WILEY InterScience
DISCOVER SOMETHING GREAT

Statistical Properties of Bibliometric Indicators: Research Group Indicator Distributions and Correlations

Anthony F.J. van Raan
Centre for Science and Technology Studies, Leiden University, Wassenaarseweg 52, P.O. Box 9555, 2300 RB Leiden, The Netherlands. E-mail: vanraan@cwts.leidenuniv.nl

In this article we present an empirical approach to the study of the statistical properties of bibliometric indicators on a very relevant but not simply "available" aggregation level: the research group. We focus on the distribution functions of a coherent set of indicators that are used frequently in the analysis of research performance. In this sense, the coherent set of indicators acts as a measuring instrument. Better insight into the statistical properties of a measuring instrument is necessary to enable assessment of the instrument itself. The most basic distribution in bibliometric analysis is the distribution of citations over publications, and this distribution is very skewed. Nevertheless, we clearly observe the working of the central limit theorem and find that at the level of research groups the distribution functions of the main indicators, particularly the journal-normalized and the field-normalized indicators, approach normal distributions. The results of our study underline the importance of the idea of "group oeuvre," that is, the role of sets of related publications as a unit of analysis.

Introduction

An appropriate application of bibliometric indicators in the evaluation of scientific research requires the consideration of the statistical properties of these indicators (Schubert & Glänzel, 1983). In particular, the uneven, skewed character of the most basic distribution, citations over publications, needs further study. For earlier work on scientometric distribution functions we refer to the studies of Lotka (1926), Price (1965), Naranan (1971), Haitun (1982), Redner (1998), and Simon (1955), and to those of Laherrère and Sornette (1998) and Tsallis and de Albuquerque (2000) for general discussions of skew distribution functions.

Seglen (1992, 1994) studied the relationship between article "citedness" and journal impact on the basis of the work of 16 Norwegian biomedical researchers at one institute. He concluded that the use of journal impact as an indicator for research performance evaluation is inappropriate as the skewed distributions result in poor correlations between article citedness and journal impact. In his work, journal impact was restricted exclusively to the ISI[1] journal impact factor; other types of journal impact indicators were not considered. Careful analysis of the relations between specific indicators, such as article citedness and journal impact, is a crucial part of the bibliometric methodology (Moed & Van Leeuwen, 1995, 1996). In this study on statistical properties of bibliometric indicators we focus on distribution functions of the individual indicators, including a more appropriate indicator for journal impact at a specific but important aggregation level, the research group.

The crucial point in the discussion of bibliometric research-performance measures is the skewness of distributions. For instance, only a small fraction of articles in a journal are cited near the average citation rate of a journal as a whole. Therefore, citation averages, and not only those of journals, seem to be inappropriate, on statistical grounds alone, as an element in the construction of indicators. However, a more differentiated approach is necessary as the targets of bibliometric measurements are often larger entities and thus the statistical properties are different from those of the basic elements such as individual publications or small sets of publications, as in the case of individual researchers.

Katz (1999, 2000) discussed scaling relationships between number of citations and number of publications across research fields, institutes, and countries. The scientific community is characterized by the Matthew effect (cumulative advantage; Merton, 1968, 1988), which implies a nonlinear increase of impact with increasing size, demonstrated by the finding that the number of citations as a function of number

Received November 22, 2004; revised January 11, 2005; accepted February 17, 2005

 ● Published online 14 December 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asi.20284

[1]The former Institute for Scientific Information (ISI) in Philadelphia, now Thomson Scientific, is the producer and publisher of the Science Citation Index, the Social Science Citation Index, the Arts & Humanities Citation Index, and the "specialty" citation indexes (CompuMath, Biochemistry and Biophysics, Biotechnology, Chemistry, Material Science, Neurosciences). Throughout this article we use the term CI (Citation Index) for the set of databases.

of publications (measured for 152 (sub)fields[2] of science) exhibits a power law dependence with an exponent larger than 1. Katz supposes that these scaling relationships probably exist across smaller entities such as research groups. In this article we show that this is indeed the case (see Fundamental Distribution Functions and Application of the Central Limit Theorem). Katz argues that the "conventional" bibliometric indicators may fail to account for this nonlinearity between size—measured by number of publications—and impact—measured by number of citations—and could result in an over- or underestimation of research performance. In other words, smaller groups have a smaller propensity than larger groups to be cited. Therefore, "conventional" indicators should be supplemented with measures that are corrected for the specified scaling relationships (Katz, 2005).

Katz, however, also indicates that size is often also an indicator of performance, in the sense that successful groups are able to attract more funds for research and thus grow. At the same time, they can develop a larger and more differentiated research program with strongly related themes, which very well may reinforce further their scientific performance and influence. Thus one could also argue that a larger impact as measured on the basis of citations cannot be simply waved aside as purely a scale-dependent effect. In this way groups are "punished" growing because the number of citations received by them should be corrected for size. Furthermore, groups with a larger number of publications may suffer from a possible scenario in which to an increasing extent researchers are making more and more choices within the complete oeuvre of the group to give citations, leading to a decreasing number of *citations per publication* for larger groups.

