Scientometrics, Vol. 27. No. 2 (1993) 157-178

RESEARCH PERFORMANCE INDICATORS FOR UNIVERSITY DEPARTMENTS: A STUDY OF AN AGRICULTURAL UNIVERSITY

A. J. NEDERHOF, R. F. MEIJER, H. F. MOED, A. F. J. VAN RAAN

Centre for Science and Technology Studies (CWTS), University of Leiden, Wassenaarseweg 52, P.O. Box 9555, 2300 RB Leiden (The Netherlands)

(Received October 12, 1992)

The present bibliometric study extends previous work by focusing on the research performance of departments in the natural and life sciences, the social and behavioral sciences, and the humanities. The present study covers all 70 departments from one agricultural university, and several veterinary departments of a second university. The impact analysis was extended by including other types of documents than journal articles. For about a third of the departments, publications not covered in citation indexes accounted for at least 30% of the citations to their total oeuvre. To deal with different citation and publication habits in the various fields, both short-term and medium-term impact assessments were made. The commonly used three year window is not universally applicable, as our results show. The inclusion of self-citations forms an important source of error in the ratio of actual/expected impact. To cope with this, the trend and level of self-citations was compared at university level with that in a matched sample of publications. Moreover, at a departmental level, self-citation rates were used to detect departments with divergent levels of self-citation. The expected impact of journals accounted for only 18% of the variance in actual impact. Comparison of bibliometric indicators with two peer evaluations showed that the bibliometric impact analyses provided important additional information.

1. Introduction

In several evaluation studies of research performance, peer review and bibliometric methods have been used in combination (e.g., *Martin & Irvine*, 1983; *Moed* et al., 1985a, 1985b; *Nederhof & Van Raan*, 1989). Both peer review and bibliometric methods have been criticised, but each has its own assets and liabilities (e.g., *Harnad*, 1982; *Nederhof*, 1988). Previous evaluations of research performance have mainly been focused upon big science facilities (e.g., *Martin & Irvine*, 1983), on the (basic) natural and life sciences (e.g., *Moed* et al., 1985a) or on particular social sciences or humanities (e.g., *Davis & Papanek*, 1984; *Nederhof & Noyons*, 1992; *Nederhof, Zwaan, De Bruin & Dekker*, 1989; *Spangenberg, Buijink & Alfenaar*, 1990). The present study extends previous work by focusing on the research performance of

Elsevier, Amsterdam – Oxford – New York – Tokyo
Akadémiai Kiadó

research units in agriculture and veterinary science, by covering both basic and applied research, and by including disciplines from the natural and life sciences, but also from the social sciences and the humanities. Moreover, the present study covers all departments from one university (the Wageningen Agricultural University (WAU)), and not only part of a university. Given this broadened emphasis, the set of bibliometric indicators usually applied in studies of academic research performance needed to be extended and improved, and tailored to a large variety of publication and citation habits.

In the application of the bibliometric approach, one has to be aware of its limitations: an exclusive focus on written output, restriction to (Social) Science Citation Index journals (at least as a citation source), the necessity to confine impact measurement (citation counting) to a particular 'citation window', and the presence of various sources of error and bias in citations (cf. *Martin & Irvine*, 1983; *Nederhof & Van Raan*, in press). The present study also focuses on written output (other forms of output have been covered in *Meijer, Nederhof & Van Raan*, 1991), but attempts to break through the confinements of one particular citation window.

It is assumed that science develops at the international research front (*Price*, 1963). Here, scientific publications (in journals, books, etc.) form an important medium of scientific communication. Research results are reported in publications, in which researchers refer to earlier work, a.o. because it is linked to their own research or because they base themselves on it. Therefore, the number of times a publication is referred to, gives a partial indication of the 'impact' of a publication (*Martin & Irvine*, 1983; *Moed* et al., 1985).

Publications in journals constitute an primary communication medium for scientific researchers. However, publications in other media can also be of great relevance (*Nederhof*, 1989; *Meijer, Nederhof & Van Raan*, 1991). An important new aspect of the present investigation concerns an extension of citation analysis to other types of documents than journal articles.

In this article, we first examined the coverage of WAU output in ISI journals. Next, we looked at some of the main citation and publication characteristics of WAU publications in order to fine-tune assessment. Then, indicators are applied at both WAU and department level. Finally, bibliometric results are compared with results of peer evaluations.

2. Method and data collection

The research project which this article resumes is aimed at the construction of bibliometric impact indicators for agricultural and veterinary research conducted in the Netherlands, at the Wageningen Agricultural University (WAU) and the Faculty of Veterinary Science of the University of Utrecht. Agricultural and veterinary research in the Netherlands covers both basic to applied research.

Our data collection is based on digitalized publication-lists provided by and carefully screened by the WAU concerning all of its departments (N=73) for the publication period 1976–1987. The departments can be divided in natural science (N=16) and bioscience (N=39) departments, and departments of the social and behavioural sciences and humanities (in short: designated as social sciences, as only few humanities departments were present) (N=18), reflecting the large scope of agricultural research. In analyses, these clusters of departments are treated as aggregated scientific areas. Publications included both those in ISI-journals (N=3966), and all scientific publications not covered by ISI, namely non-ISI journal articles (N=1278), PhD theses (N=395), scientific books (N=285), chapters in scientific books (N=1509), contributions to proceedings (except abstracts; N=1524), and external research reports (N=374). Citations from 1976 to 1989 were collected in an automatized way in SCI, SSCI and A & HCI, by linking the first six letters of the name of the first author, the first initial, publication year, volume number, and first page of the publication (*Moed*, 1989). Citations to non-journal publications (e.g. books) were manually linked to these publications, as page number and volume number are mostly missing. Based on publication lists provided by the Faculty of Veterinary Science of the University of Utrecht, similar data were collected for the period 1980–87.

