Per O. Seglen

Evaluating Biology

A Scientometric Study of a University Biology Department

NIFU skriftserie nr. 6/2001

NIFU – Norsk institutt for studier
av forskning og utdanning
Hegdehaugsveien 31
0352 Oslo

ISSN 0808-4572

Preface

The use of quantitative performance indicators in the evaluation of teaching and research at specific institutions may easily become an abuse, if the exercise is not done with care and competence. In the present report, professor Per O. Seglen deals in depth with the methodological issues that arose in connection to an evaluation of the Department of Biology at the University of Oslo in 1999. His aim is neither to criticize nor to defend the quantitative approach of the evalution, but to give insight into the considerations and decisions that were made in the process of a complex institutional evaluation.

In NIFU's view, the report has clarifying discussions of important questions that may well be neglected, but cannot be avoided, in evaluations with similar approaches. The report may thus serve as a model for the methodological work in new evalutions, in addition to being a contribution to scientometric and research policy studies in general.

Oslo, April 2001

Petter Aasen
Director

Randi Søgnen
Research Director

Author's foreword

The present report is based on the work of a committee appointed by the Department of Biology, University of Oslo, to analyze the research and teaching productivity of the various sections within that department. I want to express my sincere thanks to my fellow committee members, Inger-Hege Andersen, Yenan Bryceson, John S. Gray and Göran Nilsson, for their efforts in assembling the data, and for the many stimulating discussions we had during the study period (1998-1999). I would also like to thank the head of department during that period, Norbert Roos, for allowing me to freely use the committee's material in the preparation of this report.

Oslo, January 2001

Per O. Seglen

Contents

Summary

The research and teaching productivity of the Department of Biology at the University of Oslo has been assessed by a committee using a scientometric approach. The individual tenured member of the scientific staff was chosen as the basic unit of evaluation, and the data were subsequently aggregated to the level of fourteen sections/subsections (cell biology, biotechnology, microbiology, toxicology, genetics, physiology, botany, plant physiology, freshwater biology, limnology, marine botany, marine zoology, marine chemistry and zoology).

Scientific productivity was measured by the counting of international publications during the period 1992-1998, using a conservative fractional credit attribution that gave full credit to the project leader (the tenured staff member) regardless of the number of coauthors, except if two or more coauthors belonged to a different group/institution, in which case half credit (but never less) was given. Non-journal publications played a minor role in this department, suggesting that the ISI journal article database, which was found to be more reliable and consistent than the institutional publication lists, might be an adeqate source of publication data. Although the Botany section, with only a 63% article recovery, would need to be supplemented by publication lists, the overall 90% database recovery for the other sections would seem to be sufficient for an evaluation at the sectional level. Using a combined database/publication list indicator, the *per capita* productivity of the different department sections was found to exhibit an approximately 20-fold variability. Most sections had some unproductive staff members, but no systematic productivity differences between "field biology" and "laboratory biology" were found.

Citedness, an expression of intrascientific utility and thus a partial indicator of scientific quality, was measured by retrieval of 1992-1998 citation data for individual staff members from the ISI database, aggregated to the sectional level and expressed on a *per capita* basis. Since citedness is highly dependent on the research field, field correction factors were constructed for each department section, based on the section's weighted subject composition and the calculated mean citedness of these subjects. The field-corrected data (in which the value of a cell biology citation was reduced nearly threefold relative to a zoology citation) showed that sectional citedness varied greatly, from zero to 75 citations per staff member during the period studied, correlating moderately (c = 0.56) with sectional productivity. At the level of individual scientists, the correlation between citedness and productivity was very poor (C = 0.22). The overall citedness of the Department of Biology was calculated to be slightly above the field-corrected world average, with no obvious differences between "field biology" and "laboratory biology".

The ability of a project to attract research grants was considered as a possible partial quality indicator. The tenured staff members at the Department of Biology had, on average, an annual external grant income of 0.7 million NKK, ranging between sections from zero to 1.6

million *per capita*. Grant income correlated reasonably well with scientific productivity at the sectional level (c = 0.69) as well as at the individual level (c = 0.57), but eventually this indicator was judged not to be solid enough to be included in the final evaluation.

Three indicators were used to measure teaching performance: (1) number of lecturing/tutoring hours given by the staff; (2) undergraduate course attendance (study points x student number); (3) graduate student numbers. The first two indicators were well correlated at the sectional level (c = 0.78, and revealed a threefold difference between sections in terms of teaching performance. The variability in graduate output was even higher. At the level of individual staff members, a weak positive correlation (c = 0.38) between teaching performance and scientific productivity was observed, i.e., low scientific activity is not generally compensated by an increased teaching engagement.

Since the size of a department or a department section ought to be related to the size of its subject, an attempt was made to estimate the volume of biological knowledge associated with each deprtment/section. The volume of international research, as reflected in the annual number of scientific publications within each subject, was considered to provide a good approximation to the overall knowledge volume, since current research represents the output of past knowledge as well as the input to future knowledge. The number of publications listed within each of the 265 subject categories of the ISI database was used as the basic measure of research volume, and the subject categories were allocated, wholly or partially, to subjects corresponding to the various departments within the Science Faculty as well as to the various sections within the Department of Biology. A medical correction index was constructed to achieve a balanced partitioning of the biomedical litterature between medicine and biology.

The research volume attributed to biology was found to be about 50% larger than that of physics, although the corresponding departments had equally large scientific staffs. Biology was, furthermore, 80% larger than chemistry and nearly eight times larger than mathematics, with only 18% smaller department staffs. Even when giving equal weight to student numbers (which is particularly relevant for mathematics), it was clear that the Biology department, along with Biochemistry, was greatly understaffed relative to the other faculty departments. Within the Department of Biology, cell biology and physiology, with 30% of the staff, accounted for 67% of the biological research volume, whereas aquatic biology, also with 30% of the staff, accounted for only 7% of the research volume. There was thus a striking discrepancy between the sizes of the sections and the corresponding knowledge volumes.

The various sections' scores on each of the eight indicators (productivity, field-corrected and uncorrected citedness, grant income, teaching hours, undergraduate attendance, graduate students and knowledge volume) were expressed as per cent of the department total and presented to the department board without further processing. The relative weighting of the indicators and the final science policy decisions with regard to future sectional staff size was thus left to the board. A board-appointed committee assigned 75% weight to the three teaching performance indicators, 20% weight to the science performance indicators (weighting productivity and citedness equally), and 5% weight to the knowledge volume indicator. On the basis of the weighted indicators, the department board eventually recommended a substantial cut in the relative staff size of aquatic biology. However, where the indicators suggested a relative expansion of the cell biology staff and a relative reduction in the botany staff, the board instead recommended the opposite. Scientometry can thus be expected to have an impact on science policy only to the extent that its results are in consonance with the prevailing power balance.

Introduction

Evaluation of science and scientists is an important, yet complex issue. Although evaluation-based allocation of resources in large measure determines the extent, composition and direction of scientific activity in a society, there is no general agreement as to how science evaluation best should be done. Scientific performance can be rated in terms of productivity, quality or relevance, but all these aspects may be defined in various ways, and at least the latter two are are difficult to measure in an objective manner.

The evaluation issue becomes even more challenging when a complex institution such as a large university department is considered. Even within a department, scientists may work in widely different research fields which may be difficult to compare, and they may collaborate internally or externally to various extents. Furthermore, in addition to research, university scientists will be engaged in teaching and administration, which must be given weight when the overall distribution of resources is discussed.

The present study is an account of an evaluation of the Department of Biology at the University of Oslo, performed in 1999 at the request of the department board. A four-member committee was appointed, with the mandate of "analyzing the scientific and teaching productivity within the various sections of the department, and to present criteria and scenarios for dimensioning of the sections in the next decade." The committee chose to adopt a scientometric approach, attempting to find quantifiable indicators that would allow a fair comparison between the different sections of the department. Since many of the principles, considerations and decisions involved were of a quite general nature, the evaluation of the Department of Biology could be of interest as study case of a complex institutional evaluation.

Chapter I. Units of Evaluation

Choice of departmental units for evaluation

At the time of the evaluation, the Department of Biology was organized into eight major sections: Zoology, Botany, Physiology, Genetics, Cell Biology, Limnology, Marine Botany and Marine Zoology & Chemistry. Several of these sections could be further subdivided into distinct thematic and organizational subunits ("study directions"), on the basis of their teaching responsibilities. The Cell Biology section thus included the study directions Cell Biology, Biotechnology, Microbiology and Toxicology; the Botany section could be subdivided into Botany and Plant Physiology, the Limnology section into Limnology and Freshwater Biology, and the Marine Chemistry & Zoology section into Marine Chemistry and Marine Zoology. Since the department board wanted the evaluation to be carried out at the subsectional (study direction) level, altogether 14 organizational units, with a total tenured staff of 50 scientists, were chosen to be evaluated independently (Table 1), and will henceforth be referred to as sections.

Table 1. *Sections at the Department of Biology, University of Oslo*

Major section/subsection	Tenured scientific staff
Cell Biology	
Cell Biology	4
Biotechnology	1
Microbiology	3
Toxicology	1
Genetics	3
Physiology	6
Botany	
Botany	6
Plant Physiology	3
Limnology	
Freshwater Biology	3
Limnology	2
Marine Botany	4
Marine Chemistry & Zoology	
Marine Zoology	4
Marine Chemistry	2
Zoology	8
All sections	50
Common facilities	6

In addition to the above-mentioned sections, the Department of Biology encompasses several "common facilities" that are administered independently of the sections: an Electron Microscopy laboratory, a Phytotron (an insulated laboratory providing controlled conditions for ecological experiments), a DNA laboratory, and two biological out-stations (for Marine Biology and Alpine Ecology, respectively). These units, with a tenured staff of 6 scientists, provide service functions that are not directly comparable to the science and teaching obligations of the sections, and they have therefore been exempt from evaluation. This is clearly a debatable decision: the scientists attached to these facilities do perform science, and if the purpose of the present evaluation had been to allocate scientific resources to individual scientists or research groups, they should have been included (as they have been in one of the tables). For an evaluation at the sectional level, however, it was considered most practical to leave them out.

Choice of personnel to be evaluated

The personnel situation at a large department is very complex. Science, in particular, is not only performed by tenured scientific staff, but also by research fellows, students, visiting scientists, retired scientists, external cooperating partners, and part-time employees. Supportive (technical and administrative) staff should probably be regarded as a basic resource rather than as a means of production in the scientific-educational process. Particularly important within the part-time category are the professor-II positions, i.e., scientists who have their main position elsewhere, but who have a part-time (1/5) engagement at the Department of Biology that involves teaching (in particular graduate training) and sometimes research. Most of the professor-IIs are financed from external sources, but some are also paid by the department. Since the professor-IIs are usually appointed on the basis of their scientific merit, they can be very productive, and might account for a major fraction of a section's scientific production if they were to be included. The various sections followed different principles in this respect when compiling their sectional publication lists, and some felt that at least professor-IIs paid by the department should be fully included (the possibility of a 1/5 inclusion was apparently never considered). The evaluation committee, however, decided to exclude professor-IIs altogether from the scientific evaluation, since their research is for the most part financed and carried out elsewhere. Research papers co-authored with tenured staff would appear on the publication lists of the latter anyway.

In the evaluation of scientific performance, a "section" has thus been defined as its tenured scientific staff. This is probably the only scientifically relevant indicator of section size that can be unequivocally quantified. Non-tenured personnel (students, research fellows etc.) is usually engaged in a tutorial or collaborate relationship with tenured personnel, and their scientific output will, as a rule, be co-authored with the latter. Nevertheless, in all sections, a significant fraction of the scientific output was produced by non-tenured personnel, independently of tenured staff (Fig. 1). The fraction was only about 12% for the Department of Biology as a whole, but as high as 25% in Zoology, and 40% in the Marine Botany section. The committee reasoned that the section as such had contributed little to these publications, except by offering working space (internal expenditure resources being virtually non-existent at the Department of Biology), and therefore should not be accorded scientific merit. At this point, the present evaluation deviates from most previous scientometric reports, which tend to credit an institution for all publications that carry the institutional address. The above-mentioned sections felt, naturally, that their scientific output had been somewhat underestimated by including staff-authored publications only.

Fig. 1. Scientific productivity of tenured vs. non-tenured scientific personnel. *The number of international publications 1996-97 authored by tenured (open columns) and non-tenured (closed columns) presonnel was recorded, on the basis of institutional publication lists, for each of the major sections at the Department of Biology, University of Oslo (CEL, cell biology, biotechnology, toxicology and microbiology; GEN, genetics; PHY, physiology; BOT, botany and plant physiology; LIM, limnology and freshwater biology; MAB, marine botany; MCZ, marine chemistry and marine zoology; ZOO, zoology).*

The committee decided to evaluate *present* tenured personnel, on the basis of their past publication record, even if the latter included work done before they became affiliated with the Department of Biology. By the same token, personnel retired before 1998 were not included, altough they might have contributed substantially to the sectional output during the evaluation period. The concept of the section as a current group of people rather than as an institutional address was felt by the committee to be more future-oriented than the alternatives. Furthermore, this definition carried major methodological advantages in terms of interrelating the various indicators used in the evaluation, since most of them could be associated with actual, physical persons. The use of a personnel-based section definition is well related to the purpose of the evaluation, which was to assess how the sections ought to be dimensioned, *in terms of tenured scientific personnel*, on the basis of scientific and teaching performance. As a general rule, the goals of an evaluation should always be clearly defined at the outset, to allow the choice and construction of indicators optimally suited for their purpose.

Teaching performance was handled somewhat differently: at the sectional level, contributions from professor-IIs were included, since the section carries the major administrative burden of education, and should be given full credit for it. It was also necessary to consider fractional engagement (due, e.g., to leave of absence), and thus to relate the total teaching performance to the "net staff" involved.

Chapter II. Scientific Productivity

Scientific productivity can, in principle, be measured relatively simply by the quantification of published material. In practice it is more difficult, as a number of issues have to be faced: the choice and weighting of publication types, attribution of author credit, contents quantification, choice of time window, source of information etc. Each of these issues had to be addressed by the evaluation committee.

