Research Evaluation, 16(2), June 2007, pages 93–105
DOI: 10.3152/095820207X227529; http://www.ingentaconnect.com/content/beech/reev

Modelling of bibliometric approaches and importance of output verification in research performance assessment

Thed N van Leeuwen

This paper presents a model describing the various possibilities in the application of bibliometric techniques in evaluation processes. The model distinguishes between the goals and functions of a bibliometric analysis, and clearly indicates the limits of the various bibliometric approaches. In the so-called top-down or bottom-up approaches, the importance of verification of publication material is indicated. Another important feature of the paper is the description of the problem of the interpretation of the relationship between on the one hand organizational structures, and on the other hand fields of science, in the light of the two main approaches in bibliometric studies. Finally the paper contains a number of case-study examples of the model described.

BIBLIOMETRIC TECHNIQUES have been applied in evaluation procedures worldwide over the last decades. A wide range of applications has been used in various countries — in the USA, for example, the National Science Foundation (NSF); in the UK, for example, the Research Assessment Exercise (RAE); and in the Netherlands, for example, the disciplinary studies by the Dutch Association of Universities (VSNU) — on different levels of aggregation (country, organization, research group, and even individual researchers). The last couple of years have shown a strong focus on 'benchmarking' and 'scientific excellence'. While this focus might be justified by the wish of society and politics to obtain some sort of accountability of science, not all bibliometric techniques have the same level of expressiveness, and they are not equally applicable on all levels of aggregation mentioned above.

Therefore, it seems necessary to distinguish between various bibliometric approaches, and the results and consequences of the application of these techniques. In applying bibliometrics in evaluation procedures (of any kind), we need to make a number of distinctions. In the following we will present a model according to which the different methods and approaches can be 'classified', and a correct application of bibliometric techniques will follow.

Thed N van Leeuwen is at the Centre for Science and Thechnology Studies (CWTS), Leiden University, Wassenaarseweg 52, PO Box 9555, 2300 RB Leiden, The Netherlands; Email: leeuwen@cwts.leidenuniv.nl.

Modelling bibliometric approaches

If bibliometric techniques are applied as a diagnostic tool (for example, How is this university performing scientifically?), we distinguish two different *functions*, namely *descriptive* and *evaluative*. In bibliometric studies, data-collection can start at a very high level of aggregation (for example, on the country or organizational level), to the level of research groups. Depending on the degree of accuracy in the data-collection process, the results will have farther-reaching consequences. This relates to the next distinction. If bibliometric analysis is applied with the intention to measure scientific impact, we distinguish two main *goals*, namely gaining insight into the 'past performance' of a certain entity, or gaining insight into the 'current research potential' of a certain entity. Again, the data-collection procedure is making the difference.

An important aspect in this procedure concerns verification of publication data in a study. If the collected publication material is compared to lists of output by an entity, or if the researchers involved in the bibliometric analysis are asked to verify the underlying publication material themselves, the results of the study will be a stronger representation of the output used in the research performance assessment. In other words, verification by ‘a second party’ (that is, not the bibliometric analyst, but for instance the researchers involved in the evaluation, or research managers) will make the underlying publication material more reliable, and consequently can be analyzed and interpreted in more detail.

While distinguishing these functions and goals, we can ‘mix’ them to a certain extent. Bibliometrics can be applied to evaluate the past performance as well as the research potential of a research entity, if the verification of the underlying data-material has been carried out accurately and precisely, thereby identifying and linking the publications to the researchers of the entities involved. However, insufficient care in this verification process as an integral part of research performance assessment procedures would mean that the diagnostic function of bibliometrics can only have its consequences on the insight into the ‘past performance’, while the underlying publication data do not allow interpretation of the current ‘research potential’ available.

The above distinction between functions and goals relates strongly to a distinction often made in the bibliometric practice: a ‘top-down approach’ versus a ‘bottom-up approach’. In the top-down approach, the highest level of aggregation determines the starting point for data-collection. So for example, data are collected on the level of a whole country or university, based on the country name, or a university defined by a unified ‘main organizational’ name. Consequently, on the basis of the collected publication data, bibliometric comparisons are made between countries or large institutions such as universities. While the data-collection on the level of countries is in most cases straightforward, the data-collection on the level of main research organizations (for example, universities, companies or public research institutes) requires a previous effort to unify and ‘clean up’ the institutional names and the underlying addresses attached to publications. This ‘unification’ procedure requires a thorough insight into the structure of both the address material attached to scientific research output as well as insight into the research system currently analyzed (at any level of aggregation). Eventually, a lower level of aggregation can be reached, but only after a careful cleaning of the relevant parts of addresses attached to scientific research output, and additional information derived from personnel of the initiating organization.

The bottom-up approach starts at the other end of the organizational spectrum, namely at the lower levels of aggregation (that is, researchers and research groups). For this approach, the publication oeuvres as far as covered by the Citation Indexes, or nowadays the Web of Science (WoS) of Thomson Scientific, the former Institute for Scientific Information (ISI), need to be collected on each required level, and verified by the researchers involved, or any knowledgeable third party. This verification process can result in deleting certain publications that were selected erroneously, or adding publications that were previously missed in the data-collection procedure. In a later section we will return to this procedure. From the collected and verified sets of publications, higher levels of aggregation can be reached by de-duplicating the gathered publication material on this higher level.

The model in schemes

A schematic presentation of both the top-down approach as well as the bottom-up approach will be shown in the following. We first introduce the basic building blocks of the model, components of nearly any bibliometric study. These building blocks are presented in Figure 1.

In Figure 1, we find the main organization (here indicated by a university), the underlying organizational parts (in this case, four laboratories), the

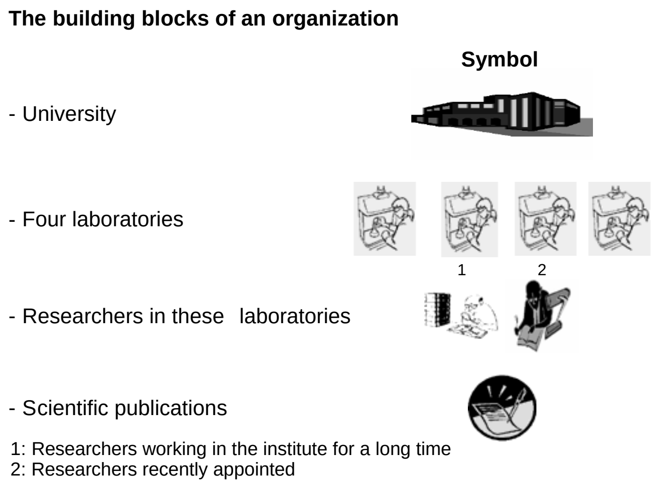

Figure 1. Overview of the organizational components relevant in bibliometric analysis