We have developed output-indicators based on number of publications, and impact indicators based on number of citations. Output indicators include the number of publications in journals covered by ISI ('ISI-publications'), and the mean number of these ISI-publications normalized by input per 'full time equivalent' devoted to research (fte). The fte indicates the percentage of time a person can formally spend on research, according to his or her appointment. Furthermore, we monitor the number of publications in other media (publications in journals not covered by ISI, books, chapters and contributions to books, contributions to proceedings, research reports, and PhD theses).

Impact indicators include the total number of citations to ISI-publications, the total number of citations to other publications ('non-ISI publications'), as well as the mean number of citations per publication, determined by type of publication medium. Impact is measured with a citation window of at maximum five years starting from the year of publication. In this study, short-term impact is determined by 'citation windows' from the publication year up to two years after the publication year, and medium-term impact by citation windows from the third year to the fourth year after publication (see Section 3). The mean number of citations is described as 'impact', as discerned from the absolute impact. These impact analyses excluded self-citations.

The actual impact is also compared with an expected value. This value is based on the mean citation score of the journals in which the departments publish their articles (cf. *Moed* et al., 1985). The computation of Journal Citation Scores (JCS) may serve as a baseline for comparison of performance of research units. This procedure entails comparing the number of citations a set of articles receives in the second or third year after publication, (frequently the 'top' year of citation) to that obtained by the average article in the same journals. For instance, the 225 articles appearing in Journal A during 1982 received 502 citations in 1983 and 1984, so the JCS is 2.23. Now, assume that a set of five articles published in 1982 consists of two articles published in Journal A and three articles in Journal B (with a JCS of 1.03). Then, the estimated JCS value of that set would be $2 \times 2.23 + 3 \times 1.03 = 7.56$, and the JCS for an article, the 'expected impact', would be $7.56/5 = 1.512$. If the five articles combined are cited 8 times in 1983 and 1984, the 'actual impact' would be 1.6, or rather close to the expected impact. However, if the five articles were cited 100 times, actual impact would have been 20, and it clearly would have exceeded the expected impact.

For the JCS analyses only, the actual impact is calculated including self-citations, because the expected values can only be calculated including self-citations. This JCS procedure is limited to ISI journals, as expected values can be derived only from the Journal Citation Reports. For reasons discussed below, we also computed expected and actual 'fourth-fifth year' impact scores, and second-fifth year impact scores. To gain insight in the relation between the impact level of journals and the impact of articles, we calculated Spearman's rank order correlation coefficients between expected impact and actual impact.

3. Results

3.1. ISI-coverage

To gauge the appropriateness of the ISI citation indexes for assessing the research performance of the Wageningen Agricultural University (WAU), the coverage of WAU publications by ISI was examined. The WAU publishes a minority of its scientific publications in ISI-journals (43%). ISI-coverage varies among the three aggregated disciplinary areas. Bioscience departments publish almost as much ISI-publications (48%) as non-ISI publications, while natural science departments publish more often in ISI- (55%) than in non-ISI media. In both cases, the percentage remains stable during the whole period. However, more than 90% of the social science publications appears in non-ISI media.

ISI covers foremostly journal literature, and has an Anglo-Saxon language bias in many fields. When we consider only scientific journal articles, 76% of WAU articles is published in ISI-journals. Taking into account only journal publications written in another language than Dutch, even 84% is covered by ISI, while then the ISI-coverage for social science departments varies between 45% (1976–79) and 70% (1984–87). These figures indicate that ISI adequately covers publications of WAU directed at a scientific public, particularly in the natural and biosciences, and particularly when the journal literature is concerned.

3.2. Selection of citation windows

In bibliometric analyses, citations are often collected during the first three years after publication. Previous research has indicated that in many fields, citations peak in the third year after publication, but there are many exceptions (e.g., *Moed* et al., 1985; *Nederhof & Noyons*, 1992; *Nederhof & Van Raan*, in press). To determine the optimal citation window, we have examined the speed of diffusion of knowledge – as measured by changes in amount of citations over time, and the number of years it takes publications to gain a significant peak of citations.

ISI-publications (N=1139) from 1976–80 show a fast increase of citations until the third year of their lifetime. When the citation scores are ranked, a non-parametric statistical test, Duncan's range test, shows that the third year differs significantly ($p < 0.05$) from the first and second year, but not significantly from the other years ($p > 0.05$).

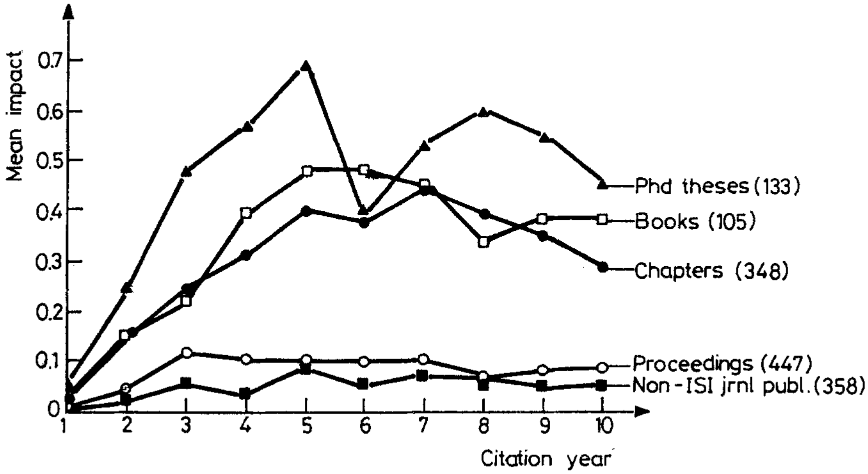

Fig. 1. Impact per non-ISI document type from the first after the publication year to the tenth year of their lifetime (publication period 1976–1980).
Note: The number of publications is indicated between brackets

The aggregated non-ISI publications (N = 1439) from 1976–80 level off only in the fifth year. Diffusion speed differs among non-ISI publications (see Fig. 1). Publications in other media than journals attain a citation plateau in the fourth or fifth year of their lifetime, but contributions to proceedings attain a peak in the third year just like journal articles.