Sources of information: publication lists

The most obvious sources of productivity information would be the publication lists, as provided by the institution or by the authors themselves. The Department of Biology produces an annual report that contains categorized publication lists from each major section; these publications could be sorted to individual (tenured) authors and subsections. Unfortunately, it soon became evident that there was a lack of consistency in the way data were reported from the various sections (e.g., with regard to the inclusion of professor-IIs and other external affiliated personnel), as well as numerous errors. Some publications were doubly reported, i.e., in two consecutive years; other publications (found in international databases) were missing, and errors in title or author composition were not uncommon. Problems of this type occur universally in bibliometric studies, reflecting the sad fact that authors are not very exact in compiling and updating their publication lists. For example, a paper originally recorded by the author e.g. as " Nordmann, O., Fiskvik, G. & Torske, P.S (1997), Altered levels of hepatic drug metabolism in resident fish populations near oil platforms, *J. Arctic Marine Biol.*, submitted" may be later become extensively revised both with regard to title and author order/composition, and may even end up in a different journal and/or with another year of publication. If the entry is updated by adding the final page numbers only, trouble arises. Even individual author publication lists are, therefore, somewhat unreliable. Furthermore, they cannot be expected to be consistent as to how different publication types are classified.

The strength of the publication list *versus* databases is that all types of publication can be included. In an evaluation context, this is particularly pertinent in the case of conference proceedings, books and book chapters, popularized science and other publication categories that are not well covered by databases. The committee, therefore, did make use of the institutional publication lists, but only after a thorough (and time-consuming) cleanup of the data which effectively limited the chosen time window to two years (1996-1997).

Sources of information: databases

As an alternative to bibliometric information based on publication lists, several international databases are available, and can be accessed through the internet. The database most suitable for biology would probably be the one compiled by ISI (Institute for Scientific Information, Philadelphia, USA), accessible through the University of Oslo network at http://www.bibsys.no/isearch. The ISI database extends back to 1992, thereby providing a wide time window if desired, and it also furnishes citation data specified for each article. A database search was made for each individual tenured staff member 1992-1998, and the results compared with the institutional publication lists. Every article was checked to control author addresses and to eliminate homonymies, and to enable fractional credit allocation (see below). Problems were encountered only in the case of a single author, due to extensive homonymy and the lack of a Department of Biology address prior to his departmental engagement. This oversight was duly criticized upon the first presentation of the evaluation, and duly corrected in the final version.

The ISI database includes several types of document from international journals, i.e., original articles (including short original communications sometimes designated "letters"), review articles, editorials and book reviews. The latter two categories are not really scientific contributions, and were excluded from the present ISI-based evaluation, given below.

The ISI database was found to provide very good coverage of the international biological litterature, although some differences between the sections were noted. 86% of all international articles were recovered from the database, the discrepancy being mainly due to the Botany section, which had a coverage of 63%. The database coverage for biology has thus improved greatly over the fifteen years that have passed since a Dutch study found a coverage of only 30% (Moed *et al.*, 1987). In the future, it should be no problem for the Botany section to select adequate journals from the extensive ISI portifolio, in which case the ISI database would be sufficiently representative to serve as the sole source of bibliometric information for this type of evaluation.

Attribution of credit

The majority of research publications are multiauthored, and many are the result of collaborative efforts involving several research groups or institutions. Bibliometric studies, often performed at high aggregate (organizational) levels, tend to use "normal counting", i.e., with no fractional attribution of credit (everyone gets full credit), or "straight counting", where all credit is given to the first author (Lindsey, 1980; Pravdic & Oluic-Vukovic, 1986; Vinkler, 1996c). Straight counting is clearly incompatible with contemporary publishing practices, where the project leader tends to be listed last. Normal counting, on the other hand, will cause a systematic overestimation of productivity, and cannot be applied to scientific units with highly variable degrees of collaboration, as in the case of the Department of Biology.

A third bibliometric standard alternative is "adjusted counting", where the credit for a publication is divided equally between all the authors (Pravdic & Oluic-Vukovic, 1986). This would be inappropriate in the present case, since all tenured scientists at the Department of Biology are project leaders who should be credited for their papers independently of the number of coworkers included as authors. It is the scientific output of the research group that should be measured, regardless of how the research is organized internally. (If the individual coworkers were to be evaluated, the project leader should probably be attributed a constant fraction of the credit, say 40%, the remainder being divided between the other authors with, e.g., double credit to the first author). The present evaluation therefore accorded full credit to the group leader (the tenured scientist) for papers where all authors were group members. If a paper was co-authored by two different groups within the Department, each group was accorded equal (half) credit. Athough this will be unfair in cases where one group has contributed much more than another, it is beyond the capacity of a commiteee to make such fine distinctions (which will be subjective anyway).

Intra-departmental collaborations involving as many as three groups were extremely rare; in these few cases the committee chose a conservative fractionation and accorded half credit to each group, as an encouragement to cooperation. By the same argument, the principle of conservative fractional counting was also applied to extra-departmental collaborations. Collaborations are usually skewed, in the sense that the project as such is generally located at one institution, the other partner providing methods or personnel. For example, a visiting scientist in the Department of Biology would usually function as a group coworker despite the appearance of an additional institutional address on a collaborative paper; accordingly, full credit was given to the Department group. In contrast, a Department scientist working abroad was not given full credit for a collaborative paper, but was accorded no less than half credit according to the conservative fractionation principle. As a practical rule, the Department group was accorded full credit for a paper if no more than one coauthor had an extradepartmental address; if there were two or more, the Department group received half of the credit. This "conservative fractional counting" causes a moderate overestimation of overall productivity: cooperation at the national and international level is rewarded, but not excessively so. With more detailed knowledge about cooperative relationships, a more precise credit attribution might have been possible, but at a greatly increased labour investment on the part of the committee.

The conservative fractional attribution of credit was apparently well received by the staff of the Department of Biology, as no objections were raised.

Delimitation and weighting of publication types

The scientific staff of a university department may generate a wide variety of publications, ranging from original scientific papers to computer software (Murphy, 1995). There are no standard rules as how to classify, count and weight various publication types (Schwartz & Lopez Hellin, 1996). The official publication lists from the Department of Biology used 14 different publication categories, but the classification was not entirely consistent from year to year. As a starting point, the committee used a general classification of publication types as shown in Table 2. Non-scientific publications were not considered, the line being drawn at popular science and science-related public debate.

The relevance of a classification such as the one given in Table 2 will depend on its purpose. In a purely scientific evaluation, many of the categories listed will be irrelevant. However, in a comparison of applicants competing for a position, e.g., a professorship with mixed scientific, teaching and administrative duties, all of the categories may be considered. Since the mandate of the present committee was to assess "scientific productivity", a simplified science-oriented classification was adopted, with only three publication categories. *Class I* publications included original international articles and reviews (1-2 and 10-11 in Table 2), which were given full weight (1.0), and other items in international journals (12), which were given half-weight (0.5). *Class II* publications included electronic articles, proceedings and book chapters, editing and patents (3-5, 13-14 and 23), which were given full weight, as well as books (6 and 15), which were given double weight (2.0). *Class III* publications included scientific and popular articles/chapters (single-weighted) and books (double-weighted) written in Norwegian language (16-18). Since many national faunistic and floristic journals contain a mixture of original and popular articles, it was considered too difficult to distinguish between the two. Other publication categories, including abstracts and theses, were not considered.

Table 2. *A classification of scientific/professional publications*

Original scientific publications (written in international language)

1. Original articles in ISI-indexed journals (including short communications/letters and accepted manuscripts)
2. Original articles in other refereed international journals
3. Electronic original articles (published in databases)
4. Conference reports (proceedings chapters) that include original material
5. Book chapters or reviews containing original material
6. Books containing original material
7. Abstracts
8. Extended abstracts
9. Unpublished scientific manuscripts (e.g., submitted but not yet accepted) or reports

Other scientific/professional publications

10. Review articles (without original material) in ISI-indexed journals
11. Review articles (without original material) in other international journals
12. Editorials, commentaries, book reviews, letters to the editor etc. in international journals
13. Book chapters, reviews or proceedings chapters without original material
14. Editing of books or proceedings
15. Books (witout original material), including textbooks
16. Norwegian-languaged journal articles, book chapters, proceedings chapters, editorials, book reviews etc.
17. Norwegian-languaged books
18. Reports, public or institutional
19. Teaching compendia (printed or xeroxed)
20. Graduate theses
21. Doctoral theses
22. Debate articles, newspaper chronicles etc. related to science and science policy
23. Patents
24. Software (programs)
25. Popular science articles, book chapters or books

The committee eventually chose not to include Norwegian-languaged publications (class III) in the final evaluation. Science is an international endeavour, and very much a matter of communication. Authors who publish in a national language clearly do not have the international scientific community as their intended audience. If universality is accepted as an essential attribute of science, it can, therefore, be questioned whether a publication written in a minor national language should really be regarded as a scientific contribution in the strictest sense. Publications written in Norwegian may still be of high quality and of considerable national, local or popular interest, but were considered by the committee to fall outside the scope of science evaluation. The same argument would apply to locally published reports, theses and compendia that do not have a general (worldwide) distribution, regardless of the language used. In contrast, journals published in Norway, but in an international language (English) and with proper peer review procedures, were considered as international journals.

There was general agreement that abstracts should not be included in the evaluation. Although an abstract, when published, may contain interesting original information, the value of an abstract is highly transient, and it will soon be superseded by a more extensive report (if not, it is probably best forgotten). The inclusion of abstracts in the institutional publication lists was, furthermore, somewhat random. Extended abstracts, of several pages' length, should probably be included in the proceedings chapter category, but no examples of these were found in the present material. The argument regarding transient value may also to some extent apply to proceedings chapters (and indeed to scientific publications in general), but the latter usually contain full data documentation, and may sometimes remain as final publications. To check the extent of data duplication between proceedings and subsequent articles was considered to be beyond the capacity of the committee.

At the time of data assembly for this study (spring 1999), institutional publication lists were only available up to 1997, and it was felt that 1998 data were needed to make the evaluation reasonably up to date. The publication lists were, therefore, supplemented by information from the ISI database. This database includes original articles, review articles, editorials and book reviews, which were accorded fractional credit and weighted as described above, and added to the 1996-97 publication list data. The inclusion of this more selected publication sample provides additional weighting in favour of international journal papers, which was regarded as an advantage by the committee. With this in mind, it was decided to combine class I and II publications, adding the values up with the ISI data. The resulting numbers were divided by the number of tenured staff in each section at the Department of Biology, to provide a sectional productivity indicator. Although this entity has a somewhat chimaeric character, it can be regarded as a primarily publication list-based indicator, according to which there was a 25-fold variation in scientific productivity between the various department sections (Table 3).

Table 3. *Scientific productivity at the Department of Biology 1992-1998*

		Publication lists 1996-1998					ISI 1992-98	Total		
	Tenured	Int.	Proc.,	Norw.	Int +	I+P	ISI ISI	Prod.	Fraction	
Dept. section	staff	art.	books	art.	proc	/staff	92-98	/staff	index	(%)
Cell biology	4	34	5	0	39	9.8	57	14.3	12.0	12.0
Biotechnology	1	9	0	1	9	9.0	20	20.0	14.5	3.6
Microbiology	3	3	0	0	3	1.0	14	4.7	2.8	2.1
Toxicology	1	4	0	2	4	4.0	5	5.0	4.5	1.1
Genetics	3	23	14	2	37	12.3	34	11.3	11.8	8.9
Physiology	6	38	7	4	45	7.5	70	11.7	9.6	14.4
Botany	6	43	18	38	61	10.2	56	9.3	9.3	14.6
Plant physiol.	3	1	3	0	4	1.3	7	2.3	1.0	1.4
Freshw. biol.	3	19	3	22	22	7.3	33	11.0	9.2	6.9
Limnology	2	1	0	0	1	0.5	2	1.0	0.8	0.4
Marine botany	4	12	1	5	13	3.3	25	6.3	4.8	4.7
Marine zoology	4	12	7	1	19	4.8	24	6.0	5.4	5.4
Marine chem.	2	3	0	0	3	1.5	5	2.5	2.0	1.0
Zoology	8	71	4	19	75	9.4	114	14.3	11.8	23.6
All sections	50	273	62	94	335	6.7	466	9.3	8.0	100.0

The initial experience with the ISI database soon made it clear that it might provide not only a valuable supplement, but actually an excellent alternative to the use of publication lists. With the committee's agreed emphasis on international science, the database coverage of around 90% (excepting the Botany section) would seem adequate, the omissions representing very minor journals with a limited distribution. A separate counting, weighting and credit attribution of ISI-recorded publications 1992-98 was therefore undertaken. The resulting productivity index, expressed as the number of ISI publications per staff member (Table 3), showed a 20-fold variation between department sections, i.e., of the same order as the publication list-based index, despite the wider time window used (7 vs. 3 years). The two indices were well correlated (r = 0.84), i.e., there were obviously real differences between the sections. The indices also correlated well at the level of individual staff members (r = 0.89, Fig. 2), but with discrepancies, particularly among the low producers, large enough to suggest that a window wider than three years is needed to avoid random year-to-year fluctuations.

Fig. 2. Correlation between departmental publication lists and the ISI database. *Abscissa: International publications 1996-97 (including books, proceedings and book chapters) authored by individual members of the tenured scientific staff at the Department of Biology, supplemented by ISI-registered journal articles from 1998. Ordinate: ISI-registered journal articles 1992-98 authored by the same staff members.*

As a working compromise, the two indices were combined, using their arithmetic mean value as the final productivity index (Table 3). This combi-index gives extra (approximately double) weight to journal articles versus other publication types, and added (approximately double) weight to newer (1996-98) publications, which was regarded as reasonable. In future evaluations, the committee would, nevertheless, recommend a purely ISI-based productivity index for use at the sectional and higher aggregate levels. This index is easy to compile, and can be directly compared with a corresponding citation index constructed from the same database. The ISI database also contains author addresses, which is a prerequisite for proper credit attribution according to the rules established in the present study. At the level of *individual scientists*, however, an evaluation would still have to be supplemented with publication lists for comparisons to be made within a field as heterogeneous as biology.

It should be noted that the moderate time weighting applied here was introduced in a rather circumstantial manner: a direct time weighting of pure ISI data would have been a more rational approach. In any case, the time weighting illustrates an important technical point: that bibliometric indicators can be tailored, by parameter weighting, to suit the purpose of a particular evaluation. A warning is probably also in order: "customized bibliometry" obviously provides a possibility to introduce systematic bias, and it will be the ethical responsibility of the evaluator to ensure that misuse does not occur.

Fig. 3. Scientific productivity of individual scientists within different sections at the Department of Biology. *Each symbol represents the productivity of a single member of the tenured scientific staff, as measured by the combined productivity index (cf. text). CB, cell biology; BT, biotechnology; MI, microbiology; TX, toxicology; PH, physiology; GE, genetics; BO, botany; PP, plant physiology; FW, freshwater biology; LI, limnology; MB, marine botany; MC, marine chemistry; MZ, marine zoology; ZO, zoology.*

Fig. 3 summarizes the productivity data for individual scientists as well as department sections, using the combined productivity index. It can be seen that there are large differences between individuals within most of the sections. Practically all sections, even those with a high average productivity, have staff members who produce very little. About one-quarter of the staff at the Department of Biology had published only 0-2 ISI-registered articles during the seven-year study period, which must be regarded as unacceptable for university employees with an obligation to perform research.