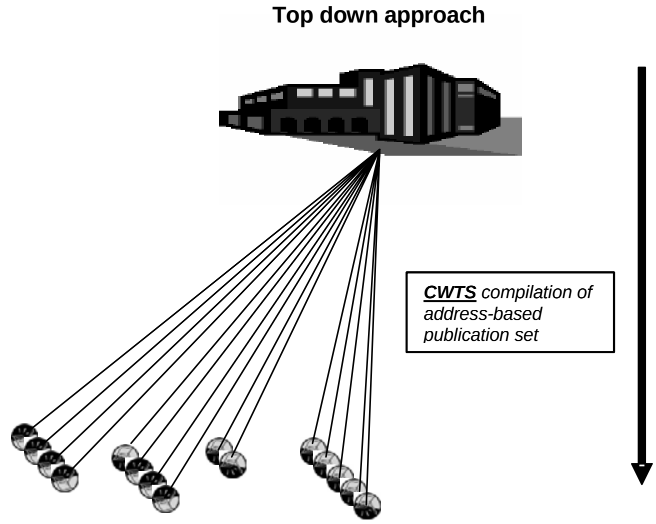

Figure 2. Schematic overview of the top-down approach in bibliometric analysis

researchers working in these labs, and their research output (publications) In Figure 1, two types of researchers are indicated, researchers who have been working for a long time in a research organization (1), and researchers who have been appointed only recently (2). This ‘distinction’, which in practice will always be a mix, allows a clearer explanation of the earlier discussed distinction between ‘past performance’ and ‘current research potential’ analysis. Finally, we distinguish between scientific publications.

A scheme of the top-down approach is presented in Figure 2. In this scheme, we observe a large ‘distance’ between the organizational level and its underlying research output. Besides the direct link between the main organization and its research output, no other apparent organizational units ‘exists’ between the main research organizational level and the underlying research output. On the basis of searches for names and variants of research organizations, sets of publications for different organizations or institutions can be compiled, and as such can be the starting point for bibliometric comparison of research organizations. This methodology is applied in benchmarking or main organizations in science, and the resulting ranking of main organizations.

Figure 3 shows the schematic overview of a bottom-up approach. As stated before, the starting point of this procedure is the individual researcher. For each researcher, the publication oeuvre published in internationally refereed journals as far as covered in the Citation Indexes is collected and verified. As becomes apparent from Figure 3, both types of staff members are involved. This implies that with respect to the *goals* of bibliometric analysis, both a past performance analysis as well as a research potential analysis are possible. The bottom-up approach allows for analyses that also include the publications of recently appointed scientists, including their research output generated elsewhere, while in the top-down approach this can never be accomplished.

Aggregation of the groups(s) of researchers to higher levels of organizational structure becomes possible by combining and de-duplicating research

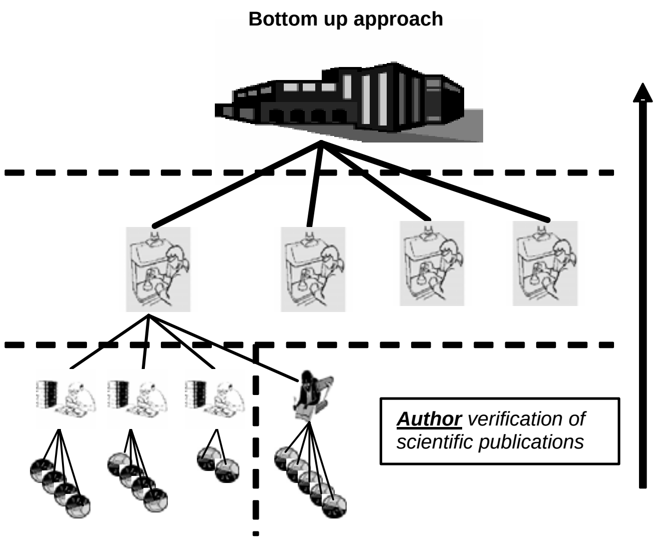

Figure 3. Schematic overview of the bottom-up approach in bibliometric analysis

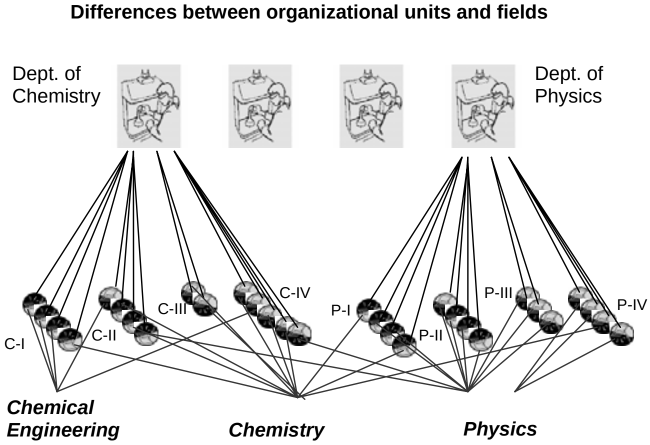

Figure 4. Schematic overview of the structure between research units, their publication output and their fields

output oeuvres. At each of the higher levels within the organizational structure, the underlying publication material (based on [verified] individual researchers' output) allows for farther-reaching conclusions and consequences.

The discrepancy between the bottom-up and the top-down approaches becomes most clearly visible in the following hypothetical situation. In Figure 4, a schematic overview is given of the structure of the scientific process, with actors (research units), their output (publications in scientific journals), and the disciplinary structure into which this output fits in (fields of science). We observe two departments, one in chemistry and one in physics, each having four scientists (indicated by C-I to C-IV and P-I to P-IV, respectively) with their oeuvres, and three scientific fields (chemistry, physics, and chemical engineering).

This schematic structure clearly displays the situation that is practically always found in research evaluation procedures: scientists publish their results in a variety of scientific journals, mostly belonging to more than one field, not necessarily corresponding to the field-specific name of their department. Consequently, no 'organic' link exists between the fields of science available for researchers, and the organizational structure often found in universities and research institutes.

One of the often-applied analyses of the Centre for Science and Technology Studies at Leiden University (CWTS) relates to so-called 'research profiles'. Such a profile provides insight into the distribution of the publication output over fields of science. Within a top-down approach, the interpretation that occurs often is the 'direct linking' between a certain field (for example, condensed matter physics) and a specific organizational unit apparently related (for example, the department of condensed matter physics) to that scientific discipline (as if a one-on-one relation exists).

Figure 4 thus illustrates a very frequent misunderstanding in the correspondence of fields to research units in the organization: linking. While in the data-collection of this hypothetical top-down situation, the relation between research units and their fields was not covered, the interpretation of the data-material easily exceeds the possibilities of the chosen approach.