To be able to adjust for period effects, we also compared ISI-publications from 1976–1980 with those of 1981–1985, using 5 year citation windows. It was observed that a citation peak was reached only in the fourth year in 1976–1980, while in 1981–1985, the peak was reached already in the third year, and differed significantly from the fourth year ($p < 0.05$). However, at a departmental level, in both periods only a minority actually peaked in the third year (see Table 1), although in most of these cases the third, fourth, and/or fifth year did not differ significantly. In 1981–1985, 7% of the departments (all in the natural sciences) even had a citation peak as early as the second year. The number of social science departments attaining the minimum number of publications (N = 8) required to be included, was too small in both periods to yield reliable results. When all social science publications were combined, a non-significant shift from the fourth year (1976–1980) to the third year (1981–1985) was observed.

Table 1
Percentage of departments at the Agricultural University of Wageningen without citation peak in the third year

Field	1976–80	1981–1985
Life sciences	83%	73%
Natural sciences	69%	57%

To determine the resemblance of 2nd–3d year and 4th–5th year citation windows, Spearman's rank order correlations were calculated. For the WAU as a whole we found a coefficient of 0.64. The bioscience and natural science departments show comparable coefficients (0.63 and 0.64). Social science departments deviate greatly, however, with a coefficient of 0.38.

We conclude that a short-term citation window is applicable most of the times. However, we found a substantial number of departments with rather strong differences among the citation windows. Moreover non-ISI media also show rather large differences between the citation windows. Therefore, we used multiple windows in the present study.

3.3. Productivity and impact analyses

3.3.1. Production and impact of ISI publications. First, we studied the productivity and impact at the level of WAU, as measured by ISI articles (see Fig. 2). The total number of ISI publications increased with nearly 200% from 190 in 1976 to 563 in 1987. However, when normalized at the number of man hours available for research (fte), productivity increases at a more moderate rate from 0.7 publications/fte to 0.9 publications/fte on average. While the 1976 publications were cited 180 times in 1976–1978, those from 1987 were cited 973 times. The short-term impact increased from 0.9 c/p (citations per publication) in 1976 to an average of around 1.7 c/p during 1978–1987. Recently, the short-term impact has decreased slightly. However, at this level of aggregation, changes in impact may be rather trivial in nature and may be due to changes in organization, such as a reduction in researchers active in a field characterized by a low impact (e.g., sociology) and an increase in researchers active in a high impact field (e.g., biomedical research) rather than changes in research performance.

At the next level of aggregation, trends for the biosciences closely resemble that of the WAU. Not surprisingly, the agricultural WAU is dominated by the biosciences, which account for about 61% of research hours (fte). For the natural sciences, the output nearly triples from 60 to 171 ISI-publications, while normalized on fte, in 1978 a stable level of around 1.1 publication/fte is reached. The average impact increased between 1976 and 1985 from 2.3 c/p to 3.9 c/p, but has in recent years decreased somewhat to 3.6 c/p.

Productivity and impact are relatively low for the social sciences. Until 1983, yearly less than 9 ISI publications were produced, which nearly doubled to 17 in 1987. Normalized on research hours, the average doubled from 0.1 to 0.2 publications per fte. The average impact increased from 0.1 c/p to 0.9 c/p.

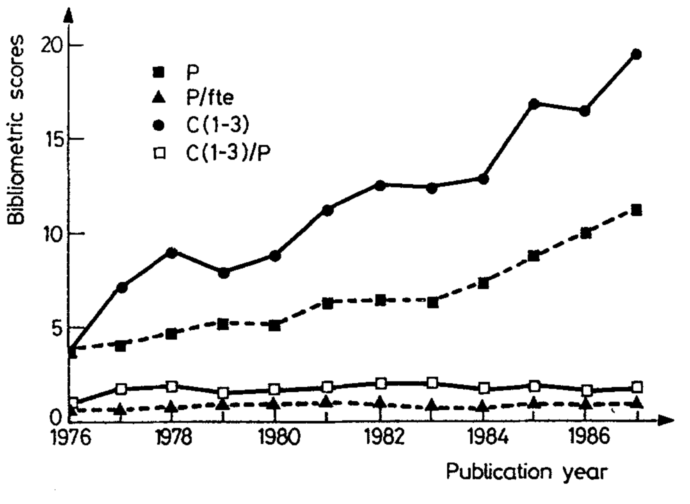

Fig. 2. Productivity and impact of WAU as measured by its ISI-publications.
Note: The number of publications (P; N = 3968) has been divided by 50, as has been the number of citations (C(1−3)). Self-citations have been excluded. The mean impact is represented by C(1−3)/P, while the average number of publications per fte is indicated by P/fte

3.3.2. Role of language in output and impact. Given the potential important role of language for a Dutch university, the percentage of publications in Dutch on the total number of publications was calculated. We found that the use of Dutch varies substantially among the aggregated scientific areas. On average, the WAU publishes about 13% of its scientific publications (N=823) in Dutch during the whole period. Social science departments publish more than 60% in Dutch. Bioscience departments are highly oriented to an international audience, and become even more internationally oriented recently, as is indicated by the decrease of the percentage publications in Dutch from 13% in 1976–79 to 8% in 1979–82. After this period the percentage remains stable. Finally, the natural science departments are most strongly internationally oriented, as the publish less than 3% of their scientific work in Dutch.

Figure 3 illustrates one effect of language on the impact of publications. For all document types relatively few publications in Dutch are cited at all. This is probably due to the relatively small language area of Dutch. Possibly, citations by other articles in the same journal raise the percentage of cited articles above 50% in the few Dutch language journals covered by ISI. Furthermore, even among cited publications, the impact of foreign publications (mainly in English) is much higher than the impact of publications in Dutch. For example, among the Dutch cited non-ISI items, 94% was cited less than 5 times, while the same figure for the foreign language items amounted to 69%.