Interestingly, no systematic differences between "laboratory biology" and "field biology" were observed in terms of productivity, nor were there any obvious differences with regard to the other bibliometric indicators used, somewhat contrary to the expectations among the department staff. Indeed, the present evaluation has apparently helped to eradicate some of the prevailing mythology regarding the incommensurability of these two major divisions of biology, by showing that they can in fact be measured by the same yardstick.

Choice of evaluation window

All members of the tenured staff at the Department of Biology had publication records beginning well before 1992; the 1992-98 window offered by the ISI database was thus suitable for the evaluation. If staff members with a shorter publication record had been included, a "per year" index could have been used. A seven-year index was considered adequate, although one staff member with a strong past publication record argued that the evaluation window should have stretched even further back in time. The committee was rather inclined to think that a shorter time period might have been better, to place more emphasis on current research, but such emphasis was in fact achieved by the combined index eventually used (with added data from 1996-1998). Ideally, an evaluation window should be no shorter than five years, to avoid random year-to-year fluctuations. A reasonably wide window is also recommendable for the measurement of citedness, to allow due credit to durable articles relative to more ephemeral contributions (Moed *et al.*1987).

During the work with the ISI database, it was noted that 1998 articles continued to accumulate throughout the first half of 1999. It is, therefore, important that evaluations using this database should allow at least a half-year lag before closing the file for a given calendar year.

Contents quantification

In the present evaluation, all journal articles were given equal weight, regardless of their length or contents. "Salami slicing" (the splitting of a scientific study into as many short publications as possible) is, however, a well-known publication strategy, and it has been shown that journal articles may vary as much as a 100-fold in their actual scientific contents (Seglen, 1996). An assessment of scientific productivity should, therefore, ideally include contents quantification. However, within the Department of Biology, the types of research performed were considered to be too heterogeneous to allow comparable quantifications.

Chapter III. Citations

What is measured by a citation?

A citation to a scientific document means that its contents has been used, and referred to, in some other scientific document. The citation can thus be regarded as a receipt for usage, and the overall citedness of a document can be regarded as an indicator of its overall *utility* in scientific research (Garfield, 1979). Since intrascientific utility is an aspect of scientific quality, the citation score can be used as a partial quality indicator. It should be stressed, however, that citedness tells nothing about other, more important aspects of scientific quality, like originality and solidity.

To place citedness in its proper place as a performance indicator, a classification of scientific quality aspects may be useful (Table 4). There is general agreement about *originality* and *solidity* as essential quality aspects (Chase, 1970; Gulbrandsen & Langfeldt, 1997). We have also added *informativity* (presentation quality), to emphasize the importance of scientific results being presented in a clear, intelligible and balanced manner (Hemlin, 1993). Whether intra- and extrascientific *utility* should be regarded as quality aspects is more arguable: they are both highly dependent on factors outside the control of the performing scientist. The argument is stronger for intrascientific utility, in the sense that any scientific contribution is an integral part of the overall scientific process. Extrascientific utility, on the other hand, is almost totally determined by outside society, but may be judged the central quality aspect and ultimate justification for science by those who finance it. Since both the conception and possible categorization of scientific quality is thus clearly perspective-dependent, we have chosen middle ground and classified the two utility aspects as *extrinsic* qualities, to distinguish them somewhat from the more fundamental intrinsic qualities of science.

The practical value of a classification such as the one given in Table 4 is that different aspects can be weighted independently, depending on the context and purpose of the evaluation. For example, extrascientific utility is of paramount importance in applied research, but plays little or no role in basic research. The present evaluation, which concerns basic research, should ideally have emphasized intrinsic scientific qualities, but unfortunately no suitable bibliometric indicators are available, and the time and resources at the committee's disposal were insufficient for subjective quality assessments. Our quality evaluation was, therefore, limited to the aspect of intrascientific utility, as expressed by article citedness. This is a major weakness which should be remedied in future evaluations.

Table 4. *Aspects of scientific quality*

Intrinsic qualities

Originality
- Theme (offroad, niche, novel combination, unexplored area)
- Problem (novel problem)
- Difficulty (attacking hard problems)
- Methods (novel developments, improvements, novel applications)
- Theory (original - and well supported! - hypotheses and theories)
- Results (new knowledge)

Solidity
- Data quality (clear, obvious, large effects, many experiments, adequate statistics)
- Methodological quality (adequate methods; advanced methods)
- Control (adequate positive and negative control experiments, checking and excluding alternative explanations)
- Information power (well-defined and solid conclusions)

Informativity
- Clarity (well organized and comprehensible problem formulation, results and conclusions)
- Objectivity (critical evaluation of own data; balanced evaluation of other research, fair credit attribution)
- Knowledgeability (expertise, broad knowledge and insight, relevant and representative reference choice)
- Technical quality (clear, well-organized and informative figures and tables)

Extrinsic qualities

Intrascientific utility
- Relative citedness within own field
- Overall citedness; journal impact
- Accessibility (type of publication)
- Cooperativity (national and international collaborations)
- Invited lectures
- Research prizes and grants
- Honorary titles and positions

Extrascientific utility/relevance
- Inventions and patents
- Products (including software)
- Prizes and grants for applied research
- Private and governmental assignments
- Industrialization based on inventions

Citation bias

Even if regarded purely as a utility measure, the use of citedness as an indicator in scientific evaluation is fraught with problems (Table 5). A scientific article usually rests on a knowledge base consisting of hundreds of other articles, but due to space limitations only a fraction of these are actually acknowledged by a reference (Margolis, 1967; MacRoberts & MacRoberts, 1996). This prepares the ground for considerable bias as to who is cited and who is not, and uncited does not necessarily mean unused. The selection of references is neither random nor fair: whereas some knowledgeable scientists may honour originality and quality by citing those who deserve it, this is by no means the rule. Inexperienced scientists tend to refer to a recent rather than to an early report of a phenomenon, and to chose review articles and other secondary sources rather than original articles for documentation (MacRoberts & MacRoberts, 1996). The increasing use of litterature databases with an incomplete coverage of the past tends to aggravate the problem. It should, however, in all fairness be stressed that the primary purpose of a reference is to document, not to reward.

Table 5. *Motives, problems and biases in reference selection*

1. The primary criterion is not quality, but utility within research
2. Incomplete referencing due to journal space limitations
3. Poor knowledge of primary literature
4. Citation of secondary sources (e.g., reviews) rather than of primary publications
5. Reference copying
6. Established knowledge is not cited ("obliteration by incorporation")
7. Argumentative citation (mainly self-supportive)
8. Flattery (citation of editors, potential referees etc.)
9. Show-off (citation of "hot" papers)
10. Conventions (methods are cited; reagents are not)
11. Personal communications are not recorded as citations
12. Self-citation
13. In-house citation (friends and close colleagues)

It is, unfortunately, relatively common to lift references from other articles without reading them. This is particularly the case with references to old articles which are difficult to obtain in original, and is sometimes revealed by misspelling variants occurring more frequently than the bibliographically correct reference (Seglen, 1998). Both reference copying and reference-based litterature retrieval will tend to favour well-established references. On the other hand, some research results are so well established as to be regarded as common knowledge, no longer referred to. This phenomenon is known as "obliteration by incorporation" (Merton, 1968; Cole & Cole, 1973), and has the paradoxical effect that the most important scientific articles are rarely cited. Other citing conventions likewise create bias: in the biochemical litterature it is, for example, customary to give full reference to methods, but not to reagents used. Frequently used methods can generate very high citation scores: some of the most highly cited articles both in Norway and in the world are methodological. On average, however, methodological articles are not cited more than other articles: methods which are not used by others are not referred to, no matter how ingenious they may be (Garfield, 1979).

A large fraction of the references in a scientific article are associated with the discussion and interpretation of the results. Ideally, this should be a balanced *pro et contra* discourse, but in practice scientists strive to find references that support their own conclusions (Gilbert, 1977; Brooks, 1985) - often a necessity in order to get the results published. Original, unconventional papers may therefore be less cited than mainstream research. Contrary to common belief, negative citations are rare (references to Fleischmann and Pons' herostratic "cold fusion" paper, Fleischmann & Pons, 1989, being a notable exception): disagreement tends to breed silence rather than debate (Chubin & Moitra, 1975; Price, 1981).

It seems quite common to include complimentary references (e.g. in the introductory general review of the field) to influential colleagues who are likely to be used as referees. If journal editors or board members have performed relevant research, they are guaranteed to be cited. Many authors also try to be fashionable by referring to the "latest news" within a field regardless of its relevance, thereby attempting to place themselves in the research front (Line & Sandison, 1973; Gilbert, 1977).

A scientific paper is usually built upon the authors' previous work, which makes self-citations scientifically relevant (Garfield, 1979), but for the purpose of evaluation they are not very useful. Self-citations make up about 50% of all citations on a world basis, and may account for the majority of citations to little-cited articles (Seglen, 1989a). ISI does not distinguish self-citations from other citations, and they may also be difficult to recognize in the database indices, where articles are listed under first author only (the project leader is usually last author). In the net version of the database, however, citing documents are presented with a full bibliographic reference, which makes it possible to make the distinction. In the present evaluation, self-citations, which made up about 30% of the citations to publications from the Department of Biology, were excluded.

In-house citations are relatively common, reflecting the exchange of information and methods between colleagues within an institution. The citations are however, often complimentary and therefore biased: reference may be given to the colleague's use of a method rather than to the original description of the method. In highly socialized scientific communities, where transfer of information occurs mainly through personal contact rather than through the litterature, even a national bias may be prominent. It is obvious that agreements on mutual citation could have large effects on citation scores, but there is no evidence that such agreements actually exist.

Database limitations

The recording and retrieval of citations in the ISI database is associated with several technical problems (Table 6). Although the database regularly indexes 7-8,000 journals (about 5,200 in the expanded Science Citation Index, about 1,400 in the Social Science Citation Index, and about 1,100 in the Arts & Humanities Citation Index, with some overlap between the indices), this is still a small fraction of the estimated world total of about 130,000 scholarly journals (Andersen, 1996a). The degree of coverage varies considerably between different fields: some years ago, it was estimated to be about 90% for chemistry but only 30% for biology (Moed *et al.*1987). In an Australian study, fewer than 20% of the country's academic publications were found to be printed in journals covered by the ISI database (Murphy, 1995). The situation has improved considerably in recent years, both in terms of increased journal coverage in the database, and due to a globally increased tendency to publish internationally. For a research-oriented institution such as the Department of Biology, the overall coverage (for articles written in English) in the ISI database was thus about 85%.

A major limitation of the ISI database is that books are not included as source items, despite their prominent role in many research fields. Whereas only about 5-15% (depending on the field) of the citations within the natural sciences are given to books or book/proceedings chapters, as many as 50-70% of the citations within the social sciences may be to book items (Andersen, 1996b; Sivertsen, 1993). Among a set of highly ranked mathematical publications, about 40% were not included as source items in the ISI database, mainly because they were published in books. Interestingly, these non-ISI publications were cited twice as highly as the rest !

Table 6. *Technical problems with the ISI citation database*

1. The journal coverage in the database is incomplete
2. Different research fields are unequally covered by the database
3. Books are not included as source items in the database
4. The journal set included in the database may vary with time
5. Different database products differ in contents
6. The database has a language bias in favour of English language
7. The database is biased in favour of US journals
8. Delayed registration of citations
9. Many misprints (up to 25%)
10. Inconsistent foreign language spelling (e.g., æ, ø, å)
11. Synonymy (several variants of the same article)
12. Homonymy (several authors with the same name, e.g., in Japan)

The portifolio of journals included in the ISI database varies somewhat from year to year, which may have a disturbing influence on long-term statistics. For example, an apparent decline in the productivity of Norwegian chemists some years ago, causing considerable science-political consternation, was eventually shown to be due to the removal of the Norwegian chemistry journal *Kjemi* from the ISI database (Sivertsen, 1991). It should also be known that different database products from ISI (e.g., partial, national science databases) may include different journal sets, thus compromising comparisons across databases (Moed, 1996).

The incomplete database coverage may promote various types of bias, e.g. in relation to field, nationality or language. The ISI database has a clear preference for English language and for North American journals (Moed *et al.*, 1987), thus discriminating against countries with a significant number of national-language journals (Germany, France, Russia). For example, the Social Science Citation Index was shown to include only two German journals, as compared to 542 in a German social sciences database (Artus, 1996). As many as 85% of the citations in the two German journals were to other German journals; similarly, French journals had about 70% national citations. A national/language bias (in part due to self-citations) is thus not basically an Anglo-American phenomenon (Lange, 1985; Andersen, 1996a; Narin & Hamilton, 1996), but the Anglo-American dominance of the only internationally used citation database does create a problem. Even among the journals included in the ISI database, a random set of American journals were cited twice as highly as the German journals, and five times more than the Russian journals (Seglen, 1997a). US scientists receive about one-half of all citations (Braun *et al.*, 1996; May, 1997), and have an overall citedness 30% above the world average (Braun *et al.*, 1996), undoubtedly in large measure reflecting the national and language bias of the ISI database (Møller, 1990; Narin & Hamilton, 1996; Andersen, 1996a).

A number of more technical database problems may significantly affect the outcome of a citation-based evaluation. The recording of citations can be considerably delayed, and citations are not always indexed under the year they were given. Misprints are fairly frequent (reportedly of the order of 25%), often due to errors in the original references (Wade, 1975; Evans *et al.*, 1990). One highly cited paper was thus found to be entered under 70 different synonyms in the ISI database (Seglen, 1989a). Complex author names may generate several false synonyms, as may the inconsistent transliteration of unusual letters, such as the Norwegian æ, ø and å. Even the authors themselves may not be entirely consistent in this respect. A related problem is *homonymy*, i.e., several authors with the same name, including initials. Homonymy is particularly extensive among the japanese: for example, K. Suzuki is indexed in the ISI database with several hundred articles per year.

Research field effects

A number of factors that determine the overall citedness of a publication are highly dependent on the research field (Table 7). For example, the citedness within a field is a direct function of the average number of references per article in that field. Biochemical publications, which tend to contain twice as many references as mathematical publications, are thus cited twice as often on this basis alone. In addition, the citedness is dependent on the obsolescence of the article relative to the time window used to record the citations. A short-term window (e.g., the last three years) captures twice as large a fraction of the citations to the relatively short-lived biochemical articles as it captures citations to the more durable mathematical articles. Reference number and obsolescence thus combine to make the average biochemist cited four times as much as the average mathematician (Moed *et al.*, 1985). Within the arts and humanities, article references are used infrequently, leaving these and related research fields virtually uncited (Hamilton, 1991).