A further application of the top-down approach can be found in Figure 5. Here a more sophisticated process of data-collection and verification leads to

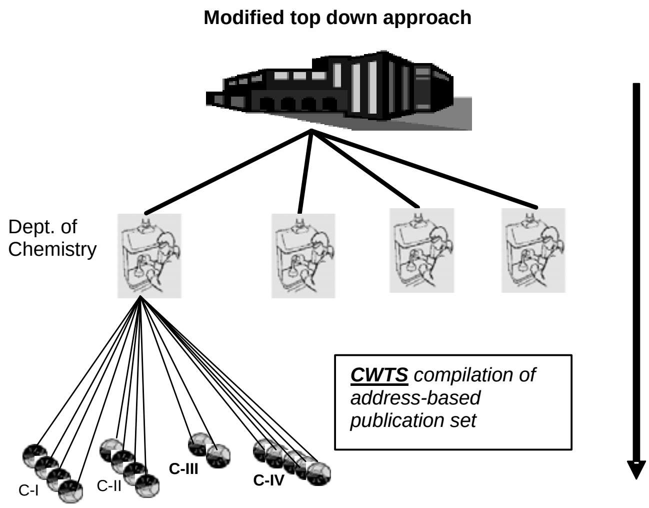

Figure 5. Schematic overview of the more sophisticated top-down approach

analyses on the level of organizational units. In this approach, the publication data are collected on the main organizational level, and a further analysis on lower levels of the addresses attached to scientific research output, including a thorough verification and cleaning of this data-material, leads to reliable information on lower levels of the organizational structure.

As can be concluded from a comparison of Figures 3 and 5, the bottom-up approach and (adjusted) top-down approach move closer towards each other in terms of the level of aggregation and the level of reliability; still these two approaches are not equal to each other.

The verification process

In the beginning of the 1990s, the period when CWTS began to apply research performance studies on a large, national scale (for instance, national analyses in astronomy and astrophysics, biology and chemistry), the idea occurred to streamline the process of data-collection. Software routines were developed to collect publication data-material, and to calculate indicators automatically. Although at a very rudimentary stage, this was the foundation for current routines. One of the most important topics raised was the problem of validity in collecting publication material from large groups of scientists in a particular field. Initially, this was solved by sending each individual scientist the publication lists created by entering his/her name into the ISI-based publication data-system, with a request to control and verify the lists. Although this worked remarkably well, and scientists were willing to cooperate with this procedure, the problem of ‘continuity’ in the data-collection popped up at the same time. By continuity, we mean the issue of not bothering scientists with the same request in any future study, overlapping with a previous study. Therefore, a method was developed to allow scientists to verify their publication material in an electronic environment.

Thus the lists of publications of each scientist were loaded on a floppy diskette, and sent for verification, in most cases via a person in the organization with a coordinating task. After verification, the diskettes were sent back to CWTS, and re-entered into the data-system. This offered the possibility of creating a ‘history’ for each scientist who has cooperated in CWTS verification procedures, thus avoiding verification of the same material twice. This was experienced by the scientists involved as a useful improvement in the CWTS procedure. Keeping pace with technological development, CWTS is nowadays offering the scientists involved in research performance analyses the possibility of verifying their publication material via an Internet connection, thereby ‘shortening’ the distance between the scientists involved and the analysts at CWTS. This is further facilitated by a helpdesk, which offers help in questions related to coverage and technical problems.

All this is related to research performance studies, in which the scientists involved play a significant role and, as such, it becomes clear that this procedure is not (very easily) applicable on a large scale (for example, the national level).

Results

Case 1: Biochemistry and molecular biology at the Delft University of Technology

The first two cases described below both relate to a study conducted by CWTS for the Delft University of Technology (TUD; van Leeuwen, 2001). In order to fully understand the problems resulting from the chosen approaches, we first need to clarify the starting position. As mentioned above, many of the interpretation problems arise when one is misinterpreting our earlier discussed ‘research profiles’. These profiles represent the distribution of the publication output over fields of science (represented by ISI journal subject categories). Although by far from an ideal classification system, it is currently the only available system that fits the multidisciplinary nature of the ISI Citation Indexes.

In the research profiles, the impact score is indicated by *CPP/FCSm*, which stands for a comparison of the actual citation score (*CPP*, or citation per publication ratio) with a worldwide field-normalized reference value (*FCSm*, or mean field citation score, a weighted average field impact score in which the weights are being determined by the total number of papers published in each field by the entity). *CPP* is an important impact indicator, as it represents the ‘visibility’ of an organization normalized to its output, while *FCSm* is the representation of the impact in the fields in which a unit is publishing its work. So by comparing the actual impact with mean impact per field, in the light of the distribution of the output over fields of science, users of these research profiles are provided with an overview of the scientific scope and interdisciplinarity of research entities.

In Figure 6, we observe the distribution of the total output of the TUD in the period 1980–1997 over fields of science. Both the basic sciences as well as the engineering sciences are well presented in the research profile of the TUD. The length of the bars is proportional to the output shares of the TUD in each field, and the ‘colour’ of the bar represents the value of the impact.

An impact value ranging between 0.8 and 1.2 is indicated with shading, and above 1.2 with black bars. From Figure 6, it becomes clear that TUD has, within the top-30 ranking fields covered in terms of output shares, no fields with a low impact score.

Some parts of this research profile, however, raised questions within the commission supervising the study. A first problem arose in the interpretation