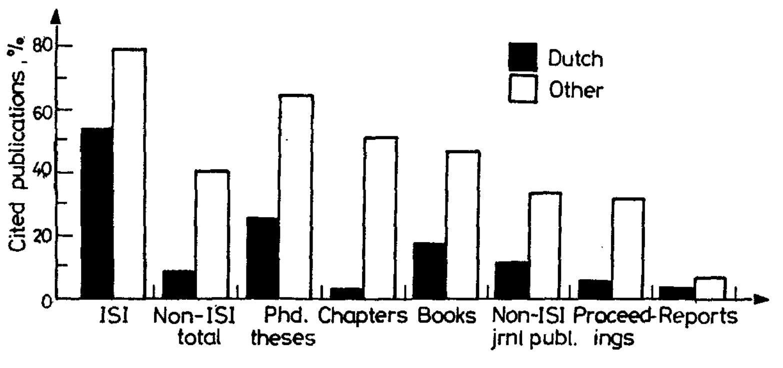

Fig. 3. Percentage of cited publications on the total number of publications by language and document type

3.3.3. Comparison of productivity and impact as measured by ISI- and non-ISI publications. The majority of scientific publications of WAU is not covered by ISI sources. As Figure 4 shows, the number of non-ISI publications always exceeds that of ISI-publications during 1976–1987. However, during the whole publication period, ISI-publications accounted for more than 80% of the total number of citations to natural and bioscience departments. The share of ISI publications in the total impact of social science departments increased substantially from 20% to 50%–60% in recent years. About a third of the departments which were cited at least 15 times (short-term window) obtained at least 30% of their citations from non-ISI publications (including two natural science departments, and more than 10 bioscience departments).

Document types differ substantially in mean impact. The impact of ISI-publications is substantially higher than the impact of non-ISI publications (see Fig. 5). Among non-ISI items, books, PhD theses, and chapters were cited relatively well, while contributions to proceedings and articles in non-ISI journals had a relatively low impact. Reports (not shown in Fig. 5) were hardly cited at all.

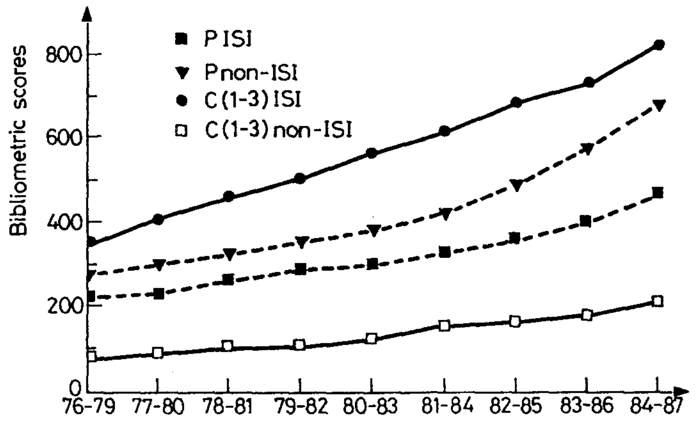

Fig. 4. Short-term impact of ISI- and non-ISI publications of WAU for 1976–1987.
Note: P ISI indicates the number of ISI-publications; P non-ISI indicates the number of non-ISI-publications; C(1–3) ISI indicates the short-term impact of ISI publications; C(1–3) non-ISI indicates the short-term impact of non-ISI publications

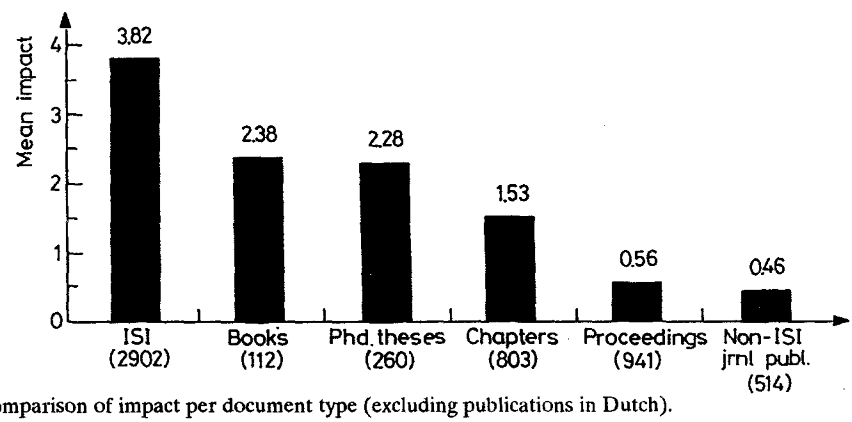

Fig. 5. Comparison of impact per document type (excluding publications in Dutch).
Note: The number of publications is indicated between brackets

3.3.4. Comparison of actual and expected impact of ISI publications. The above discussion deals primarily with trends in impact. So far, there is no *reference value* to decide whether the levels of the measured impact is 'low' or 'high'. We discussed in Section 2 that in our 'Leiden method' this problem is tackled by comparing the actual (measured) impact with an average value for the articles in the journals used by the group for publication. This latter value, called the average JCS (Journal Citation Score, see *Moed* et al., 1985a), can be seen as an expected impact value. Thus, the actual impact of ISI articles in the second or third year after publication is compared with that of the average paper in the same journal set (the expected impact). We have also taken the fourth and fifth year after publication as an alternative reference point.

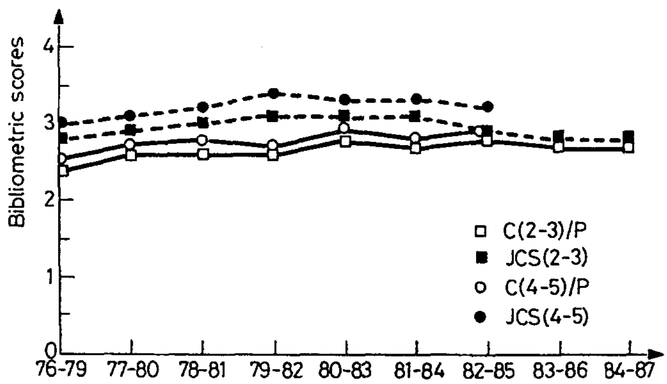

Fig. 6. Short- and medium-term actual and expected impact for the WAU.
Note: The actual short- and medium-term impact are indicated in the legend with respectively C(2 – 3)/P and C(4 – 5)/P, the expected impact with JCS(2 – 3) and JCS(4 – 5)

Figure 6 presents the trends on the level of WAU. For both windows, the expected value is highest in the early 1980s, but in recent years, WAU scientists have published, on average, in journals with somewhat lower impact values. During the whole period, the actual impact lies somewhat below the expected impact, but since 1982 – 1985, actual and short-term impact are nearly equal. This has been mostly due to the drop in expected value, while the actual value has increased slightly.