The size and dynamics of a research field may affect citation rates in various ways. In very small, closed fields, where the volume of citable material is too small to saturate the article reference lists, the field citation rate (the number of citations per article per year within the field) will be proportional to field size (the number of publishing research groups/authors). Once the reference list capacity has been exceeded, the field citation rate will be independent of field size: the numbers of citing and cited articles are both proportional to the field size, hence their ratio (corresponding to the citation rate) will be constant (Gomperts, 1968). However, a large field will display a greater range in absolute citedness than a smaller field; the maximal citation score attainable will thus be higher in the large field (Seglen, 1989b) (the minimal score will be the same, namely zero). The citation scores of top groups within different fields are, therefore, not comparable, although the average groups are.

These field size considerations apply only to fields of constant size. If a field is rapidly expanding, the number of citing articles will be high relative to the amount of citable material, and the field citation rate (the chance of being cited) will similarly be high (Hargens & Felmlee, 1984; Vinkler, 1996a). In a declining field, the inverse condition prevails, and the citation rate will be low. Citation-wise it is, therefore, a profitable strategy to jump onto the current bandwagon, although scientifically an original angle is likely to bring more progress.

Table 7. *Research field effects that affect citation rates*

1. Number of references per article within the field
2. Obsolescence of references relative to time window for citation recording
3. Field size (determines the maximal citation rate obtainable)
4. Field dynamics (field expansion or contraction)
5. Interfield relations (e.g., basal vs. applied)
6. Subfield microheterogeneity

The most important research field effect is probably the ability of a field to become cited by adjacent fields. For example, clinical medicine draws heavily upon basic biomedical research, but not *vice versa*. As a result, articles within basic medical fields are cited several times more often than articles within clinical medicine (Narin *et al.*, 1976; Folly *et al.*, 1981; Seglen, 1989b; Seglen, 1997a). In Norwegian biological and paramedical disciplines, our national bias towards applied sectors like fisheries and fish farming has been proposed as one possible explanation for our relatively low national citation averages (Aksnes *et al.*, 2000; Sivertsen & Aksnes, 2000).

Field effects may extend even to the subdiscipline level, meaning that even a moderately complex project will define its own specific citational field, which *a priori* determines the probability of being cited (Seglen, 1992a). The citation rates of scientists working on different subjects are, therefore, not directly comparable, and the development of objective field corrections at the level of individual scientists/groups would be unduly demanding. It has been suggested that a correction for field effects could be made by simply dividing a citation score by the weighted average citedness of the journals in which the cited papers are published (Moed *et al.*, 1987; Schubert & Braun, 1986), but this would effectively punish authors for publishing in the most highly cited journals within their field. Field factors based on authors' reference lists (Vinkler, 1996b) are better, but still selective and author-dependent. A first requirement for field correction factors must be that they be objective, i.e. calculated on the basis of the *total* journal portifolio that defines a field. For the evaluation of an actual scientific unit, it will, furthermore, be necessary to chart that unit's fractional composition relative to pre-defined scientific fields (usually defined by a given journal set). In the present evaluation, these principles have been used to construct field correction factors at the sectional level of the Department of Biology, as will be described below.

Citational variability

A final technical problem in the use of citation scores for evaluation purposes is the intrincic variability in citation data, even for a single author/group. Articles by the same author have been found to exhibit a large variability in citedness, distributing in a lawful manner according to a negative exponential function (Seglen, 1992b). This does not mean that citedness is a random affair, since a given article is cited at a highly consistent rate from year to year (Seglen, 1994). What it apparently does mean is that the intrascientific utility of a document is determined by its contents rather than by its author. The distribution of citedness within any journal follows the same negative exponential law, confirming that the article contents, rather than its journal address, determines its citedness (Seglen, 1992b; Seglen, 1994).

The intrinsic variability in article citedness means that there is always an overlap between authors, and that a large number of articles (of the order of fifty) is required to obtain a consistent citedness value (Seglen, 1994) and to establish the significance of, e.g., a twofold difference (Seglen, 1992a). Analysis of large materials, or of author groups, makes it clear that systematic differences in citedness between authors do exist (Seglen, 1994) , but it is not certain whether these reflect actual differences in the utility of the research, or field effects, database biases etc. Most evaluations are applied to a limited time period, during which very few research groups will produce a publication volume large enough to make a citation analysis meaningful. The evaluation of citedness should, therefore, be confined to higher aggregate levels (sections, departments, institutions, nations).

Citation data from the Department of Biology

As will be evident from the discussion above, the present evaluation committee had strong reservations regarding the use of citation-based performance indicators. However, since the staff at the Department of Biology expressed great interest in use of citations, it was decided to include such data in the evaluation. All citations to ISI-registered articles from the department staff during the period 1992-98 were therefore recorded, each published paper being examined individually to allow the subtraction of self-citations (which made up about 30% of the citations). Like in the case of publication (productivity scoring), a conservative fractional credit attribution was applied.

***Table 8.** Citations to journal articles from the Department of Biology 1992-1998*

Dept. section	Tenured staff	ISI-art. 92-98	Citations (-self)	Cit. /staff	Cit. (%)	Field factor	Corr. cit.	Corrected cit/staff	Corrected cit. (%)
Cell biology	4	57	442	111	**19.4**	2.79	158	40	**10.4**
Biotechnology	1	20	85	85	**3.7**	1.96	43	43	**2.8**
Microbiology	3	14	145	48	**6.4**	2.20	66	22	**4.3**
Toxicology	1	5	15	15	**0.7**	1.50	10	10	**0.7**
Genetics	3	34	145	48	**6.4**	2.67	54	18	**3.5**
Physiology	6	70	293	49	**12.9**	1.96	150	25	**9.8**
Botany	6	56	30	5	**1.3**	1.11	27	5	**1.8**
Plant physiol.	3	7	115	38	**5.0**	1.28	90	30	**5.9**
Freshw. biol.	3	33	257	86	**11.3**	1.14	225	75	**14.8**
Limnology	2	2	0	0	**0.0**	1.14	0	0	**0.0**
Marine botany	4	25	68	17	**3.0**	1.16	59	15	**3.9**
Marine zoology	4	24	93	23	**4.1**	1.10	85	21	**5.6**
Marine chem.	2	5	5	3	**0.2**	1.27	4	2	**0.3**
Zoology	8	114	585	73	**25.7**	1.06	552	69	**36.2**
All sections	50	466	2278	46	**100.0**	1.54	1479	30	**100.0**

The net numbers of citations thus compiled for each staff member were added up for each section, and the sectional citation score was calculated on a *per capita* (i.e., per staff member) basis, and also as a percentage of the Department total (Table 8). The sectional *per capita* citedness varied from zero to over one hundred, with the highest score for the most basic field, cell biology. The high score for freshwater biology was due entirely to one highly cited staff member, whereas zoology had three highly cited staff members. Clearly, with the small number of groups involved here, sectional performance is highly influenced by individual performance.

Since the focus of the present evaluation was the dimensioning of sections, the number of citations accumulated by each section relative to the whole Department, expressed in per cent, was felt to be the most relevant way of expressing the citedness data (Table 8, fifth column, boldfaced).

Table 9. *Mean citedness (impact factor) 1995-96 of biologically relevant ISI subject categories (from Journal Citation Reports 1996; number of journals in parentheses)*

ISI subject category	Category impact factor
Biochemistry & molecular biology	3.25 ± 0.32 (227)
Biochemical research methods	1.82 ± 0.18 (21)
Biology	1.64 ± 0.29 (56)
Biology miscellaneous	1.28 ± 0.27 (21)
Biotechnology & appl. microbiol.	1.47 ± 0.14 (71)
Chemistry	1.37 ± 0.21 (110)
Cell biology	3.41 ± 0.52 (107)
Developmental biology	3.80 ± 1.12 (23)
Ecology	1.37 ± 0.14 (79)
Entomology	0.79 ± 0.16 (57)
Genetics	2.95 ± 0.49 (78)
Limnology	1.05 ± 0.29 (11)
Marine & freshwater biology	1.04 ± 0.09 (66)
Microbiology	2.37 ± 0.38 (65)
Microscopy	1.08 ± 0.13 (10)
Mycology	0.70 ± 0.12 (12)
Ornithology	0.61 ± 0.08 (14)
Physiology	2.07 ± 0.42 (60)
Plant sciences	1.22 ± 0.15 (126)
Toxicology	1.49 ± 0.24 (53)
Virology	2.18 ± 0.38 (19)
Zoology	0.66 ± 0.06 (96)

Sectional field correction factors

As discussed above, article citedness is highly dependent on the research field; citations within fields as disparate as botany and cell biology are, therefore, not strictly comparable. In an attempt to correct for this, factors expressing the relative citedness of each relevant field were constructed. The necessary field information was obtained from ISI's *SCI Journal Citation Reports 1996* (JCR), which classifies scientific journals under 245 subject categories and provides impact factors (mean number of citations per article per year during the first two years following the publication year) for each journal. The journal impact average for each subject category was taken as an indicator of overall category citedness (category impact factor; Table 9). This approximation was considered adequate for the present purpose, but for a more exact computation of category impact factors, the publication volume of each journal (available as 1996 source items in the JCR) should also be taken into account.

In a second step, the research profile of each section within the Department of Biology was examined (on the basis of publication data), to see how the category impact factors of Table 9 could best be combined to construct appropriate field correction factors for each section. The result is shown in Table 10. The SCI subject categories *Biochemistry & Molecular Biology* and *Biochemical Research Methods* have been merged under the name *Biochemistry* (impact 3.13), and the categories *Biology* and *Biology miscellaneous* have been merged under the name *Biology* (impact 1.54). This general Biology category was included in all sectional field factors with a weight of 0.2; other categories and weightings are given in Table 10.

It should be noted that the ISI/SCI journal impact factors (and hence the computed field correction factors) are extremely short-term, based only on citations given during the first two years after the publication year. Dynamic fields like cell biology and genetics have, therefore, inflated journal impact factors relative to "slow" fields like botany and zoology. Since the actual citation data collected in the present study were relatively long-term (1992-98), the more dynamic fields will be somewhat overcorrected by the application of journal impact factor-based field factors (a computation of proper field factors for the whole seven-year study period was considered to be beyond the capacity of the committee). On the other hand, other field differences within the Department of Biology are so large that uncorrected citation data would be highly misleading. The committee chose, therefore, to present both field-corrected and uncorrected citation data (Table 8). Eventually, the department board decided to accord equal weight to both data sets, thus effectively opting for a partial (50%), and relatively conservative, field correction.

Table 10. *Field correction factors (estimated average field citedness) for the various sections within the Department of Biology*

Section	SCI subject category	Subj.categ. impact factor	Weight	Field correction factor (subj.cat.i.f. x weight)
Cell biology	Cell biology	3.41	0.4	1.364
	Biochemistry	3.13	0.2	0.626
	Microscopy	1.08	0.1	0.108
	Developmental biol.	3.80	0.1	0.380
	Biology	1.54	0.2	0.308
	Total		**1.0**	**2.786**
Biotechnology	Biotechnology	1.47	0.4	0.588
	Biochemistry	3.13	0.2	0.626
	Virology	2.18	0.2	0.436
	Biology	1.54	0.2	0.308
	Total		**1.0**	**1.958**
Microbiology	Microbiology	2.37	0.8	1.896
	Biology	1.54	0.2	0.308
	Total		**1.0**	**2.204**
Toxicology	Toxicology	1.49	0.8	1.192
	Biology	1.54	0.2	0.308
	Total		**1.0**	**1.500**
Genetics	Genetics	2.95	0.8	2.360
	Biology	1.54	0.2	0.308
	Total		**1.0**	**2.668**
Physiology	Physiology	2.07	0.8	1.656
	Biology	1.54	0.2	0.308
	Total		**1.0**	**1.964**
Botany	Plant sciences	1.22	0.2	0.244
	Mycology	0.70	0.4	0.280
	Ecology	1.37	0.2	0.274
	Biology	1.54	0.2	0.308
	Total		**1.0**	**1.106**

Table 10, ctd.

Plant physiology	Plant sciences	1.22	0.8	0.976
	Biology	1.54	0.2	0.308
	Total		**1.0**	**1.284**
Freshwater biol.	Marine freshw. biol.	1.04	0.8	0.832
	Biology	1.54	0.2	0.308
	Total		**1.0**	**1.140**
Limnology	Limnology	1.05	0.4	0.420
	Marine freshw. biol.	1.04	0.4	0.416
	Biology	1.54	0.2	0.308
	Total		**1.0**	**1.144**
Marine botany	Marine freshw. biol.	1.04	0.7	0.728
	Plant sciences	1.22	0.1	0.122
	Biology	1.54	0.2	0.308
	Total		**1.0**	**1.158**
Marine zoology	Marine freshw. biol.	1.04	0.7	0.728
	Zoology	0.66	0.1	0.066
	Biology	1.54	0.2	0.308
	Total		**1.0**	**1.102**
Marine chemistry	Marine freshw. biol.	1.04	0.4	0.416
	Chemistry	1.37	0.4	0.548
	Biology	1.54	0.2	0.308
	Total		**1.0**	**1.272**
Zoology	Zoology	0.66	0.3	0.198
	Ecology	1.37	0.3	0.411
	Ornithology	0.61	0.1	0.061
	Entomology	0.79	0.1	0.079
	Biology	1.54	0.2	0.308
	Total		**1.0**	**1.057**

What is the level of citedness at the Department of Biology in an international perspective? The 466 ISI-registered articles published by the department staff during 1992-98 received a total of 2278 citations during the same period, i.e., 4.9 citations/article. Most articles published in 1998 remained uncited, leaving six real citation years and an average of 2.8 citation years per article (assuming a uniform age distribution), i.e., 4.9/2.8 = 1.7 citations per article per year. This compares favourably with the estimated departmental field factor of 1.54 (the average of the field factors for all 50 staff members; Table 8), which should represent the global annual citedness of a research field of the same subject composition as the department. A citation score equal to, or above the world average (at a ratio of 1.1) must be considered satisfactory, considering that the US dominance of the database tends to place the rest of the world below this average.

Relationship between productivity and citedness

At the level of individual staff members, the average article citedness values ranged from zero to more than twenty citations/article. As shown in Fig. 4A, the correlation between article citedness (without field correction) and overall scientific productivity (ISI articles) was poor (correlation coefficient = 0.22), in agreement with previous studies where self-citations were similarly subtracted. The relatively low numbers of articles dealt with here makes article citedness a somewhat spurious parameter; cf. the great spread at the lower productivity levels. At higher levels of productivity, the values tend to converge towards a level somewhat above the departmental and global average (at 4.9 and 4.3, respectively).