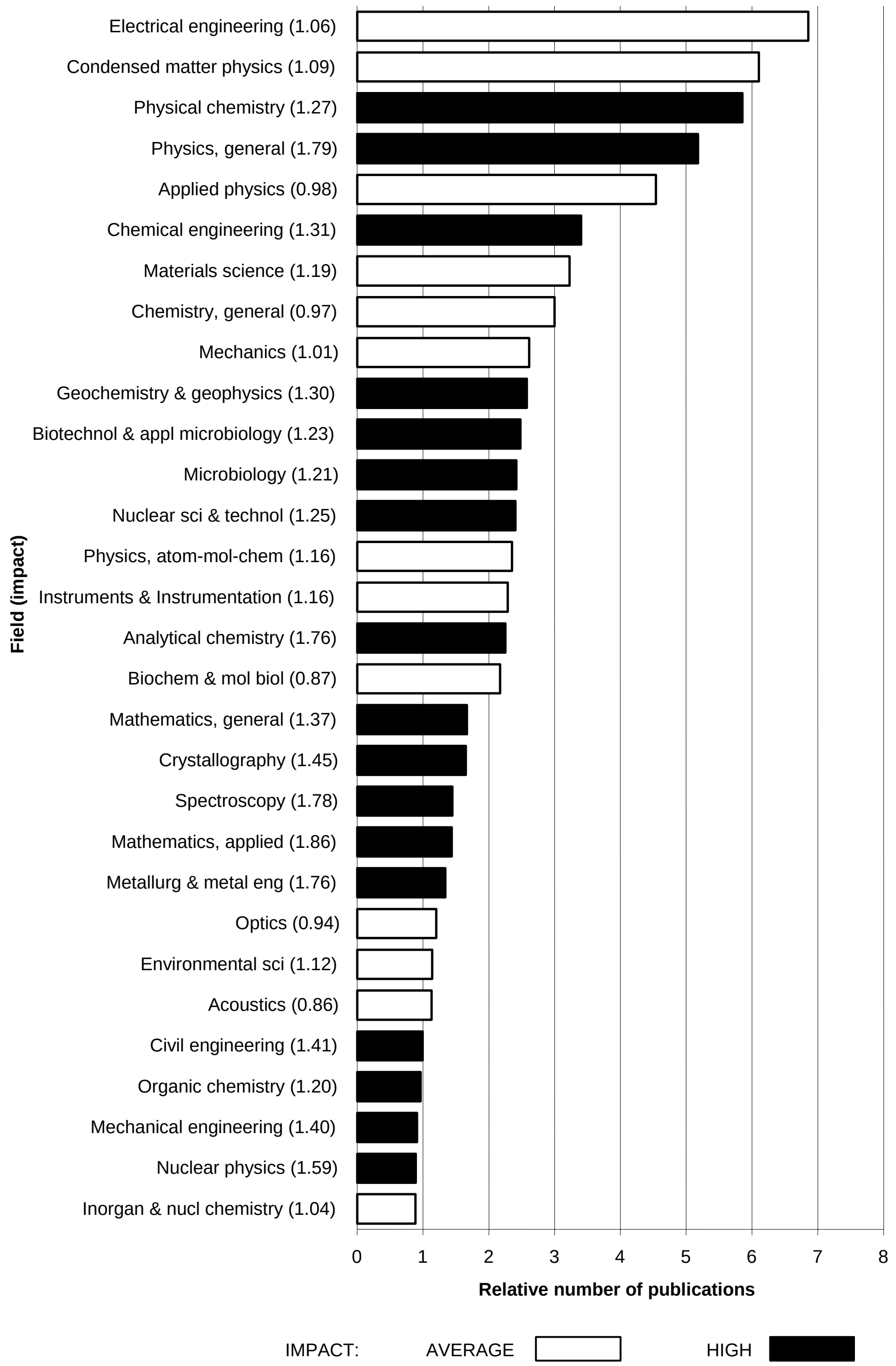

Figure 6. Research profile: output and impact per field, Delft University of Technology, 1980–1997

of the TUD research profiles for the field of biochemistry and molecular biology. Although represented in the profile with a shaded bar, indicating an average impact score (the actual *CPP/FCSm* of this field for TUD was 0.87), this outcome was experienced as rather awkward, because TUD has a high international reputation in the field of biochemistry and biotechnology (particularly, the Kluyver Laboratory).

How is it possible to have internationally highly renowned research groups in a specific field, and a ‘mediocre’ impact score in that field in the research profile of the university as a whole? In order to solve this problem, the following procedure has been followed.

The publications in the field of biochemistry and molecular biology have been carefully screened and labelled as originating from groups active primarily in this field (in the case of the Delft University of Technology, mostly groups from Kluyver Laboratory), and groups only partially active in this field. We have labelled the first group of publications as originating from the *core*, whereas the second set of publications is labelled as *non-core*. These two sub-sets of publications are analysed according to the same method as the entire field of biochemistry and molecular biology. The results of this analysis are displayed in Figure 7.

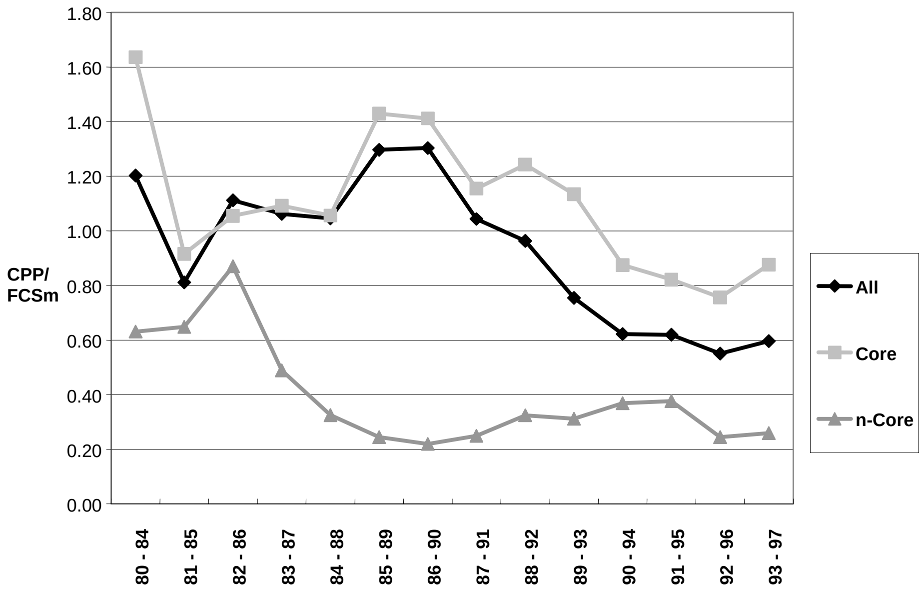

Figure 7. Impact analysis in biochemistry and molecular biology, for the entire University of Technology Delft, and for the core and non-core publication sets, trend analysis 1980–1997

Analysis of this field for the entire Delft University of Technology shows, in comparison with worldwide field scores, an average impact (with *CPP/FCSm* around 1), with a steady decline in impact from 1985–1990/1986–1991 onwards, resulting in a significant low impact score (0.60) in the second part of the period covered. The analysis of the publications labelled as *core* shows a somewhat different perspective. These publications also show a fluctuating impact, with peaks in 1980–1984, 1985–1989, 1986–1991, and 1988–1993, and a decline in impact from 1989–1993 onwards, reversed in the last period of the trend-analysis 1993–1997. The impact scores of the publications labelled as *non-core* show a totally different result. These publications have in general a very low impact, and contribute only marginally to the *impact* of the entire field for the Delft University of Technology. However, these publications do contribute for a considerable extent to the output of the Delft University of Technology in this field, between 20% and 40%, thereby 'diluting' the impact of Delft as a whole in biochemistry and molecular biology.

A conclusion of the macro-level study was the declining impact of the Delft University of Technology in biochemistry and molecular biology from 1986–1990 onwards. On a macro-level, this conclusion is correct. However, one should not conclude from this measurement that all research activities in Delft in this field are below the worldwide impact score. This case study shows that the data-sets on a macro-level (as a result of the top-down approach chosen for the Delft University of Technology) and a micro- or meso-level (as resulting from this more thorough investigation) are not mutually exchangeable. It also shows how the bibliometric results, based on data used on a macro-level, will change if the data-set is disaggregated into sub-sets directed at meso- or micro-levels.