A potential problem with JCS-indicators is that they contain self-citations, which may distort trends. One way of coping with this has already been applied by excluding citation data from the first year after publication, which are mostly self-citations. If we look at self-citations among WAU publications, we find that these have increased from 31% to 38% during the observed period (using the short-term window), a level which might explain the observed shift in trend. However, random samples (both N=75) of non-WAU papers in the same journal set (the number of articles sampled from a journal was in proportion to the number of WAU papers in that journal) showed a similar increase in self-citations. Among the 1980 papers, 26.5% self-citations were counted, but among papers from 1987 the self-citation rate increased to 37.9%. As highly cited papers tend to contain less self-citations, we also compared the self-citation rates among papers cited less than 8 times (an increase from 32.3% to 43.2%) and more than 8 times (an increase from 22.1% to 30.9%). The medium-term impact is less influenced by self-citations (these increased for WAU papers from 22% to 26%), and thus may offer a better window. The ratio of actual/expected impact also increased, however (from 0.8 to 0.9 in 1976–1985). Thus, it seems that self-citations are not responsible for the increasing ratio of actual/expected citations at the WAU. At a lower level of aggregation, the trends among the biosciences again closely resembled that of the total WAU. The actual impact of publications in the natural sciences remains below expected level until 1982–1985, when they draw close. Until 1980–83, the actual impact of publications from the social sciences remains far below the expected impact. More recently, actual and expected impact approach one another, mostly due to a large increase in the level of the actual impact. Even after the large increase in self-citations from 0% to 48% has been discounted, a considerable increase in actual/expected ratio has been achieved.

3.3.5. Relation between actual and expected impact. We examined the correlation between expected and actual impact at three levels of aggregation. For the 27 most publishing authors at the WAU (publication period 1976–1985) correlation coefficients vary between −0.30 and 0.80. For departments, correlations ranged between −0.08 to 0.91 (excluding departments with less than 10 publications). For the WAU as a whole we found coefficients amounting to 0.43 and 0.41, respectively for 2nd–3d year and 4th–5th year windows.

3.4. Application of the indicators on departments: an example

Our bibliometric instruments are primarily directed at and most useful at monitoring departments rather than more aggregated bodies. As shown before, the WAU is characterized at various levels of aggregation by substantial differences in ISI-coverage, and in publication and citation characteristics. With this in mind, we have tuned the monitoring instruments for departments. To illustrate the approach, some of the main results of a bioscience department are presented.

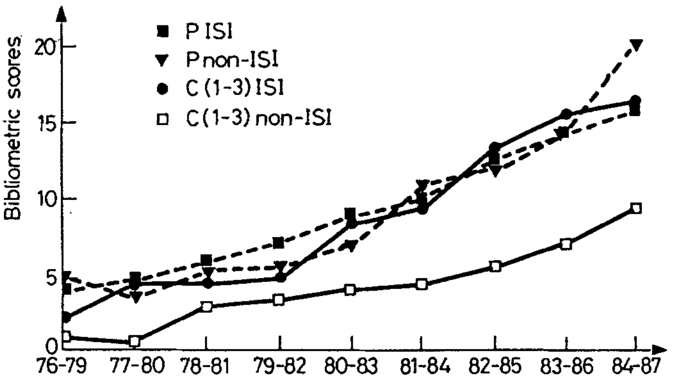

Fig. 7. Number of ISI-publication and absolute impact of these publications, compared to the number of non-ISI publications and the absolute impact of non-ISI publications (bioscience department, period 1976 – 1987).
Note: The absolute impact (C(1 – 3)) is calculated without self-citations for a window from year 1 to year 3. The aggregated non-ISI publications (P non-ISI; N = 128) concern articles in non-ISI tijdschriften (N = 23), contributions to books (N = 34), books (N = 1), PhD theses (N = 11), contributions to proceedings (excluded abstracts; N = 58), and external reports (N = 1). The data concern the year-average values for four-year publication blocks

The bioscience department publishes similar quantities of ISI and non-ISI publications during the whole period, although recently the number of non-ISI publications increases relatively strong (see Fig. 7). For both types, the output increases strongly, and still doubles after normalization on research hours (not shown). The department is mostly cited because of its ISI publications, although its non-ISI publications account for about 30% of its 'citation-income'. In Figure 8, the impact is specified by document type. The impact of non-ISI publications remains constant after an increase at the start of the period from 0.5 to 1.9 c/p, while that of ISI publications shows a slight upward trend. PhD theses and publications in non-ISI journals score highest, the latter in decreasing numbers. In several periods, the impact of publications in non-ISI journals exceeds that of publications in ISI journals. Until recently, in the Netherlands PhD theses were not produced by graduate students but by experienced researchers. The medium-term impact indicators (not shown) yield comparable results, with the exception of contributions to books which in recent years realize an impact comparable to that of PhD theses.