Fig. 4. Correlation between citedness and scientific productivity for individual staff members at the Department of Biology. *(A) Correlation between the number of ISI articles published 1992-98 and the average (uncorrected) article citedness during the same period. (B) Correlation between productivity (index values from Fig. 3) and total, field-corrected citedness 1992-98.*

The total (cumulative) number of citations collected by each staff member during the seven-year period would be a more robust (and more fair) measure of citedness. This parameter is also a (weighted) measure of productivity, and as such naturally correlates quite well with the individual publication index (correlation coefficient = 0.70; Fig. 4B), calculated a previously outlined (Table 3). The differences in total (field-corrected) citedness between individual staff members are, in absolute terms, much greater than the differences in productivity, ranging from zero (five cases) to more than 100 (nine cases).

There are also great differences in citedness between the different departmental sections. These differences are brought out very clearly in a correlation diagram of sectional citedness *versus* productivity (per staff member) (Fig. 5). The two sections with the lowest productivity - limnology and marine chemistry - are also the lowest in citedness, and an additional cluster of five sections - plant physiology, microbiology, toxicology, marine botany and marine zoology - shows low scores on both indicators. The citedness of the relatively productive botany section is also surpringly low, even after correction for the dominant position of mycology within this section. The high citation score of freshwater biology is due to a single staff member, who is cited a hundred times more often than the rest of his section.

Fig. 5. Correlation between scientific productivity and citedness (per capita) for the different sections within the Department of Biology. *The diagram is based on the productivity indices of Table 3 and the citedness indices (corrected citations/staff member) of Table 8. CB, cell biology; BT, biotechnology; MI, microbiology; TX, toxicology; PH, physiology; GE, genetics; BO, botany; PP, plant physiology; FW, freshwater biology; LI, limnology; MB, marine botany; MC, marine chemistry; MZ, marine zoology; ZO, zoology.*

Chapter IV. Journal Impact

A widely used indicator in evaluations of science is the *journal impact factor* (Garfield, 1972), which represents the average citedness of the articles in a journal. This indicator owes its popularity to the fact that impact factors for individual journals are readily available through ISI, and that adding them together is the easiest form of science quantification conceivable. The rationale for using impact factors is the belief that publication in high-quality journals reflects the quality of the science itself.

Table 11. *Problems associated with the use of journal impact factors*

1. Journal impact factors measure intrascientific utility (citedness), not overall scientific quality
2. Journal impact factors do not correspond numerically to perceived journal quality
3. Journal impact factors are not representative of the individual journal articles, correlating poorly with actual article citedness
4. Journal impact factors are highly research field-dependent
5. The impact factor is a function of the number of references per article in the field
6. Research fields with short-lived litterature are favoured
7. The impact factor depends on research field dynamics (expansion/contraction)
8. Small research fields tend to lack high-impact journals
9. Relations between fields strongly determine the journal impact factor (e.g., clinical vs. basal)
10. The database does not correct for author self-citations
11. A journal's selective self-citation will inflate its impact factor
12. Incomplete database coverage: journals not included as source items may be assigned an impact factor, but are deprived of journal self-citations
13. Books are not included as database source items
14. Citations to "non-citable" items are erroneously included in the database, inflating the impact factors of e.g. journals with a correspondence section
15. Reviews are much cited and inflate the journal impact factor
16. Supplement volumes are little cited and deflate the journal impact factor
17. Long articles are favoured
18. Short publication lag is favoured
19. The database is dominated by American and other English-language journals
20. Preference for national-language references favours English-language journals
21. Variable journal set in database
22. Authors' choice of journals is (was) not primarily based on impact factors

While there may be a grain of truth in this assertion, converting it into a quantitative indicator is problematic. Firstly, although journal impact factors may be acceptable (if not very good, see below) indicators of journal utility, they are unlikely to reflect other quality aspects of the journals' articles (Table 4). Furthermore, whereas journal impact factors within a field may span a hundred-fold numerical range, expert ratings of journal quality rarely exceed a twofold range (Korevaar & Moed, 1996). The direct use of journal impact factors as numerical indicators of scientific quality would, therefore, be highly misleading, even if a general correlation were to be present.

Secondly, a major problem in the use of impact factors, even as a utility measure, is the extreme heterogeneity of a journal's contents with regard to the property expressed by the impact factor, i.e., citedness (Seglen, 1992b). The individual articles in a journal tend to show a very skewed (negative exponential) citedness distribution, the general rule being that 15% of the articles account for 50% of the citations. The most cited half of the articles is in fact cited ten times more often than the least cited half. The journal's impact factor is, therefore, representative only for a small fraction of its articles, and applying it to all of them would seem like a relatively irrational form of evaluation. "Science deserves to be judged by its contents, not by its wrapping" (Seglen, 1994).

Thirdly, journal impact factors are fraught with a number of technical problems (Table 11), some of which are the same as in the use of citations, discussed in the previous chapter. Prominent among these are the field effects: journals in fields with low average citedness, like botany and zoology, will have low impact factors, whereas journals in fields with high average citedness, such as cell biology and genetics, will have high impact factors (Seglen, 1997b). Since the impact factor is based on short-term citedness (citations during the first and second year after the publication year), journals in dynamic fields will have higher impact factors than journals in more stable fields. Field size may introduce a large bias in the sense that scientists in a large field may have access to journals of a much higher impact than the top journals of a small field (Seglen, 1989b); this cannot be compensated for by field correction factors.

Many of the database problems are also evident at the journal impact level: for example, the ISI database favours English-language journals, and particularly American journals (Moed *et al.*, 1987). "Foreign" journals may not be assigned an impact factor at all, although they may be as highly cited as the Anglo-American journals (Korevaar & Moed, 1996). National citation bias, evident e.g. as high journal self-citation rates, will tend to inflate the impact factors of the latter (Lange, 1985; Andersen, 1996a; Narin & Hamilton, 1996).

A particular systematic error in the way ISI calculates impact factors may have large effects: ISI records citations to all types of document, but only regular articles and reviews are classified as citable documents and entered into the denominator of the impact factor (Magri & Solari, 1996; Moed *et al.*, 1996). Journals with a significant element of editorials, correspondence etc. may thereby inflate their impact (Moed & Van Leeuwen, 1996).

Similarly, journals which contain many highly citable items like reviews and other long articles will receive high impacts (Seglen, 1992a; Moed *et al.*, 1996; Bourke & Butler, 1996). One journal recently stopped publishing supplement volumes, because these, although much in demand, contained many little-cited abstracts which lowered the overall impact factor of the journal (Zetterström, 1999).

Since journal impact factors were judged to be neither representative nor reliable as indicators of any aspect of science, the committee decided not to use them at all in the present evaluation. This was the aspect of the evaluation report that received the most comments: several staff members felt that publication in high-impact journals would be a good indicator of scientific quality, which may to some extent, and in a very general sense, be true. The basic problem is that research published in journals of lower impact may be just as good.

Chapter V. Research Grants

The ability of a scientist or a project to gather research grants, prices/awards and other honours is often regarded as an expression of project quality. Most of these honours are too spurious or incomparable to be useful as quantitative science indicators, but research grants are a regular and necessary aspect of scientific activity which can also be readily quantified. It could be argued that grants to free, scientist-initiated research to some extent reflect project quality, whereas program grants, industrial support etc. might indicate extrascientific utility. The committee had considerable reservations about this, since grant acquisition may depend as much on the time and energy invested, the financial channels sought, personal connections etc. as on the intrinsic qualities of the project. Furthermore, the spending of a large amount of money on an unproductive project is not necessarily an asset. Ideally, an evaluation should balance scientific results against the resources used, in which case grant income belongs on the input side rather than on the output side.

At the request of one committee member it was, nevertheless, decided to include grant income as a scientometric indicator in the present study. The Department of Biology staff was asked to supply information about all types of external support, i.e. for running expenses, equipment, research fellow salaries etc., received in 1998. No attempt was made to distinguish between program grants (earmarked for specific topics) and free grants, since the distinction between the two can be quite blurred, especially within the Norwegian Research Council system. The requested information was supplied willingly, suggesting that a total input/output analysis of research establishments should be feasible. In the present study, the major institutional costs (facility maintenance, running expenses for administrative and supportive functions, staff salaries etc.) were not considered. Furthermore, the internal resources for research at the Department of Biology are so small, and evenly distributed, that they can safely be ignored as a research input variable. Grant income can, therefore, be regarded as a good indicator of project-specific research expenditure.

The Department of Biology staff demonstrated an impressive capability of acquiring external grants, with an average of 0.77 million NKR per staff member in 1998 (Table 12). Cell biology and zoology both had a grant income of about 1.5 million NKR per staff member. The three least productive sections, limnology, plant physiology and marine chemistry, also had the lowest grant income. Physiology had a surprisingly low income score relative to productivity (Fig. 6), and could probably collect more grants. In general, the correlation between scientific productivity and grant income at the section level was quite good (correlation coefficient = 0.69; Fig. 6).

***Table 12.** 1998 grant incomes to the various sections at the Department of Biology*

Department section	Tenured staff	Grant income total (NKR)	Grant income per staff (KKR)	Fraction (%)
Cell biology	4	6 449 283	1 612	16.8
Biotechnology	1	600 000	600	1.6
Microbiology	3	1 516 260	505	4.0
Toxicology	1	688 445	688	1.8
Genetics	3	2 856 227	952	7.5
Physiology	6	2 392 356	399	6.2
Botany	6	3 331 514	555	8.7
Plant physiology	3	65 449	22	0.2
Freshwater biology	3	2 685 627	895	7.0
Limnology	2	0	0	0.0
Marine botany	4	1 876 922	469	4.9
Marine zoology	4	3 030 357	758	7.9
Marine chemistry	2	742 029	371	1.9
Zoology	8	12 118 920	1 515	31.6
All sections	50	38 353 389	767	100.0

***Fig. 6. Correlation between grant income and scientific productivity for the different sections within the Department of Biology.** The diagram is based on the 1992-98 productivity indices of Table 3 and the 1998 grant incomes of Table 12. CB, cell biology; BT, biotechnology; MI, microbiology; TX, toxicology; PH, physiology; GE, genetics; BO, botany; PP, plant physiology; FW, freshwater biology; LI, limnology; MB, marine botany; MC, marine chemistry; MZ, marine zoology; ZO, zoology.*

At the level of individual staff members, the correlation between grant income and scientific productivity was more moderate (correlation coefficient = 0.57; Fig. 7). Viewed as an input/output relationship, there would still seem to be a clear tendency for the productivity to increase as a function of the investment. However, it should be noted that the time scales are not comparable, and any causal relationship would be the other way round: the more productive projects during the period 1992-98 received more grants in 1998. In general, there is a linear relationship between investments and output in science (McAllister & Wagner, 1981), and at the research group level it has been shown that scientific productivity is directly proportional to group size (Seglen & Aksnes, 2000). Since fellowship salaries are part of the grant income in the present study, the correlation with income in Fig. 7 would to some extent reflect a correlation with group size. However, the relationship between project productivity and project investments is, in practice, reciprocal, and can only be fully characterized through a longitudinal study.

***Fig. 7. Correlation between grant income and scientific productivity of individual staff members at the Department of Biology**. The diagram is based on the 1992-98 productivity indices of Fig. 3 and the 1998 grant income of each staff member.*

It should be pointed out that there are some striking exceptions to the general correlation in Fig. 7: for example, two of the projects with an income of less than one million NKR were at least as productive as the singular 7-millon NKR project and other multi-million projects. It is well known that the cost requirements of different types of research can be widely different, and it is, therefore, difficult to draw any conclusions relating to grant income from a data set as heterogeneous as the present one. This was also the opinion of the department board, who chose to ignore the grant income indicator in their final implementation of the results from the present evaluation.

Chapter VI. Teaching

Three indicators were chosen to document teaching performance: Firstly, the total number of hours spent on lectures and tutoring, at all study levels, by each staff member and section at the Department of Biology in 1998 was used as a general teaching performance indicator. Secondly, an "undergraduate level" indicator was calculated by multiplying the weight (number of study points) of each undergraduate biology course with the number of students taking that course in 1997. This indicator, which measures undergraduate course attendance expressed as "student points", correlated well with the number of teaching hours given at each section (correlation coefficient = 0.78; Fig. 8). Thus, there seems to be a good balance between the educational needs, as expressed by the study volume (student points), and the education offered (teaching hours). The student points could not be broken down to the subsection level, but are presented at the sectional level in Table 13. Thirdly, the number of graduates (cand. scient.) under supervision at each department section in 1996-97 (presented as the annual mean number) was used as a "graduate level" indicator.

Fig. 8. Correlation between undergraduate course attendance (student points) and total teaching hours (all levels) at the major sections of the Department of Biology. *For each undergraduate course taught in 1997, the course weight (number of study points) was multiplied by the number of students enrolled, to obtain a "student points" value. The correlation between student points and the hours of teaching given per staff member in 1998 at each of the major Department of Biology sections is shown (CEL, cell biology, biotechnology, toxicology and microbiology; GEN, genetics; PHY, physiology; BOT, botany and plant physiology; LIM, limnology and freshwater biology; MAB, marine botany; MCZ, marine chemistry and marine zoology; ZOO, zoology).*

Table 13. *Teaching performance of the various sections at the Department of Biology*

		Teaching hours, all levels			Undergraduate level			Graduate level		
Department section	Teaching staff	Teaching hours	Hours /staff	Fraction (%)	Student points	Points /staff	Fraction (%)	Graduate students	Grad. /staff	Fraction (%)
Cell Biology	8.1	7331	3496	**23.0**	985	122	**25.4**	21.0	12.2	**21.7**
Cell biology	3.4	3243	954	**10.2**				10.0	2.5	**10.3**
Biotechnology	1.0	673	673	**2.1**				6.0	6.0	**6.2**
Microbiology	2.5	2255	902	**7.1**				2.0	0.7	**2.1**
Toxicology	1.2	1160	967	**3.6**				3.0	3.0	**3.1**
Genetics	2.2	1903	865	**6.0**	369	167	**9.5**	6.5	2.2	**6.7**
Physiology	5.7	4030	707	**12.6**	487	85	**12.6**	11.0	1.8	**11.3**
Botany	8.5	5680	1315	**17.8**	723	85	**18.7**	15.5	3.2	**15.9**
Botany	4.5	3794	843	**11.9**	491	109	**12.7**	11.5	1.9	**11.8**
Plant physiology	4.0	1886	472	**5.9**	232	58	**6.0**	4.0	1.3	**4.1**
Limnology	5.0	1733	687	**5.5**	197	39	**5.1**	3.5	1.5	**3.6**
Freshwater biol.	3.0	1079	360	**3.4**				1.5	0.5	**1.5**
Limnology	2.0	654	327	**2.1**				2.0	1.0	**2.1**
Marine Botany	3.5	1940	554	6.1	87	25	**2.2**	3.5	1.0	**3.6**
Mar. Chem. Zool.	6.4	4288	1212	**13.4**	263	41	**6.8**	11.5	3.4	**11.8**
Marine zoology	4.4	3418	777	**10.7**				9.5	2.4	**9.7**
Marine chemistry	2.0	870	435	**2.7**				2.0	1.0	**2.1**
Zoology	7.2	4989	693	**15.6**	763	106	**19.7**	25.5	3.2	**26.2**
All sections	47.1	31894	677	**100.0**	3874	82	**100.0**	97.5	2.0	**100.0**

Data on staff fraction engaged in teaching (1998), teaching hours (1998), undergraduate student points (course points x student no., 1997) and graduate students under education (1996-97) were obtained from the department administration. In the calculation of graduates/staff, integral rather than fractional staff numbers were used.