Yet another result of this focusing of the different organizational units contributing to a whole field for a university is of a completely different nature. Figure 7 also displays a somewhat low impact situation for the so-called core research units, from the beginning of the 1990s onwards, with the start of a rising trend in the last period of the analysis. From conversations with research managers from Delft University of Technology, the cause for this decreasing impact of the core research units in this field had to be sought in the situation of the research staff and, more particularly, in professor rank level. It became clear that, in a few research groups, professor vacancies have been unfulfilled for a number of years. This might significantly contribute to the somewhat low impact trend of Kluyver Laboratory research output between 1990 and 1996.

Case 2: Mechanics and mechanical engineering at the Delft University of Technology

Case 2 relates to the same study conducted for the Delft University of Technology. In this case, the fields of mechanics and mechanical engineering are scrutinized more closely. While TUD is represented rather well in studies concerning the national R&D system (NOWT, 2000), the research profile of TUD (Figure 6) indicates a prominent output position for mechanics (the ninth field, with 2.5% of the total output of the university in the period 1980–1997), combined with a modest impact score (*CPP/FCSm* = 1.01), and relatively low output position for mechanical engineering (less than 1% of the output) combined however with a rather high impact score (and *CPP/FCSm* = 1.40). This asked for a more thorough investigation, especially in relation to the national studies, in which the field of mechanics (in fact, a combination of the ISI journal category-based fields mechanics, and mechanical engineering) at TUD was presented as outstanding.

In this case, the publications labelled as mechanics or mechanical engineering were analyzed in more detail, and especially the addresses attached to these publications. The addresses were used to distinguish the different faculties at TUD, thereby allowing a more detailed insight into the distribution of both output shares and impact over faculties. Contrary to the previous analysis, a break-down into so-called core and non-core shares was not possible, because of the large number of different organizational units within Delft University of Technology that publish in both fields. As an organizational unit, the faculty that is mentioned in the heading of the publication (as far as possible/available) was used to indicate the origin of the publication.

Figure 8 displays the distribution of research activity of faculties of the Delft University of Technology in both fields, as well as the impact. Again, this special research profile clearly shows that there are no core and non-core shares. However, it does show that in terms of strength or weakness in these two fields, we can distinguish between organizational units having a relative stronger position than others, both in terms of output as well as in terms of impact (mechanical engineering, civil engineering, and applied physics, with output shares ranging between 25% and 11%, and impact scores above worldwide average).

With respect to the comparison with the national evaluation, one needed to consider the different time-periods covered by both studies. While the study for TUD covered a 17-year period, from 1980–1997 onwards, the national study covered only the last part of this period. The differences in results between both studies (focusing on either the national level or the organizational level) can thus be explained from choosing different time-intervals for the impact measurement of this specific discipline within the Delft University of Technology, thereby indicating the importance of the parameters chosen in different types of bibliometric analysis.

Case 3: Personnel management at the Delft University of Technology

The third case focuses on the differences between the two main goals of a bibliometric analysis, namely, assessment of past performance and assessment of current research potential. Both goals strongly contribute to the insight of research leaders in the situation of their organization. While assessment of past performance serves as an 'accounting'

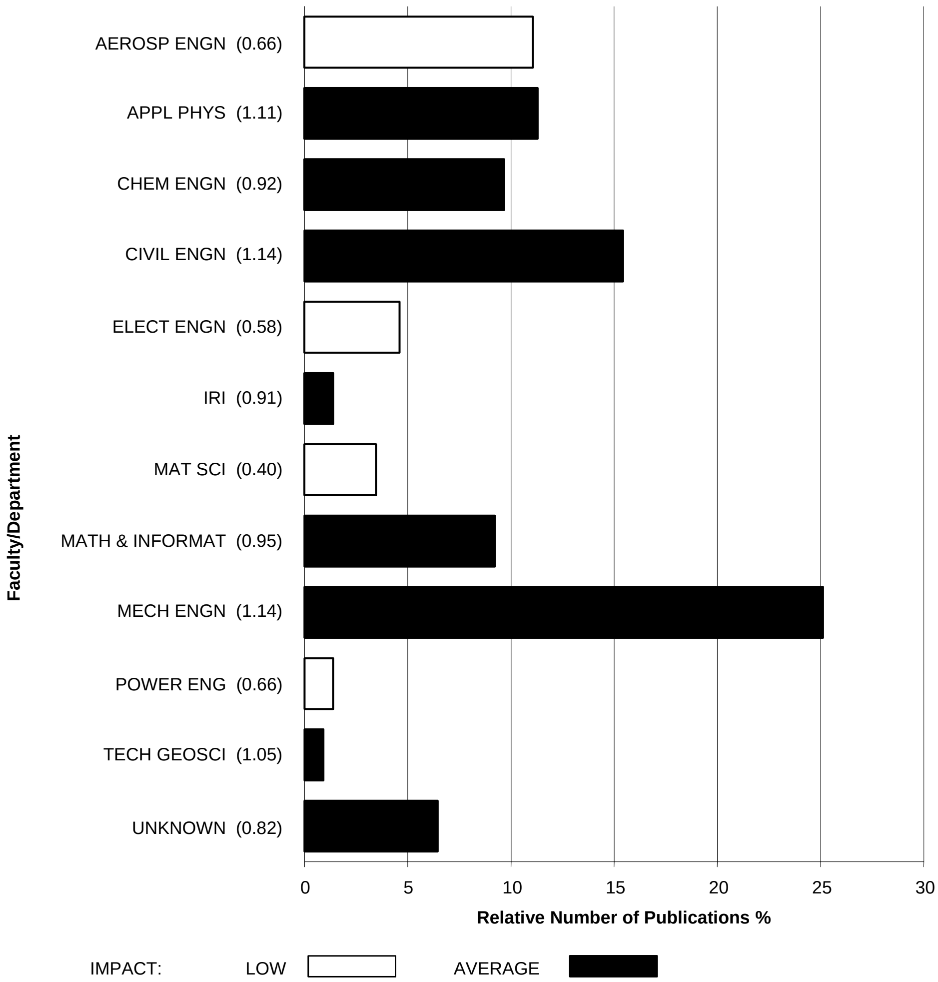

Figure 8. Research profile: output and impact per organizational unit in mechanics, and mechanical engineering, Delft University of Technology, 1980–1997

Table 1. Analysis of publication output, Faculty of Applied Science, Delft University of Technology, 1999

Goal	Target group	Number of staff
Assessment of past performance	All research staff appointed	203
Assessment of current research potential	Only currently appointed research staff	170

tool, that is, previous research efforts and directions can be valued; the assessment of current research potential analysis provides insight into the future prospects of a research group, based on the hypothesis that the most recent results are the best predictor for future research results.