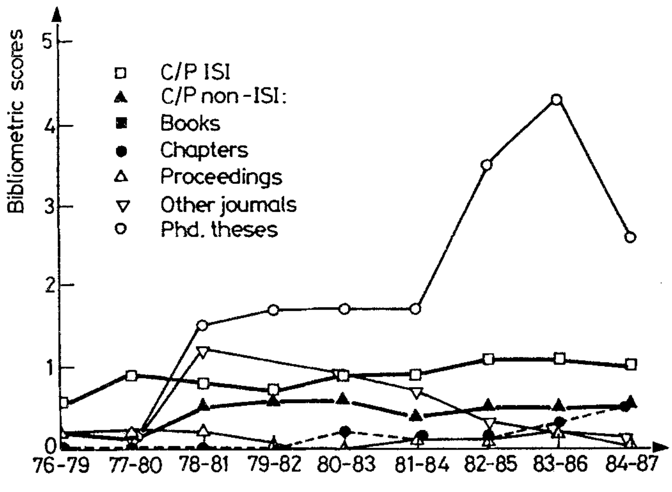

Fig. 8. Impact of ISI-publications compared with the impact of non-ISI publications (bioscience department, period 1976–1987).
Note: The impact (C/P) is calculated excluding self-citations for the short term (year 1 tot jaar 3). The number of ISI-publications amounts to 113. Non-ISI publications include: 23 publications in journals, 34 contributions to books, 1 book, 11 PhD-theses (internally prepared) and 58 contributions to proceedings (no abstracts)

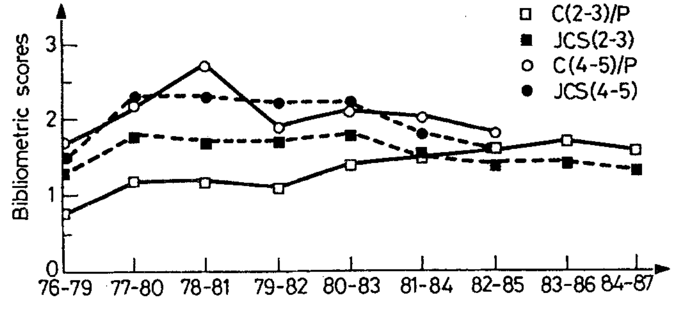

Fig. 9. Actual versus expected short- and medium-term impact of ISI-publications (bioscience department, period 1976–1987).
Note: The actual short- and medium-term impact are depicted in the legend with respectively C(2–3)/P en C(4–5)/P, the expected impact with respectively JCS(2–3) en JCS(4–5). The total number of ISI-publications amounts to 113

Finally, Fig. 9 shows the actual impact of the bioscience department compared to the expected impact. The actual short term impact increases in the 1980s, while the medium-term impact remains more or less constant during the whole period, with exception of a maximum in the year block 1978–81. Due in part to a decrease of the expected value, and in part to an increase of the actual impact, the department exceeds the expected impact from the year block 1981–84 on. When only the short-term impact is considered, the actual impact is clearly below expected level until 1981–1984, and one would have been lead to conclude, wrongly, that the research performance of this department tends to be below average. However, the medium-term impact is close to the expected level, indicating an internationally competitive impact. The percentage of self-citations of this department is slightly higher than the WAU average, but not alarmingly so. In general, the output of this department has increased considerably, while its impact increased slightly, and the ratio of actual to expected impact is around the international average, and tends to improve.

3.5. Validation of the indicators

To discuss the validity of our indicators, we have interviewed eight heads of (sub)departments from five departments at WAU, and one at the Veterinary Faculty of the University of Utrecht. The interviewees recognized both broader developments and more specific events which influenced output and/or input. Among the broader developments were increased pressure to publish from the university policy level, a shift in the university towards basic research, start-up periods with few publications, and a world-wide change of approach in one specialty, which diminished the impact of 'old-style' research. More specific events include the departure or long-time absence of a key professor, a publication in a prestigious medium, a 'cum laude' dissertation, and a specific congress. When asked, seven out of eight respondents stated that they agreed to a careful use of the present set of bibliometric indicators. One respondent indicated that he was on principle opposed to any use of indicators.

Finally, departments with research in tropical countries retrieved only partly the (mostly economic) usefulness of their research in the bibliometric indicators. This is a.o. a consequence of the strong application-oriented character of their research, which is honed towards development aid. In general, however, the indicators were found to give a good to very good reflection of the research efforts of the departments, on the condition that indicators are treated as interrelated.

3.6. Comparison of bibliometric results with a productivity and peer evaluation

We have tentatively compared results of our bibliometric impact analysis with a productivity evaluation by the 'Permanent Committee for the Conduct of Science' (PCCS), and also with an evaluation by a 'Visitation Committee' (VC). In particular, we have studied to what degree the bibliometric Journal Citation Score (JCS) analysis supports or conflicts with both PCCS and VC judgements. It should be noted that we see a good actual/expected ratio as a (necessary) condition for good research performance, rather than as sufficient evidence.

The PCCS judgements are mainly based on an analysis of both scientific productivity and (the PCCS assumed, quality related) internal and external funding during 1976–1985. The PCCS compared departmental performance figures at the university level rather than field-specific international levels. Based on the relative position of a department in the distribution of figures at the university level, scores (range indicated between parentheses: 1 = well above average; 2 = about average; 3 = well below average) were awarded on six parameters: # dissertations/fte (1–3); # publications/fte (1–3); # publications in refereed international journals/fte (1–3); # research assistants (2, 3); indirect funding by the national research council (1, 2); and acquisition of commissioned research funds (1, 2). Top departments scored at least twice a '1', while departments were considered to be below average if they scored at least twice a '3'.

The VC-evaluation assessed PCCS results by peer judgement. The VC-evaluation involved a substantive judgement of a set of publications selected by the departments (publication period 1983–1986), as conducted by one internal expert and up to ten external experts. This VC-evaluation can therefore be characterized as a peer evaluation. However, in two cases, the VC changed its final judgement by focusing on other aspects than research performance: the contribution of one department to teaching, and international recognition of another department.

Given the strong emphases of both evaluations on productivity and on local or national rather than international research performance, we do not feel that these evaluations can be used directly to validate our bibliometric results. However, the peer results can be used to gauge the extent to which bibliometric indicators focusing on international comparison can provide additional input to peer evaluations. Therefore, we compared the JCS results with those of both other evaluations.