For a fair assessment of the teaching performance of individual staff members, staff with an extended leave of absence were supplemented with corresponding data from the previous year (1997). When aggregating the data for teaching hours to the sectional level, contributions from non-staff members (e.g., professor-IIs) were also included. The sectional teaching hours were expressed relative to the size of the staff actually engaged in teaching, i.e., correcting for leave of absence and the inclusion of professor-IIs. The number of graduates was given relative to the tenured staff of the section.

As shown it Table 13, the total teaching performance, as measured in hours per year per staff member, exhibited a threefold variation between sections. The variability in graduate output was even higher. Fig. 9 depicts the covariation of these two indicators for each department section. Although the overall correlation was poor (correlation coefficient = 0.38), it is clear that the scientifically weak sections, limnology, marine chemistry and plant physiology, did very little teaching as well. The teaching performance of the marine botany and freshwater biology sections was also very low.

***Fig. 9. Correlation between graduate student output and overall teaching performance at the various sections of the Department of Biology**. Abbreviations as in Fig. 6.*

Fig. 10. Correlation between teaching performance and scientific productivity of individual staff members at the Department of Biology*. The diagram is based on the 1992-98 productivity indices of Fig. 3 and the number of teaching hours in 1998 (supplemented with 1997 data for staff members with a reduced teaching engagement in 1998).*

The correlation between the teaching performance and the scientific productivity of each individual staff member is shown in Fig. 10. A weak positive correlation can be seen (correlation coefficient = 0.38), which effectively kills the prevailing myth that University staff members who produce little science do so because they spend most of their time teaching. Although there are notable exceptions, the tendency is that those who are most active in research are also most active in teaching.

Chapter VII. Dimensions of Knowledge: The Sizing of University Departments and Sections

University functions relevant to the size of departments and sections

The role of a University is basically to produce, preserve and propagate knowledge. Society's needs for education tend to place major emphasis on the last of these three functions, as reflected in the fact that the Universities' budgets are in large measure based on student numbers. Since the undergraduate teaching at the Department of Biology is given as open courses with no student number restrictions, course enrolment (measured here as student points, which correlate well with the number of teaching hours offered, cf. Fig. 8) can be taken to indicate the "market need" for specific types of biological education. Graduate production (Table 13) would serve the same purpose. The indicators used in the present study should, therefore, provide a fair basis for evaluating section size in relation to teaching performance.

The function of Universities as knowledge producers is defined by the obligation of scientific staff members to spend 40% of their time on research. The research performance indicators presented above, will, presumably, provide some indication as to how well each section fulfils its research obligation, and hence as to whether the section size is scientifically justified.

However, the third function of the University - as a custodian of existing knowledge - should also be taken into account when the dimensioning of departments and sections is considered. The current expansion of biological knowledge can best be described as explosive, and it has become a major challenge to maintain competence in all subjects required for adequate teaching of biology at the University level. The sheer volume of knowledge that needs to be handled by a section, in particular in relation to teaching, should to some extent be reflected in the size of the staff.

Research volume as an indicator of knowledge volume

How can the relevant knowledge volumes be measured? One possibility is to use the volume of published research as an indicator of knowledge volume. Published research represents new knowledge, and can be regarded both as a function of existing knowledge and as a premise for future knowledge generation. Since, furthermore, University teaching is supposed to be research-based, the volume of research produced within a certain field may be an acceptable indicator of the volume of knowledge relevant to research and teaching within that field.

Information about the numbers of publications within various fields can be found in the ISI database, which classifies journals (and their article numbers) under 245 different subject categories. An "indexed search" through the *Bibsys* link (http://www.bibsys.no/isearch) for all 1992-98 articles within each subject category (stepwise retrieval using the truncated subject category terms A*, B* etc.) was, accordingly, performed. Since new articles are added daily to the database, the total search operation had to be carried out within a short time span (a few days).

Altogether, 47 of the indexed subject categories were found to be applicable to the Department of Biology. Several other biology-related subjects were considered to be too far removed from the subject profile of the Department of Biology to be included in the present analysis (*agriculture, agriculture dairy animal science, agriculture soil science, biophysics, clinical neurology, dermatology venereal diseases, environmental studies, food science technology, forestry, horticulture, infectious diseases, nutrition dietetics, obstetrics gynecology, ophthalmology, otorhinolaryngology, rheumatology, transplantation, water resources*). Many of the subject categories were relevant to several of the department sections, and some were also relevant to other departments (specifically, the Department of Biochemistry) and even to other faculties. Distribution of knowledge credit at the subsectional level was considered too difficult in the case of the marine, botanical and cell biological subjects, necessitating the use of the higher-order units Aquatic Biology (freshwater biology, limnology, marine zoology, marine chemistry and marine botany), Botany (botany and plant physiology) and Cell Biology (cell biology, toxicology, microbiology and biotechnology).

Since much of the basic biological research and knowledge is shared with the medical field, which (in Oslo) belongs to a different faculty, some distribution of knowledge credit between the faculties must be performed. For this purpose, a "medical index" for each subject category was introduced, based on the medical keywords *human, patient, disease, therapy* and *clinic** (truncated term). The percentage of articles within each subject category that contained one or more of these keywords was defined as the category's medical index, and the articles were subtracted from the total, leaving "non-medical", i.e., biological, articles. The index was calibrated against a large study material of Norwegian microbiological articles that had been expert-classified as medical (38%) or non-medical microbiology (Aksnes *et al.*, 2000; Seglen & Aksnes, 2000). When applied to the ISI subject category "Microbiology", our medical index identified 37% of the articles as medical, i.e., in agreement with the expert classification. In comparison, the index identified 84% of oncological articles as medical, but only 1% of the articles in subject categories like ornithology or limnology. The index would, therefore, seem to be reasonably valid.

Some of the ISI subject categories were combined to obtain subjects suitable for partitioning between the departments and sections sharing them. These shared subjects included *molecular biology, medical biology, ecology* and *general biology* (Table 14).

Based on an assessment of the curricular volumes taught at the different departments/sections, molecular biology was partitioned with 2/3 to the Department of Biochemistry and 1/3 to the

Cell Biology section at the Department of Biology, minus 1/3 of the biotechnology subject which was given to the Genetics section. Medical biology was divided between the sections of Physiology (2/3) and Cell Biology (1/3). Ecology was partitioned with 1/3 each to Aquatic Biology, Botany and Zoology, whereas general biology was partitioned with 1/6 to each of the major sections included in this analysis. Other ISI subject categories were assigned individually to their appropriate section.

The resulting "knowledge volume assignment" is shown in Table 15. The basic biological subject Cell Biology accounted for an astonishing 42% of the international biological research published during the study period, and Physiology for another 25%. These figures undoubtedly reflect the fact that the current "biological revolution" takes place mainly within molecular and cellular biology/physiology. The research activity and knowledge growth in this area proceeds at an unprecedented rate, placing heavy demands on the sections responsible for the control and transmission of this knowledge.

Table 14. *Sorting of ISI-registered 1992-98 articles under biological subjects shared between several University departments and sections*

Shared subject	ISI subject category	No. of articles	Medical index (%)	Biological articles
Molecular biology	Biochemistry molecular biology	411 376	25.8	305 169
	Biochemical research methods	12 082	27.7	8 735
	Biomethods	20 314	22.9	15 661
	Biotechnol appl microbiol	79 543	19.1	64 324
	Total	**523 315**		**393 889**
Medical biology	Cardiovascular system	111 186	52.7	52 586
	Cardiac cardiovascular systems	35 036	58.1	14 685
	Endocrinology metabolism	91 595	44.9	50 513
	Gastroenterology & hepatology	35 875	49.2	18 233
	Hematology	129 948	49.9	65 168
	Neurosciences	171 940	35.0	111 843
	Reproductive biology	8 704	37.0	5 484
	Reproductive systems	13 823	40.0	8 289
	Respiratory system	37 561	62.0	14 277
	Urology nephrology	70 742	53.4	32 988
	Total	**706 410**		**374 066**
Ecology	Ecology	**48 359**	3.9	**46 478**
General biology	Biology	93 886	13.1	81 582
	Biology miscellaneous	15 329	16.0	12 879
	Total	**109 215**		**94 461**

All articles registered in the ISI database 1992-1998 under the given subject category are listed. The medical index and the net number of biological articles are calculated as described in the text.

These attempts to calculate the dimensions of biological knowledge should be regarded as exploratory, and can undoubtedly be done better. Nevertheless, the majority of the committee endorsed the basic notion of knowledge as a quantifiable entity, relevant to the sizing of University units. One committee member disagreed with the use of international publication volume as an indicator, and felt that academic priorities should be based on other factors such as national research interests (e.g., with a Norwegian emphasis on marine biology). Opinions among the department staff were strong on this issue, and eventually the department board decided to give a rather moderate weight (5%) to the knowledge volume indicator.

Table 15. *Distribution of ISI-registered articles 1992-98 between subject categories corresponding to Department of Biology sections*

Department/section	Subject category	No. of articles	Meidical index	Biological articles	Fraction (%)
Cell Biology	Anatomy morphology	8 121	23.9	6 177	
	Cell biology	137 777	30.2	96 231	
	Cytology histology	1 215	43.0	692	
	Developmental biology	20 556	17.5	16 957	
	Microbiology	76 055	37.3	47 652	
	Microscopy	7 187	15.1	6 103	
	Immunology	137 470	59.4	55 858	
	Toxicology	41 536	31.8	28 321	
	Virology	25 633	64.1	9 197	
	Oncology	121 293	83.5	20 007	
	General biology (1/6)			15 744	
	Medical biology (1/3)			124 689	
	Molecular biology (1/3)			124 149	
	Total			**551 777**	**42.0**
Genetics	Genetics heredity	93 096	42.5	53 546	
	Biotechnology (1/3)			21 441	
	General biology (1/6)			15 744	
	Total			**90 731**	**6.9**
Physiology	Physiology	80 034	18.9	64 899	
	General biology (1/6)			15 744	
	Medical biology (2/3)			249 377	
	Total			**330 020**	**25.1**
Botany/plant physiol.	Plant sciences	99 275	6.1	93 237	
	Mycology	7 125	23.6	5 447	
	Ecology (1/3)			15 493	
	General biology (1/6)			15 744	
	Total			**129 921**	**9.9**
Aquatic biology	Marine freshwater biology	40 716	4.0	39 078	
	Limnology	8 205	1.4	8 090	
	Fisheries	16 426	6.3	15 399	
	Ecology (1/3)			15 493	
	General biology (1/6)			15 744	
	Total			**93 804**	**7.1**
Zoology	Zoology	45 089	5.5	42 627	
	Ornithology	6 096	1.0	6 035	
	Parasitology	14 400	26.1	10 637	
	Entomology	27 867	4.1	26 724	
	Ecology (1/3)			15 493	
	General biology (1/6)			15 744	
	Total			**117 260**	**8.9**
Department of Biology	**Total**			**1 313 513**	
Dept. of Biochemistry	Molecular biology (2/3)			248 298	

Table 16. *ISI-registered articles 1992-98; non-biological subject categories*

Subject	ISI subject category	No. of articles
Pharmacy	Pharmacology pharmacy	**169 767**
Astrophysics	Astronomy astrophysics	**62 284**
Physics	Biophysics	75 488
	Crystallopgraphy	33 581
	Mechanics	47 531
	Optics	64 785
	Physics	126 991
	Ph. applied	145 914
	Ph. atomic molec. chem.	58 111
	Ph. condensed matter	123 550
	Ph. fluids plasmas	26 904
	Ph. mathematical	32 139
	Ph. nuclear	40 979
	Ph. particles fields	34 919
	Spectroscopy (1/2)	18 179
	Thermodynamics (1/2)	9 644
	Total	**838 715**
Informatics	Computer sci. artific. intell.	19 726
	C.s. cybernetics	5 367
	C.s. hardware architecture	29 180
	C.s. information systems	24 524
	C.s. interdiscipl. appl.	31 423
	C.s. software graph. progr.	33 047
	C.s. theory methods	31 246
	Total	**174 513**
Geophysics	Geochemistry geophysics	35 483
	Geosci. interdiscipl. (1/2)	29 655
	Total	**65 138**
Geography	Geography	**19 884**
Geology	Geology	11 252
	Geosci. interdiscipl. (1/2)	29 655
	Mineralogy	10 338
	Paleontology	7 773
	Total	**59 018**
Chemistry	Chemistry	227 162
	Chem. analytical	78 923
	Chem. applied	30 253
	Chem. inorganic nuclear	61 656
	Chem. organic	94 769
	Chem. physical	127 758
	Electrochemistry	19 904
	Polymer science	69 170
	Spectroscopy (1/2)	18 179
	Thermodynamics (1/2)	9 644
	Total	**737 418**
Mathematics	Mathematics	78 424
	Math. applied	63 812
	Math. miscellaneous	2 998
	Statistics probability	25 186
	Total	**170 420**

Department and subject sizes within the Science Faculty

The question of organizational unit size relative to knowledge volume is of interest also beyond the Department of Biology. At the University of Oslo, the distribution of University resources to individual departments is determined at the faculty level; i.e., the board of the Science Faculty (Det matematisk-naturvitenskapelige fakultet) decides how big the Department of Biology should be relative to other science departments. It would seem quite relevant to suggest that knowledge volume should be one criterion used in this decision-making process.

Using the ISI subject categories, international publication volumes for subjects corresponding to the various Science Faculty departments can be calculated in the same manner as for the Department of Biology (Table 16). Only a few categories were considered to require partitioning: *spectroscopy* and *thermodynamics* were shared equally between Physics and Chemistry, and *geosciences interdisciplinary* was shared between Geology and Geophysics.