The data in this case result from a study for the Faculty of Applied Sciences (TNW) of the Delft University of Technology conducted in 1999. The analysis concerned the publication output of 1989–1998 from the senior research staff of four departments in two modalities:

- as far as these researchers were appointed in the ten years of the analysis, and
- by the current research staff.

The output was verified by the researchers themselves, or in cases where they no longer held appointments in Delft, by their former group leaders.

Schematically, the analysis is constructed as shown in Table 1. In such analyses, one needs to take into consideration the fact that normally a large overlap exists between the two sets of publications resulting from the two modalities defined above. Publications that do not belong to the overlap are publications from research staff who no longer hold appointments, or publications from recently appointed research staff realized elsewhere. Their research results, as published in scientific journals, from their former research 'environment', and falling within the same time-framework (that is, the period 1989–1998) are counted as part of the potential of the current research staff.

Figure 9 displays the differences in output (*P*) and actual impact (*C*) in both modalities. The output *P* is somewhat larger for the research staff appointed during 1989-1998, compared with the 'current' staff, and this is strongly influenced by the age factor: people no longer holding appointments have moved on in their career, either to appointments elsewhere or into retirement. In either way, their research output is often voluminous. On the other hand, the current research staff consists of scientists working on their research output for, on average, a shorter time. Therefore, both output and impact is often smaller/lower for the recent research staff as compared to the 'earlier' staff.

Figure 10 shows the comparison between the 'past performance' approach and the 'research potential' approach on a number of impact indicators. The output of the current research staff has a slightly higher mean impact score (*CPP*) than the previous research staff (but not significantly higher). With respect to the normalized impact scores *CPP/JCSm* (actual impact compared to average journal impact), *CPP/FCSm* (actual impact compared to average field impact), and *JCSm/FCSm* (average journal impact compared to average field impact), the current research staff performs as well as the previous staff. And although there exists a large overlap between the research staff in both modalities, this still leads to the important conclusion that the faculty has not suffered bibliometrically from the 'loss' of more senior researchers, and their younger replacements. This type of insight, provided by bibliometric analysis, can be helpful in research policy (especially on the lower levels of aggregation of departments or research groups); and it is not easily available in peer

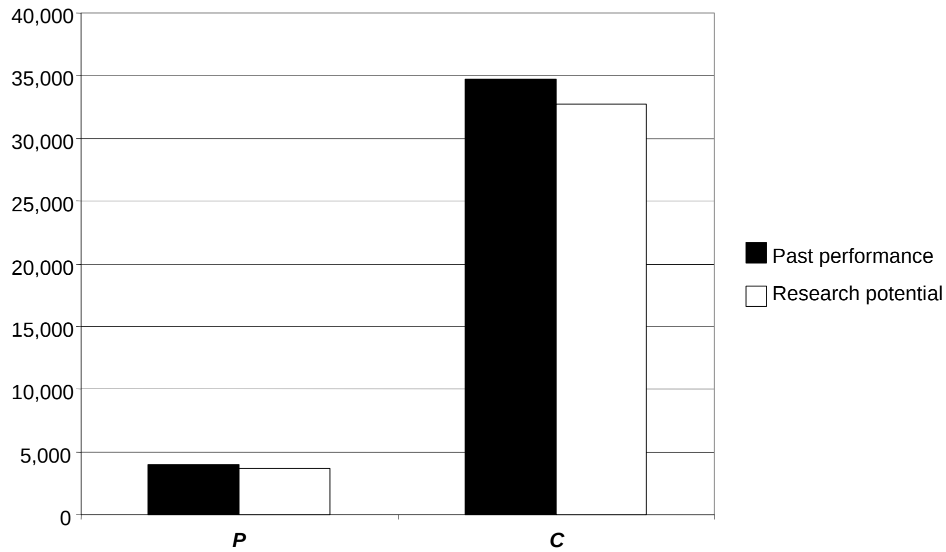

Figure 9. Output and impact volume for the 'past performance' and the 'current research potential' approach modalities at the Faculty of Applied Sciences, Delft University of Technology, 1989–1998

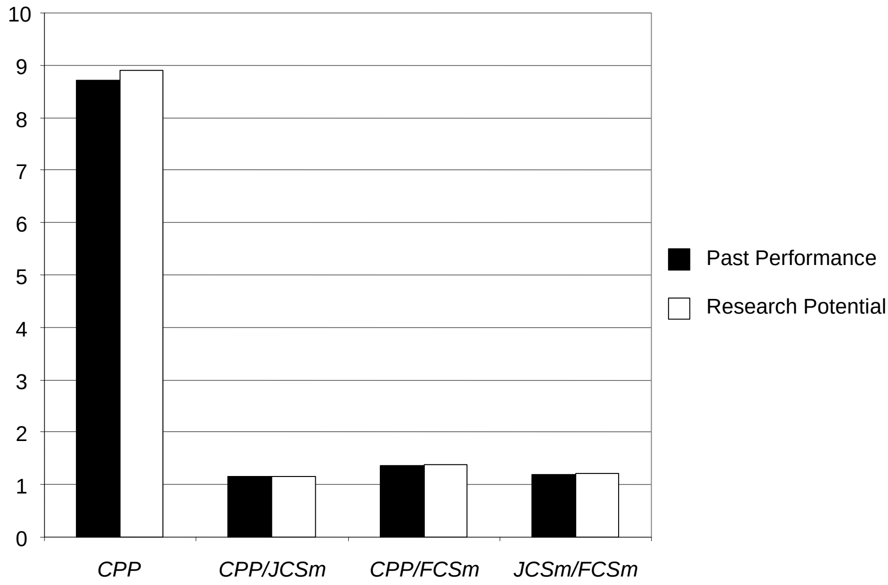

Figure 10. Impact score comparison for the 'past performance' and the 'current research potential' approach modalities at the Faculty of Applied Sciences, Delft University of Technology, 1989–1998

review, that is, qualitative analyses of the research performance.

Case 4: Field delineation in bibliometric studies

A crucial element in bibliometric analysis is the problem of field delineation. We discussed this problem already in relation to the interpretation of research profiles *vis-à-vis* the structure of the research organization involved. Field delineation can, however, play an important role in a more general manner, particularly in a top-down analysis in a certain field. In this case, the focus is on eye and vision research, or ophthalmology.

Definition or 'delineating' of research fields can be done by several methods. A quite common method is to define a research field by combining all publications in a specific area, for instance by ISI journal subject categories (that is, all publications of a large set of field-specific journals), or by selecting from a disciplinary database all the publications labelled under a specific classification code.