For the comparison with the PCCS judgements we selected the publication period 1976–1985, and for comparison with the VC judgements the period 1976–1986. The

VC and PCCS used, a.o., productivity counts of international publications as a standard: the better than average departments should publish 50% more than the WAU average, while 'below average' departments published less than 50% of the average value, and 'average departments' made up the rest. In our bibliometric assessment, a parallel rating scheme was used, to render it as similar as possible to the assessments of the visitation committees. Departments in the highest category of both PCCS and VC evaluations should have an impact which is higher that the expected value. Departments in the middle category should realize between 50% and about 100% of the expected value. Finally, when departments in the lowest evaluation category realize less than 50% of the expected value, the bibliometric correspond with PCCS and VC judgements.

A first comparison between our bibliometric indicators with two evaluations concerning the departments of the WAU (N = 70) shows that the bibliometric impact analysis supports about 50% of the judgements (see Table 2). In general, results obtained with the bibliometric classification appear to be somewhat lower on average than those obtained with the peer classification.

Table 2
Degree of support of university evaluations (PCCS and VC) by the bibliometric impact analysis, in percentages of the total number of judgements

Committee	Impact analysis results			
	Higher than	Equal to	Lower than	Total
PCCS	14% (10)	54% (38)	31% (22)	100% (70)
VC	13% (9)	49% (34)	39% (27)	100% (70)

The major part of the judgements concerning the groups in the lowest category of both evaluations is supported (see Table 3). However, the major part of the judgements concerning the departments in the highest category of both evaluations is not supported by the bibliometric analysis. Moreover, about half of the judgements falling in the middle category is not supported. Respectively two and three departments from the lowest category score higher in the bibliometric impact analysis.

Table 3
Degree of support of the University evaluations by the bibliometric impact analysis, in percentages of the total number of judgements by category

A: Support of PCCS-judgements per PCCS-category:

Impact-analysis	High	Average	Low
equal to PCCS	7% (1)	54% (19+1)	89% (16+1)
higher than PCCS	*	22% (8)	11% (2)
lower than PCCS	93% (12+1)	24% (8+1)	*
Total	100% (13+1)	100% (35+2)	100% (18+1)

B: Support of VC-judgements per VC-category:

Impact-analysis	High	Average	Low
equal to VC	21% (3)	49% (21)	77% (10)
higher than VC	*	14% (6)	23% (3)
lower than VC	79% (9+2)	37% (14+2)	*
Total	100% (12+2)	100% (41+2)	100% (13)

The total number of judgements is indicated between parentheses; a number after '+' indicates border cases from the determination of the impact, e.g. a case with a impact substantially above the expected value, due to a very low number of publications.

The discrepancies between peer evaluations and bibliometric analyses can be largely explained by the fact that both peer evaluations do not take the international scientific context into account, and did look at productivity rather than impact. Finally, differences can be explained by the fact that in several cases, peers take into account background information (a.o. the teaching load).

4. Conclusions

Based on previously conducted research (*Moed* et al., 1985; *Nederhof*, 1988) we have developed a set of bibliometric indicators for agricultural and veterinary research in the Netherlands. Important new aspects are amongst others the application of medium-term impact windows and an extension of citation analysis to other publications than journal articles.

A first issue deals with the types of documents which are included in assessing research performance. In many bibliometric analyses, only publications in journals covered by ISI are studied. However, frequently research units publish a significant (sometimes even major) share of their work in 'non-ISI media'. For the WAU, 57% of the publications are not covered by ISI. In particular, the large majority of the scientific output of the WAU social sciences was not covered by ISI. These findings indicate that it is important to include all scientific publications in assessments of scientific productivity (cf. *Nederhof*, 1989).

In general, we found that ISI-publications had a substantially higher impact than non-ISI publications, although books and chapters were relatively well cited. Often, books have the highest impact on average (*Nederhof* et al., 1989; *Nederhof & Noyons*, 1992; *Nederhof & Van Raan*, in press), but in the case of WAU many of the books were locally published with limited distribution. The impact of both contributions to proceedings and publications in journals not covered by ISI was modest, while reports were hardly cited at all. The lower impact of non-ISI publications is partly due to a greater use of Dutch, as publications in Dutch are infrequently cited. Another factor may be a weak coverage of scientific areas by ISI-journals. Given the high ISI coverage of WAU non-Dutch journal articles (84%), this might only be of some importance to the social sciences, where coverage was lower, and to a few mostly biological specialties in which books represent the predominant medium. One peer review committee noted that for one department focused on 'design', publication in refereed scholarly journals was less central than prices in competitions, and so on.

During the whole publication period, ISI-publications accounted for more than 80% of the total number of citations to natural and bioscience departments, while their share in the total impact of social science departments increased substantially from 20% to 50%–60% in recent years. This suggests that limiting citation analyses to publications covered by ISI may offer a good first approximation of the total impact, especially for the natural and biosciences. However, about a third of the departments received an important share of their citations from non-ISI publications. Taking only ISI publications into account may then result in artifactual conclusions about research units in evaluation or monitoring studies.

A second issue concerns the optimal length of citation windows. Especially for policy purposes, it is often desirable to use short-term citation windows, as then more recent publications can be included. However, although it is a good approximation in many cases, the commonly used three year window is not universally applicable, as our results show. Many differences in citation characteristics were observed among the major disciplines, and at the level of departments. Also, we found slower diffusion of publications, as measured by citations, in 1976–1980 than in 1981–1985. Moreover, citations to journal articles and contributions to proceedings peaked earlier than books or chapters. Use of both short-term and medium-term citation windows addresses these points effectively, and allows more sophisticated and versatile analysis of impact patterns.

A third issue concerns the relation between actual and expected impact. If actual and expected impact correlate very highly, one might use the expected impact as a substitute indicator for the actual impact. On the other hand, if there is no relation at all between actual and expected impact, this implies that high quality journals fail completely in reviewing articles. At the level of a university, we found a moderately high correlation (0.4) between actual and expected impact both for short-term and medium-term citation windows. Apparently, these findings conflict with a small study conducted by *Seglen* (1989), who finds only low correlation values between the journal impact values and the actual impact on both author and department level. However, at comparable levels of aggregation as used by *Seglen*, we observed a large spread of values around the 0.4 correlation. In some cases, even negative correlations were obtained. Given the modest size of Seglen's samples, his findings do not contradict our findings.