Table 17. *Relationship between department size, student number and research/knowledge volume at the Science Faculty of the University of Oslo*

	Tenured staff		Students (all levels)		Internat.publ. 1992-98	
Department	No.	%	No.	%	No.	%
Biology	54.7	17.3	623	13.5	1 313 513	34.0
Physics	53.9	17.0	376	8.1	838 715	21.7
Mathematics	44.7	14.1	1180	25.5	170 420	4.4
Chemistry	44.3	14.0	575	12.4	737 418	19.1
Informatics	36.3	11.5	921	19.9	174 513	4.5
Pharmacy	26.5	8.4	267	5.8	169 767	4.4
Geology	18.7	5.9	139	3.0	59 018	1.5
Geophysics	11.0	3.5	79	1.7	65 138	1.7
Astrophysics	10.2	3.2	84	1.8	62 284	1.6
Biochemistry	9.0	2.8	197	4.3	248 298	6.4
Geography	7.5	2.4	179	3.9	19 884	0.5
Science Faculty	316.8	100.0	4620	100.0	3 858 968	100.0

Information regarding scientific staff and students (undergraduate, graduate and doctoral students combined) has been taken from the Science Faculty's 1997 annual report; publication data are taken from Tables 15 and 16.

It is obvious from Table 16 that major science subjects like Chemistry and Physics also figure prominently with regard to research volume, but none of them reach the same level as Biology (Table 15). Table 17 compares the calculated research/knowledge volumes with the actual sizes of the corresponding Science Faculty departments, as expressed by the number of tenured scientific staff members. Data for student numbers at all levels (undergraduate, graduate and doctoral students) are also included, since these can be expected to be a major premise for the dimensioning of departments.

Table 17 reveals striking differences between the departments: Mathematics and Informatics have low knowledge volumes, but very high student numbers. Biology, Chemistry and Physics have large knowledge volumes, and also reasonably high student numbers. These five departments have the largest scientific staffs within the faculty, indicating that both parameters have played a role in determining department size. The smaller departments display low scores on both knowledge volumes and student numbers.

Department size (scientific staff) correlated quite well with both the number of student enrolled at the department (correlation coefficient = 0.69) and with the knowledge volume of the departmental subject matter (correlation coefficient = 0.79). A combined "subject size" parameter, i.e., the mean of the two percentage values for student numbers and knowledge volumes (relative to faculty total and science subject total, respectively), correlated even better with department size (correlation coefficient = 0.93), as shown in Fig. 11. However, a major discrepancy was still apparent for the Department of Biology, which would seem to be considerably understaffed relative to the other departments within the Science Faculty.

Fig. 11. Correlation between the sizes (scientific staffs) of Science Faculty departments and the corresponding "subject sizes", as calculated on the basis of student numbers and knowledge volumes (international research volumes). *The values for each department's fraction of the faculty total were taken from table 17; "subject size" is the mean of the percentages for student number and international publications. AS, astrophysics; BC, biochemistry; BI, biology; CH, chemistry, GG, geography; GL, geology; GP, geophysics; IN, informatics; MA, mathematics, PC, pharmacy; PH, physics.*

Chapter VIII.Conclusions

Overall evaluation of the Department of Biology sections

Table 18 sums up the results of the scientometric analysis of the Department of Biology at the level of sections and subsections (including the *ad hoc*-defined "supersection" Aquatic Biology). The committee chose to present the indicator list to the department board as a tool for further decision-making, leaving the academic priorities, in the form of indicator weighting, to the board.

Table 18. *Scientometric performance indicators for sections and subsections of the Department of Biology in relation to the size of the tenured scientific staff*

Section/subsection	Publ.	Citations uncorr.	Citations field-corr.	Grant income	Teaching hours	Grad. stud.	Undergr. points	Knowl. volume	Tenured staff (%)
Cell Biology	**18.8**	**30.2**	**18.2**	**24.2**	**23.0**	**21.7**	**25.4**	**42.0**	**18**
Cell biology	12.0	19.4	10.4	16.8	10.2	10.3			8
Biotechnology	3.6	3.7	2.8	1.6	2.1	6.2			2
Microbiology	2.1	6.4	4.3	4.0	7.1	2.1			6
Toxicology	1.1	0.7	0.7	1.8	3.6	3.1			2
Genetics	**8.9**	**6.4**	**3.5**	**7.5**	**6.0**	**6.7**	**9.5**	**6.9**	**6**
Physiology	**14.4**	**12.9**	**9.8**	**6.2**	**12.6**	**11.3**	**12.6**	**25.1**	**12**
Botany	**16.0**	**6.3**	**7.7**	**8.7**	**17.8**	**15.9**	**18.7**	**9.9**	**18**
Botany	14.6	1.3	1.8	8.7	11.9	11.8	12.7		12
Plant physiology	1.4	5.0	5.9	0.2	5.9	4.1	6.0		6
Aquatic Biology	**18.4**	**18.6**	**24.6**	**21.7**	**25.0**	**19.0**	**14.1**	**7.1**	**30**
Limnology	7.3	11.3	14.8	7.0	5.5	3.6	5.1		10
Freshw. biol.	6.9	11.3	14.8	7.0	3.4	1.5			6
Limnology	0.4	0.0	0.0	0.0	2.1	2.1			4
Marine botany	4.7	3.0	3.9	4.9	6.1	3.6	2.2		8
Marine chem. zool.	6.4	4.3	5.9	9.8	13.4	11.8	6.8		12
Marine zoology	5.4	4.1	5.6	7.9	10.7	9.7			8
Marine chemistry	1.0	0.2	0.3	1.9	2.7	2.1			4
Zoology	**23.6**	**25.7**	**36.2**	**31.6**	**15.6**	**26.2**	**19.7**	**8.9**	**16**

All indicators, including staff size, are expressed as % of the department total values.

A few obvious conclusions were, however suggested to the department board by the committee:

Firstly, the comparison between the various departments within the Science Faculty clearly indicated that the Department of Biology was considerably underfunded.
Secondly, all performance indicators suggested that within the Department of Biology, the sections of cell biology and zoology should be strengthened, whereas reductions were indicated for aquatic biology in general, and for plant physiology.

Thirdly, the fact that many staff members showed little or no scientific productivity should be a reason for concern, particularly since low scientific activity was not compensated by increased teaching.

Fourthly, the internal resources for research at the Department of Biology are so limited that an evaluation-based resource allocation has little meaning at present. Resources freed by staff reductions in the least productive sections might therefore be used to increase the research budget.

Implementation of the evaluation

The present scientometric evaluation was used as a basic source document by a second committee appointed by the Department of Biology board. This "staff plan committee" (stillingsplankomité) chose to discard the grant income indicator, whereas the remaining seven performance indicators were weighted as shown in Table 19, first column.

The staff plan committee felt that the department staff should be distributed primarily on the basis of teaching requirements, and gave 75% weight to the teaching performance indicators. Considering that University regulations require scientific staff to spend 40% of their time on research, the mere 20% weighting of scientific performance would seem surprisingly low. The committee recognized the value of scientific activity and the importance of maintaining and supporting a good scientific environment, but it was argued that present-day scientific performance is strongly person-dependent, and thus has limited prognostic value for future sectional research. The danger of permanently weakening scientifically important sections through staff reductions based on temporary poor scientific performance was also pointed out. However, although the distribution of weights between teaching and science is an central issue in principle, it turned out that the teaching and science indicators were so well correlated that an increased weighting of science (up to 60%) made little difference in practice.

Among the scientific performance indicators, productivity and citedness were given equal weight. The committee opted for a moderate field correction of citations by giving corrected and uncorrected citations similar weight, apparently a political compromise. It was pointed out that the lack of direct quality indicators is a major limitation to the evaluation of scientific performance, which further argues for placing only a moderate weight on scientific indicators at the present time.

***Table 19.** Weighting of performance indicators by the staff plan committee at the Department of Biology*

		Sectional score, % of department total					
Performance indicator	Weight (%)	Cell Biology	Genetics	Physiology	Botany	Aquatic	Zoology
Teaching performance	**75**	**17.5**	**5.8**	**8.9**	**12.9**	**13.1**	**17.2**
Teaching hours	5	1.2	0.3	0.6	0.9	1.3	0.8
Undergraduate points	30	7.6	2.9	3.8	5.6	4.2	5.9
Graduate students	40	8.7	2.7	4.5	6.4	7.6	10.5
Scientific performance	**20**	**4.3**	**1.4**	**2.6**	**2.3**	**4.0**	**5.5**
Publications	10	1.9	0.9	1.4	1.6	1.8	2.4
Citations, uncorrected	5	1.5	0.3	0.7	0.3	0.9	1.3
Citations, field-corrected	5	0.9	0.2	0.5	0.4	1.2	1.8
Knowledge volume	**5**	**2.1**	**0.4**	**1.3**	**0.5**	**0.4**	**0.5**
Total	**100**	**23.9**	**7.6**	**12.8**	**15.7**	**17.4**	**23.1**
		Staff size, % of department total (no. in parentheses)					
Present staff	(50)	18.0 (9)	6.0 (3)	12.0 (6)	18.0 (9)	30.0(15)	16.0 (8)
Future staff (committee)	(37)	24.3 (9)	8.1 (3)	13.5 (5)	16.2 (6)	18.9 (7)	24.3 (9)
Future staff (dept. board)	**(41)**	**17.1 (7)**	**7.3 (3)**	**12.2 (5)**	**22.0 (9)**	**22.0 (9)**	**19.5 (8)**

The "knowledge volume" indicator became a subject of intense debate in the staff plan committee, because its emphasis on basic biology deviated strongly from the present structure of the Department of Biology. In particular it was argued that marine biology, which is a small subject globally, should retain its prominent position within the department due to its high national priority. Others (the majority) felt that such priorities would be better served by external granting, and that the University's policy should be based on internal criteria. The committee finally decided to accord a modest 5% weight to the knowledge volume indicator.

The result of the indicator weighting is shown in Table 19. On the basis of budgetary and organizational considerations, the staff plan committee suggested a future 26% cut in the overall scientific staff, from 50 to 37 persons. The resources thereby freed were intended to be used mainly to increase the number of departmentally financed research fellows, from 6 to 19.

The suggestions made by the staff plan committee were submitted to the Department of Biology board, who made the final decisions. The board opted for a more moderate (18%) reduction in overall staff size, from 50 to 41. The major cut was in aquatic biology (from 15 to 9), as recommended by both of the previous committees. However, the board also suggested a surprising reduction in the Cell Biology staff from 9 to 7, whereas Botany remained unchanged, contrary to the committee recommendations. In addition to the staff reductions, the board suggested a number of organizational changes (various fissions, fusions and transfers of staff between the sections) that will not be dealt with here. It should be noted that all staff reductions were expected to occur by regular retirement, which effectively curtailed the possibilities of achieving changes within a reasonable time. This long-term implementation perspective (5-10 years) relative to the rapid changes in the composition of boards and committees means, furthermore, that on staff issues, decisions and their implementation will be performed by different people. Since the present experience suggests that science policy in the end is based more on power and influence than on objective analytical data (the latter probably being used only to the extent that they are in consonance with former), the eventual outcome of the scientometric evaluation of the Department of Biology is difficult to predict.

Scientometric problems and solutions

A major conclusion of the present study is that a scientometric evaluation of a heterogeneous field such as biology is technically feasible. The criteria, practical solutions and indicators used were well accepted, and should form a useful basis for decision-making, at least in principle.

The indicators used were determined by the mandate given to the evaluation committee (teaching and science performance); in other evaluation contexts, different indicators might have been chosen. Only objectively quantifiable indicators were used. This double requirement limits the possibilities, a major shortcoming being that no quantitative indicator of general scientific quality is available. On the other hand, not all quantitative indicators are useful: for example, the committee regarded the widely employed journal impact factors to be highly misleading as indicators of scientific performance. Similarly, the staff plan committee discarded the ability of a project to attract external grants as a useful scientific performance indicator. Thus three teaching indicators (teaching hours, undergraduate study points and graduate students) and two science indicators (productivity and citedness, the latter presented both as uncorrected and field-corrected data), plus a novel knowledge volume indicator remained.

The presentation of unweighted indicators, leaving the weighting to the user of the evaluation data, is probably the most appropriate strategy, and was well accepted. Indicator weighting is a highly context-dependent and policy-dependent matter, and clearly not the responsibility of the scientometrician. Although the evaluation committee was instructed to draw practical conlusions, it tried to avoid making explicit policy statements. In the future, to avoid any conflict of interest, it would probably be prudent to draw a clear line between evaluation and policy, and even to have the scientometric analysis performed by an external committee or agency.

The present study was a challenging exercise in practical scientometry, which had to face and solve a number of problems. Although these have been described in detail throughout this report, a brief general discussion may be useful:

(1) Unit of evaluation. In an evaluation using multiple indicators it is obviously important to chose a unit of evaluation to which all the indicators are applicable. Our choice of the individual tenured scientific staff member as the basic unit was dictated by this need. Any higher organizational unit could then easily be evaluated by defining it as its aggregate scientific staff. This strategy could be applied consistently in the case of the scientific indicators, but for the teaching indicators - which included student enrolment and output - the contribution by supplementary teachers (professor-IIs) had to be taken into account. The use of a person-based rather than an address-based institutional definition is relatively unique in a scientometric context, but carries obvious methodological advantages, at least for small-scale analyses.

(2) Choice and weighting of published material for evaluation. Scientists produce a wide array of publications with varying degrees of scientific relevance. The choice, classification and weighting of such publications will be highly dependent on the context and purpose of an evaluation. In the present case, where the emphasis was on scientific performance, only international publications were considered. Full weight was given to all original articles (in journals or in electronic media), reviews, proceedings, book chapters, edited books, and patents, whereas journal items like editorials, letters, book reviews etc. were given half weight. Books were given double weight; abstracts were ignored. With a wider evaluation mandate, it might have been relevant to construct a "knowledge transmission" indicator incorporating national-language publications, popularization, reports, textbooks, teaching material etc.; this should probably also be done in evaluations reported at the individual level. Furthermore, evaluations of individuals should consider a rough quantification of publication contents, which have been shown to exhibit a hundred-fold variation (Seglen, 1996).

(3) Sources of publication data. Institutional publication lists were used to some extent in the present study, but they were found to be relatively inconsistent and unreliable. Authors' publication lists would probably be better, but are not error-fre. On-screen searching in the ISI database gave excellent recovery: only the botany section published a significant fraction of their international articles in journals outside the database, but should have no difficulty in altering their journal set to fit the database portifolio. Although proceedings and books are not included in the database, these publication types contributed relatively little to the Department of Biology output, suggesting that future evaluations of scientific performance may be entirely database-based. The high degree of internationalization and journal-based publication was somewhat surprising, but is probably typical of contemporary academic biological research.