A second method consists of the delineation of a research field by focusing on the research area-related address-words attached to scientific publications. Publications contain addresses which indicate the organizational units the publications come from, for instance, the Department of Rheumatology within the address of an academic hospital.

In this part of our study, we compare both methods, and discuss the differences particularly in relation to the calculation of bibliometric indicators used in research performance assessments.

In this case, we define the medical field ophthalmology along the two above lines. First, the field is defined by taking all publications (articles, letters, notes and reviews) appearing in journals in the period 1999–2002, classified by ISI in the journal subject category ophthalmology. This results in a dataset of 21,862 publications. In the second case, the field is defined by collecting all publications (again the document types: articles, letters, notes and reviews) in the ISI database in the period 1999–2002 containing in the address field words related to ophthalmologic research ("OPH", "OPTH", "AUGEN", etc). This resulted in a dataset of 22,107 publications in the period 1999–2002. By comparing both sets, we find that there is an overlap of about 70% between them.

Table 2 shows clearly that the overlap between both fields is relatively large, around 69%. However, this still leaves a substantial number of publications outside the analysis when one decides to delineate a research field along one of the above lines. For instance, if one chooses a journal category based data-collection procedure - as was done in this case study for ophthalmology — one risks neglecting 6,982 relevant publications. On the other hand, using a definition based on field-specific address words, in the case of ophthalmology one misses 6,737 relevant publications in the analysis.

Next, if we look at the research profiles created with both definition methods, the situation becomes even more 'alarming'. Figure 10 shows the results of

Table 2. Results from data collection according to category or address

Field definition based on		N	A and B (N = 15,125)	A not B (N = 6,737)	B not A (N = 6,982)
Journal category	(A)	21,862	69.2%	30.8%	–
Field-specific address	(B)	22,107	68.4%	–	31.6%

the research profile analysis based on the journal category approach. One clearly observes the (expected) domination of the profile by only one field, namely the field on which the whole data-collection procedure was based, ophthalmology. The appearance of other fields is due to double classification of journals, in several other journal subject categories.

Thus, we can explain the strong appearance of ophthalmology in this research profile (90% of the publications involved), in combination with surgery (5% of the publications). The impact of the publications involved is around worldwide average level (*CPP/FCSm* = 1.04), while all other fields involved display a low impact.

Figure 11 displays the results of the data-collection procedure based on research-area-related address-words. Remarkably, now the field ophthalmology covers only 60% of the output, again with an impact score around the worldwide average (*CPP/FCSm* = 1.05). We observe fields related with eye and vision research are: neurosciences (5% of the output, combined with a *CPP/FCSm* score of 0.95), surgery (4% of the output, *CPP/FCSm* = 0.94), biochemistry and molecular biology (3% of the output, *CPP/FCSm* = 1.06). The only field in this profile with a high impact score is genetics and heredity (2% of the output and *CPP/FCSm* = 1.19). Not only is the actual share of the field ophthalmology considerably smaller than in the other data-collection procedure, but also the impact of the publications related to neurosciences is much higher. Finally, more fields play a significant role in the profile based on address-words.

These two profiles show a specific characteristic that bibliometricians encounter quite frequently, if not nearly always. This is the fact that the profile in Figure 10 displays the core of the field, in terms of the journals that are most often used by scientists working in the field (as can be concluded from the large overlap between the two sets of publications resulting from the two methods), while the research profile in Figure 11 indicates that scientists always publish their research in several different journals, which do not necessarily belong to one and the same

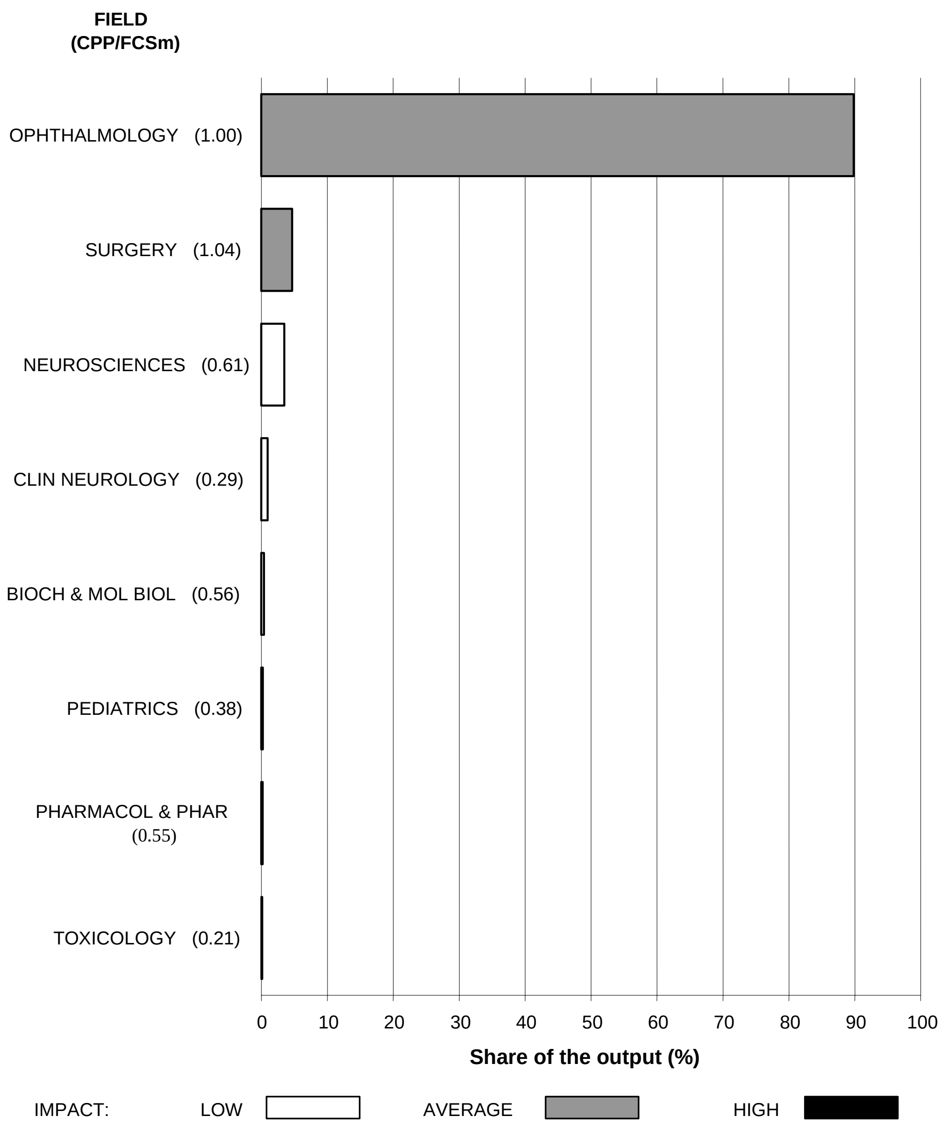

Figure 11. Research profile: output and impact per field, 1999–2002, ophthalmology defined by subject category

field. So in a way, the combination of the two methods underlines the degree of interdisciplinary research conducted in a field.