Our findings indicate that although publications in journals with a high impact tend to be more highly cited than publications in journals with lower impact, this may not be true for non-random selections of individual papers. The impact of the journal in which a paper is published corresponds to only about 18% of the variance in impact of articles. Other characteristics of the paper account therefore for about 82% of the variance in impact. Thus, research units which publish in higher impact journals may very well gain a lower actual impact than research units who publish in journals with a lower mean impact. In general, the expected impact of journals seems a relatively poor predictor of actual impact, and it is necessary to collect actual impact data. However, we do not recommend to reduce impact analysis to the ratio of actual and expected impact. It is evident that if, within a field, research units which publish in low impact journals and research units which publish in high impact journals both score a similar actual/expected ratio, that then the latter perform on a higher level than the former. Therefore, as is also mentioned in the validation interviews, the various output and impact indicators of research performance should be used in combination in order to obtain an optimal impression of research performance.

Another source of error in the ratio of actual/expected impact is due to the inclusion of self-citations. Thus, a relatively high or relatively low percentage of self-citations may distort results. The percentage of self-citations increased considerably for the WAU during 1976–1987, but in a matched sample of non-WAU publications, a similar trend was present, both for highly cited and non-highly cited papers, so the actual/expected ratio was not effected. The self-citation rates were used to spot departments with a divergent level of self-citation. Fourth, language is a potentially relevant factor in research performance in non-Anglo-Saxon countries. As noted before, publications in Dutch rarely have a substantial impact in the international literature. Research groups with a predominantly national orientation, publishing mainly in Dutch, therefore will mostly have a low impact. Of course, this does not preclude that their publications have a national impact. In these cases, impact measurement offers a good indication of the low international impact of the work, but it is problematic for the assessment of national impact (cf. *Nederhof* et al., 1989).

Finally, bibliometric findings were compared on two levels with peer judgements. First, interviews in six departments show that our indicators offer generally a good description of the research efforts of those departments in the publication period concerned. According to the interviewees, both broader developments and specific events corresponded with bibliometric results.

Secondly, results of bibliometric indicators were tentatively compared with two productivity and quality evaluations conducted at WAU. The bibliometric impact analysis supports fifty percent of the total number of judgements concerning departments. However, the major part of the judgements concerning 'top departments' were not supported, while several 'bottom departments' did well bibliometrically. This can be mainly explained from the fact that the two university evaluations focused on productivity rather than impact, and used averages based on the university level rather than the international level.

We like to emphasize that bibliometric indicators are not to be used as an isolated instrument, but need to be integrated in a dialogue with stakeholders from the research group or scientific area involved. The results show that bibliometric indicators can provide important additional information to peer evaluations of research performance.

References

DAVIS, P., G. F. PAPANEK, Faculty ratings of major economic departments by citations, *American Economic Review*, 74 (1984) 225 - 230.

HARNAD, S., (Ed.), *Peer Commentary on Peer Review*, Cambridge, Cambridge University Press, 1982.

MARTIN, B. R., J. IRVINE, Assessing basic research: some partial indicators of scientific progress in radio astronomy, *Research Policy*, 12 (1983) 61 - 90.

MEIJER, R. F., NEDERHOF, A. J., VAN RAAN, A. F. J., *Aspects of Knowledge Transfer in Agriculture in the Netherlands*..Research report to the National Council for Agricultural Research (NRLO). Leiden: CWTS, 1991.

MOED, H. F., BURGER, W. J. M., FRANKFORT, J. G., VAN RAAN, A. F. J., The use of bibliometric data for the measurement of university research performance, *Research Policy*, 14 (1985a) 131 - 149.

MOED, H. F., BURGER, W. J. M., FRANKFORT, J. G., VAN RAAN, A. F. J., A comparative study of bibliometric past performance analysis and peer judgement, *Scientometrics*, 8 (1985b) 149 - 159.

MOED, H. F., *The Use of Bibliometric Indicators for the Assessment of Research Performance in the Natural and Life Sciences*, Leiden, DSWO Press, 1989.

NEDERHOF, A. J., The validity and reliability of evaluation of scholarly performance, In: A. F. J. VAN RAAN, *op. cit.*, pp. 193 - 228.

NEDERHOF, A. J., Books and chapters are not to be neglected in measuring research productivity, *American Psychologist*, 44 (1989) 734 - 735.

NEDERHOF, A. J., NOYONS, E. C. M., International comparison of departments' research performance in the humanities, *Journal of the American Society for Information Science*, 43 (1992) 249 - 256.

NEDERHOF, A. J., NOYONS, E. C. M., Assessment of the international standing of university departments' research: A comparison of bibliometric methods, *Scientometrics*, 24 (1992) 393 - 404.

NEDERHOF, A. J., A. F. J. VAN RAAN, A validation study of bibliometric indicators: the comparative performance of cum laude doctorates in chemistry, *Scientometrics*, 17 (1989) 427 - 435.

NEDERHOF, A. J., A. F. J. VAN RAAN, A bibliometric analysis of six economics research groups: A comparison with peer review, *Research Policy*, (in press).

NEDERHOF, A. J., R. A. ZWAAN, R. E. DE BRUIN, P. J. DEKKER, Assessing the usefulness of bibliometric indicators in the humanities and the social sciences: A comparative study, *Scientometrics*, 15 (1989) 423 - 437.

PRICE, D. J. DE S, *Little Science, Big Science*, Columbia University Press, New York, 1963.

SEGLEN, P. O., From bad to worse: evaluation by journal impact, *Trends in Biochemical Science*, 14 (1989) 326 - 327.

SPANGENBERG, J. F. A., W. BUIJNK, W. ALFENAAR, Some incentives and constraints of scientific performance in departments of economics, *Scientometrics*, 18 (1990) 241 - 268.

VAN RAAN, A. F. J. (Ed.), *Handbook of Quantitative Studies of Science and Technology*, Amsterdam, North-Holland/Elsevier Science Publishers, 1988.