINE, op. cit., note 19.
24. See J. IRVINE and B. R. MARTIN, Basic Research in the East and West: A Comparison of the Scientific Performance of High-Energy Physics Accelerators, *Social Studies of Science,* (forthcoming).
25. See note 6.
26. Indeed, the early work was actively opposed by certain scientists. It was during these difficult times that Derek de SOLLA PRICE lent much needed support and encouragement to the authors, an act for which they are supremely grateful.
27. B. R. MARTIN and J. IRVINE, CERN: Past Performance and Future Prospects – III – CERN and the Future of World High-Energy Physics, *Research Policy,* 13 (1984) 311–42.
28. It might be the case that certain 'operators' learn to 'play' an evaluating system based purely on bibliometric indicators by changing their publication and referencing practices. While this would arguably have little effect in the case of highly cited papers, it would clearly be necessary to check the continued validity of the bibliometric indicators with periodic peer-review.