So if one wishes to use this type of approach for the purpose of gaining insight into volume, impact and internal structure of a specific field, one might want to choose a combination of the methods described above. Otherwise, a serious risk exists that the field that one is interested in is covered only partially in terms of the volume, with respect to the impact, might leave out related research with substantial higher impact scores, and in terms of the underlying internal field structure, leaves out important other fields. A previous study in which these approaches were used was related to German environmental medicine (van Raan *et al*, 2001). One might even wish to extend the data constituting the field with publications collected from disciplinary databases, such as *Medline* (for medicine), *Inspec* (for physics and related fields), *Chemical Abstracts* (for chemistry and chemical engineering), using from these systems their key words or classification terms, to further enlarge the data-sets used for the calculation of bibliometric scores that can be used in research performance analyses.

Conclusions

In this study, a model was developed along the lines of which bibliometric research performance assessments could be carried out. The model ("bibliometric research performance assessment model") presented in this study distinguishes between functions and goals of bibliometric studies. While the functionality is expressed by the terms *evaluative* and *descriptive*, the goals of these types of studies are presented by either the *past* or the *future*. The description of the model clearly indicates that the distinction between evaluative and descriptive bibliometrics is a fruitful

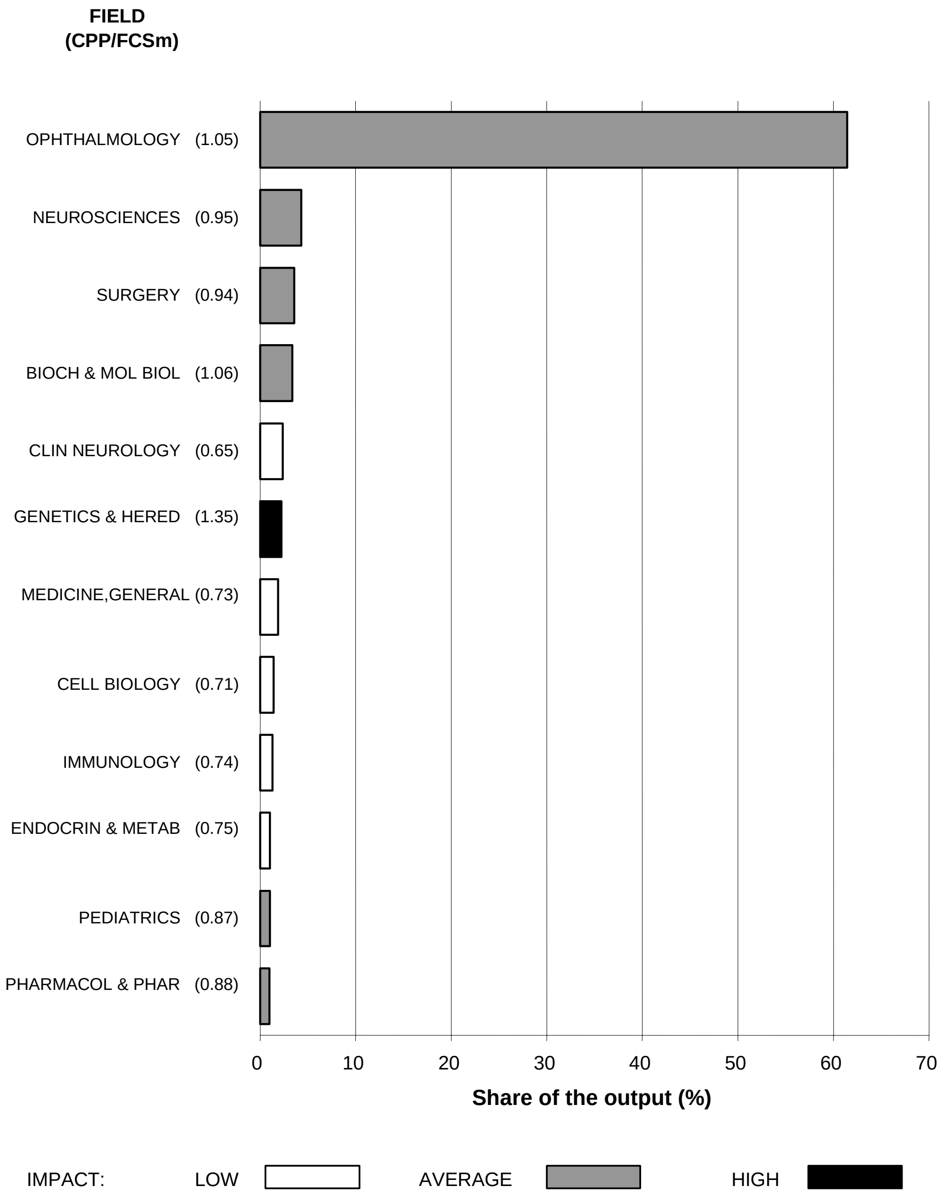

Figure 12. Research profile: output and impact per field, 1999–2002, ophthalmology defined by address

one. While bottom-up studies can have an evaluative character, due to the fact that the target-actors are strongly involved in both the reliability of the data and the 'credibility' of the applied method, top-down studies (both on the national level as well as on the level of organizations) remain in the realm of descriptive bibliometrics. In general, top-down studies lack the necessary verification and controls 'at the work floor' to allow conclusions in the framework of research performance assessment on lower levels of aggregation. An evaluative bottom-up study does provide such conclusions, particularly due to the inclusion of a verification process.

An important issue related to the discussed bibliometric research performance assessment model, its understanding, and the interpretation of bibliometric analyses in general, is the role of the person, the bibliometrician. A bibliometric researcher can and should inform the users of bibliometric data on the advantages and disadvantages of this type of study, through dialog. By answering questions regarding the function and goal of a proposed bibliometric study, the bibliometrician can guide the initiator of any bibliometric study in the direction that leads to the application of the most appropriate approach, and the related techniques.

References

NOWT 2000. Report of the Netherlands Observatory of Science and Technology.

van Leeuwen, T N 2002. The Research Profile of the Delft University of Technology. Research report to the University Board.

van Raan, A F J, M S Visser and T N van Leeuwen 2001. Bibliometrics – monitoring emerging fields. A bibliometric methodology for exploring interdisciplinary, 'unorthodox' fields of science. A case study of environmental medicine. In *Science Studies. Probing the Dynamics of Scientific Knowledge*, S Maasen and M Winterhager eds., pp. 87–123. Bielefeld: transcript Verlag.