The general attitude among the department staff suggested that the evaluation window should be as wide as possible. In the present case, all staff members had been in a position to publish throughout the seven-year window available (1992-98), but shorter tenure would pose no problem as long as productivity is expressed on a per year basis. Since database entries are often delayed, the database search should be performed no earlier than at least half a year after the last calendar year searched

(4) Conservative fractional credit attribution. In contrast to the unfractionated attribution of credit ("normal counting") used in most scientometric studies, the present evaluation has introduced a "conservative" fractional credit attribution. By this principle, a tenured scientific staff member (defined as a project leader) receives full credit for all papers coauthored with just his own coworkers, but half credit for papers coauthored with other tenured members of the scientific staff. In the case of papers coauthored with colleagues outside the Department of Biology (i.e., with two or several institutional addresses), the staff member receives full credit if only one coauthor is extradepartmental; otherwise half credit (but never less, hence the term "conservative"). This mode of crediting allows a slight overestimation of overall productivity as a compensation for not analyzing cooperative relationships in greater detail. The conservative fractional crediting received no objections from the Department of Biology staff, suggesting that an acceptable balance had been struck.

(5) Field correction of citation data. The need to make research field corrections of citation data is generally recognized by scientometricians, but such corrections are often performed in a manner that introduces novel bias, e.g., by using authors' journal choice as a reference base. The present study shows that it is possible to work out author-independent, specific field correction factors for each evaluated unit; in this case, for each departmental section. The strategy used was to calculate the average citedness of the various biological categories classified in the ISI database, each being defined by a given set of journals for which impact factors (average journal citedness) are available. The relative weight of each of these subfields within each Department of Biology section was then estimated, and a final sectional field correction factor calculated.

In principle, it should be possible to calculate specific field correction factors at any level, including the level of individual scientists. The procedure can probably be improved by using subject categories other than those defined by ISI, and by more considerate definitions of the journal portifolios that characterize each category. In particular, scientists within a field do not publish only in the field's specialist journals, but also in general, basic journals that may have been placed elsewhere in the ISI classification system (Lewison, 1996; Aksnes *et al.*, 2000). Since the latter tend to have higher impact factors than the specialist journals, the field correction factors used in the present study may tend to overcorrect citations in the basic biology sections (cell biology; genetics) and to undercorrect them in the specialist sections (botany, zoology and aquatic biology). By giving field-corrected and uncorrected citations equal weight, this possible bias was amply corrected for in the final evaluation of the Department of Biology. Until better defined citation categories have been developed, the resulting "conservative field correction" probably strikes an acceptable balance.

(6) A major shortcoming: the lack of quality indicators. Somewhat surprisingly, the staff plan committee accorded equal weight to productivity and citedness, perhaps reflecting a feeling that quantity and quality should count equally. When used cumulatively, as in the present study, citedness is in fact partially a productivity indicator, which would compensate somewhat for the limitation that it is only a partial quality indicator (measuring intrascientific utility, primarily). Although the final balance between productivity and intrascientific utility may thus be reasonable, the lack of suitable indicators of intrinsic scientific quality is an obvious shortcoming of scientometric evaluation. Intrinsic quality can probably only be assessed subjectively, which means that some effort should be made to develop subjective indicators which are sufficiently reliable to be used quantitatively. By using standardized quality criteria and score scales, it should be possible for independent specialists to achieve the reproducibility (i.e., objectivity) in quality assessment required for a quantitative indicator.

(7) Knowledge volume. The need for objective guidelines as to how large Biology Department sections ought to be, precipitated our formulation of the concept of "knowledge volume". The idea is that each subject taught at the University level represents a volume of knowledge, which may vary strongly from subject to subject. The larger the knowledge volume, the larger the staff needed to cover it adequately.

It was reasoned that the volume of contemporary research within a field should represent an integrated reflection of both the existing size of the field (on which the research is based), and its future size (which will be determined by present-day research). The number of papers registered in the ISI database under each of its 254 subject categories was, therefore, counted, and fractionally distributed into new categories corresponding to the subject composition of the department sections. A similar excercise was performed at the level of the various Science Faculty departments, in an attempt to estimate the relative knowledge volumes supposed to be covered by each department. Since some of the biomedical subject categories obviously represented knowledge to be shared with the Medical Faculty, a "medical index" was constructed, calibrated against a large study of Norwegian microbiology (Aksnes *et al.*, 2000; Seglen & Aksnes, 2000), and applied to each subject category, thus achieving a reasonable distribution of the biomedical research volume between medicine and biology.

The use of research volume as a proxy measure of knowledge volume was not contested, and the fact that biology came out as a 50% larger field than chemistry and physics was favourably received by the biologists. The relative knowledge volumes within the Department of Biology became more problematic, in particular the fact that the knowledge volume covered by the cell biology section, with 9 staff members, was six times larger than the knowledge volume covered by aquatic biology, with 15 staff members. Eventually, the staff plan committee decided to accord a conservative 5% weight to the knowledge volume indicator. It is clear that neither the Department nor the University has been able to keep pace with the explosive research and knowledge development within cell biology, but an overnight adjustment is obviously both practically and politically difficult. The knowledge volume concept is, however, an interesting scientometric innovation which probably ought to find its way to the science policy agenda.

References

Aksnes, D.W., Olsen, T.B. & Seglen, P.O. (2000). Validation of bibliometric indicators in the field of microbiology: A Norwegian case study. *Scientometrics* **49**, 7-22.

Andersen, H. (1996a). ACTA Sociologica på den internationale arena - hvad kan SSCI fortælle? *Dansk Sociologi* **2**, 72-78.

Andersen, H. (1996b). Tidsskriftpublicering og forskningsevaluering. *Biblioteksarbejde* **17**, 19-26.

Artus, H.M. (1996). Science indicators derived from databases. *Scientometrics* **37**, 297-311.

Bourke, P. & Butler, L. (1996). Standard issues in a national bibliometric database: the Australian case. *Scientometrics* **35**, 199-207.

Braun, T., Glänzel, W. & Grupp, H. (1996). The scientometric weight of 50 nations in 27 science areas, 1989-1993. Part II. Life sciences. *Scientometrics* **34**, 207-237.

Brooks, T.A. (1985). Private acts and public objects: An investigation of citer motivations. *J. Am. Soc. Information Sci.* **36**, 223-229.

Chase, J.M. (1970) Normative criteria for scientific publication. *The American Sociologist* **5**, 262-265.

Chubin, D.E. & Moitra, S.D. (1975). Content analysis of references: Adjunct or alternative to citation counting? *Social Studies of Science* **5**, 423-441.

Cole, J.R. & Cole, S. (1973). *Social stratification in science*. University of Chicago Press, Chicago & London. 283 pp.

Evans, J.T., Nadjari, H.I. & Burchell, S.A. (1990). Quotational and reference accuracy in surgical journals. A continuing peer review problem. *J. Am. Med. Ass.* **263**, 1353-1354.

Fleischmann, M. & Pons, S. (1989) Electrochemically induced nuclear-fusion of deuterium. *J. Electroanal. Chem.* **261**, 301-308.

Folly, G., Hajtman, B., Nagy, J.I. & Ruff, I. (1981). Some methodological problems in ranking scientists by citation analysis. *Scientometrics* **3**, 135-147.

Garfield, E. (1972). Citation analysis as a tool in journal evaluation. *Science* **178**, 471-479.

Garfield, E. (1979). Is citation analysis a legitimate evaluation tool? *Scientometrics* **1**, 359-375.

Gilbert, G.N. (1977). Referencing as persuasion. *Social Studies of Science* **7**, 113-122.

Gomperts, M.C. (1968). The law of constant citation for scientific literature. *J. Documentation* **24**, 113-117.

Gulbrandsen, M. & Langfeldt, L. (1997). *Hva er forskningskvalitet? En intervjustudie blant norske forskere*. NIFU-Rapport (Oslo) 9/97, 146 pp.

Hamilton, D.P. (1991). Research papers: Who's uncited now. *Science* **251**, 25.

Hargens, L.L. & Felmlee, D.H. (1984). Structural determinants of stratification in science. *Am. Sociol. Rev.* **49**, 685-697.

Hemlin, S. (1993) Scientific quality in the eyes of the scientist: A questionnaire study. *Scientometrics* **27**, 3-18.

Korevaar, J.C. & Moed, H.F. (1996). Validation of bibliometric indicators in the field of mathematics. *Scientometrics* **37**, 117-130.

Lange, L. (1985). Effects of disciplines and countries on citation habits. An analysis of empirical papers in behavioural sciences. *Scientometrics* **8**, 205-215.

Lewison, G. (1996). The definition of biomedical research subfields with title keywords and application to the analysis of research outputs. *Research Evaluation* **6**, 25-36.

Lindsey, D. (1980). Production and citation measures in the sociology of science: the problem of multiple authorship. *Social Studies of Science* **10**, 145-162.

Line, M.B. & Sandison, A. (1973). Progress in documentation: 'Obsolescence' and changes in the use of literature with time. *J. Documentation* **29**, 283-350.

MacRoberts, M.H. & MacRoberts, B.R. (1996). Problems of citation analysis. *Scientometrics* **36**, 435-444.

Magri, M.-H. & Solari, A. (1996). The SCI Journal Citation Reports: A potential tool for studying journals? I. Description of the JCR journal population based on the number of citations received, number of source items, impact factor, immediacy index and cited half-life. *Scientometrics* **35**, 93-117.

Margolis, J. (1967). Citation indexing and evaluation of scientific papers. *Science* **155**, 1213-1219.

May, R.M. (1997). The scientific wealth of nations. *Science* **275**, 793-796.

McAllister, P.R. & Wagner, D.A. (1981). Relationship between R&D expenditures and publication output for U.S. colleges and universities. *Res. Higher Educat.* **15**, 3-30.

Merton, R.K. (1968). The Matthew effect in science. *Science* **159**, 56-63.

Moed, H.F., Burger, W.J.M., Frankfort, J.G. & Van Raan, A.F.J. (1985). The application of bibliometric indicators : important field-and time-dependent factors to be considered. *Scientometrics* **8**, 177-203.

Moed, H.F., Burger, W.J.M., Frankfort, J.G. & Van Raan, A.F.J. (1987). *On the measurement of research performance: The use of bibliometric indicators*. Science Studies Unit, LISBON-Institute, University of Leiden, Leiden, 199 pp.

Moed, H.F. (1996). Differences in the construction of SCI based bibliometric indicators among various producers: a first overview. *Scientometrics* **35**, 177-191.

Moed, H.F., Van Leeuwen, T.N. & Reedijk, J. (1996). A critical analysis of the journal impact factors of *Angewandte Chemie* and *The Journal of the American Chemical Society*. Inaccuracies in published impact factors based on overall citations only. *Scientometrics* **37**, 105-116.

Moed, H.F. & Van Leeuwen, T.N. (1996). Impact factors can mislead. *Nature* **381**, 186.

Murphy, P.S. (1995). From bibliometrics to research productivity: Benchmarks and self-assessment. *Proc. 5th Conf. Internat. Soc. Scientometrics Informetrics*, River Forest, USA, 349-355.

Møller, A.P. (1990). National citations. *Nature* **348**, 480.

Narin, F., Pinski, G. & Gee, H.H. (1976). Structure of the biomedical literature. *J. Am. Soc. Inform. Sci.* **27**, 25-45.

Narin, F. & Hamilton, K.S. (1996). Bibliometric performance measures. *Scientometrics* **36**, 293-310.

Pravdic, N. & Oluic-Vukovic, V. (1986). Dual approach to multiple authorship in the study of collaboration/scientific output relationship. *Scientometrics* **10**, 259-280.

Price, D.J. de S. (1981). The analysis of square matrices of scientometric transactions. *Scientometrics* **3**, 55-63.

Schubert, A. & Braun, T. (1986). Relative indicators and relational charts for comparative assessment of publication output and citation impact. *Scientometrics* **9**, 281-291.

Schwartz, S. & Lopez Hellin, J. (1996). Measuring the impact of scientific publications. The case of the biomedical sciences. *Scientometrics* **35**, 119-132.

Seglen, P.O. (1989a). Kan siteringsanalyse og andre bibliometriske metoder brukes til evaluering av forskningskvalitet? *NOP-Nytt (Helsingfors)* **15**, 2-20.

Seglen, P.O. (1989b). Bruk av siteringsanalyse og andre bibliometriske metoder i evaluering av forskningsaktivitet. *Tidsskr. Nor. Laegeforen.* **31**, 3229-3234.

Seglen, P.O. (1992a). Evaluation of scientists by journal impact. In *Representations of Science and Technology* (Weingart, P., Sehringer, R. & Winterhager, M., eds.), pp. 240-252. DSWO Press, Leiden.

Seglen, P.O. (1992b). The skewness of science. *J. Am. Soc. Information Sci.* **43**, 628-638.

Seglen, P.O. (1994). Causal relationship between article citedness and journal impact. *J. Am. Soc. Information Sci.* **45**, 1-11.

Seglen, P.O. (1996). Quantification of scientific article contents. *Scientometrics* **35**, 355-366.

Seglen, P.O. (1997a). Citations and journal impact factors: questionable indicators of research quality. *Allergy* **52**, 1-7.

Seglen, P.O. (1997b). Why the impact factor of journals should not be used for research evaluation. *Br. Med. J.* **314**, 498-502.

Seglen, P.O. (1998). Citation rates and journal impact factors are not suitable for evaluation of research. *Acta Orthopaed. Scand.* **69**, 224-229.

Seglen, P.O. & Aksnes, D.W. (2000). Scientific productivity and group size: A bibliometric analysis of Norwegian microbiological research. *Scientometrics* **49**, 125-143.

Sivertsen, G. (1991). *Norsk forskning på den internasjonale arena: En sammenlikning av 18 OECD-lands artikler og siteringer i* Science Citation Index *1973-86*. Inst. Stud. Res. Higher Educat., Oslo, 90 pp.

Sivertsen, G. (1993) *Nordisk samfunnsforskning i internasjonale tidsskrifter: Artikler 1981-90 i Social Science Citation Index*. Nord 5, Nordisk Ministerråd, København.

Sivertsen, G. & Aksnes, D.W. (2000) Norge som forskningsnasjon - II. *Forskningspolitikk* **23**, 12-13.

Vinkler, P. (1996a). Relationships between the rate of scientific development and citations. The chance for a citedness model. *Scientometrics* **35**, 375-386.

Vinkler, P. (1996b). Model for quantitative selection of relative scientometric impact indicators. *Scientometrics* **36**, 223-236.

Vinkler, P. (1996c). Some practical aspects of the standardization of scientometric indicators. *Scientometrics* **35**, 237-245.

Wade, N. (1975). Citation analysis: A new tool for science administrators. *Science* **188**, 429-433.

Zetterström, R. (1999). Impact factor and the future of Acta P diatrica and other European medical journals. *Acta Paediatr.* **88**, 793-796.