Thus, a crucial question is whether indicators that are more complex than simple citation counts—such as our field-specific normalized indicator crown indicator *CPP/FCSm* (see Chemistry Research Over a Period of 10 Years)—still exhibit "cumulative advantage" scaling behavior. We show in this article that such is not the case.

In addition to their application as a tool for research performance evaluation, bibliometric indicators can be applied as an instrument for investigating characteristics of science as a knowledge-generating and communication system. Thus, other important questions arise: To what extent are the empirically found distributions and other statistical characteristics a reflection of certain properties of the scientific communication and reward system? What is their relation to crucial aspects of the science system? For example, how do scientists disseminate their findings as optimally as possible? How will these knowledge flows work given general network properties (e.g., coauthor networks; citation networks, see, for instance, van Raan, 2005b)? How are new ideas accepted and valued by colleagues, how do they develop, and what is the role of citations (references) in these processes? The role of references is probably a more multidimensional characteristic, ranging from recognition to "utility" or persuasion. An interesting recent finding is that scientists refer less to "authoritative" articles in the case of smaller total numbers of references (Moed & Garfield, 2004); from this observation these authors conclude that citing such articles is not a major motivation of an author. As these authoritative articles are often highly cited, the citation behavior described may cause a mitigation of the accumulation of citations to already highly cited articles (Matthew effect) and thus influence citation distribution functions.

In this article we present empirical results on the statistical properties of the standard bibliometric indicators developed and applied by our institute on the most important level in the scientific enterprise: the research group. In this sense, we present unique material, as the research group is not an entity directly available in databases such as authors or journals. Research groups are defined by the internal structure of universities, research institutions, and research and development (R&D) laboratories of companies. We use the results of two large evaluation studies: (1) all university chemistry groups in the Netherlands, for the period 1991–2000 in a total of 157 research groups, about 700 senior researchers with about 18,000 publications (Citation Index–[CI]-based) and 175,000 citations (excluding self-citations) to these publications, and (2) all 65 research groups of the Faculty of Medicine, Leiden University (Leiden University Medical Center, LUMC) in the period 1990–2001, about 400 senior researchers with around 10,000 publications and 185,000 citations. Both sets represent a volume of researchers, publications, and citations of more than one order of magnitude larger when compared to the Seglen study (Seglen, 2004), which involved 16 researchers and 907 publications (in journals to which an impact factor could be assigned).

In this article we apply several quantitative methods (Egghe & Rousseau, 1990; Newbold, 1995) to unravel the statistical properties of bibliometric indicators. The structure of this article is as follows: In the second section we discuss the data material for two sets of research groups, the application of the bibliometric method, and the calculation of the indicators. The third section addresses the statistical analysis and discusses the results of the analysis. The fourth section summarizes the main outcomes of this study.

Data Material, the Bibliometric Method, and Calculation of the Indicators

The Two Data Sets

As discussed in the previous section, we studied the statistics of bibliometric indicators on the basis of two large sets of publications, one concerning all chemistry research in a country (Netherlands) for a 10-year period, and the second concerning all research groups in a large medical

[2]We here use the definition of (sub)fields based on a classification of scientific journals into *categories* developed by ISI/Thomson Scientific. Although this classification is not perfect, it provides a clear and "fixed" consistent field definition suitable for automated procedures within our data system.

DOI: 10.1002/asi

institution (Leiden) for a period of 12 years. This material is quite unique, as to our knowledge no such compilations of very accurately verified publication sets on a large scale have been used for statistical analysis of the characteristics of the indicators at the research group level.

We stress again that the research group is the most important "work floor entity" in science. However, data at the research group level are not a trivial matter because externally stored information (such as the CI data on author names, addresses, journals, fields, citations, etc.) has to be combined carefully with internally stored data, i.e., data only available from the institutions that are the target of the bibliometric analysis. In other words, there are no data on actual research groups available externally as there are externally available data on individual scientists. The only way to study the bibliometric characteristics of research groups would be to use the address information within the main organization, for instance, "Department of Biochemistry" of a specific university. However, the delineation of departments or university groups through externally available data such as the address information in the CI databases is very problematic. We refer for a thorough discussion of this problem to Van Raan (2005a). As indicated, the data used in this study are the results of evaluation studies and are therefore based on data acquisition with strict verification procedures.

Chemistry Research Over a Period of 10 Years

The first set concerns all publications (those published in journals covered by the Citation Index, "CI publications") of all university research groups in chemistry and chemical engineering in the Netherlands (NL). Thus, publications such as reports and books or book chapters are not taken into account. However, for chemistry research groups the focus on articles published in CI-covered journals generally provides a very good representation of the scientific output (VSNU, 2002). These ("CI") publications were collected as part of a large evaluation study conducted by the Association of Universities in the Netherlands; for a detailed discussion of the evaluation procedure and the results, refer to the evaluation report (VSNU, 2002). In the framework of this evaluation study, we performed an extensive bibliometric analysis to support the international peer committee (van Leeuwen, Visser, Moed, & Nederhof, 2002). An executive summary of the bibliometric results is included in the evaluation report. The period covered is 1991–2000 for both publications and the citation received by these publications. In total, the analysis covers about 18,000 publications and about 240,000 citations of 157 chemistry groups. We applied the Center for Science and Technology Studies (CWTS) standard bibliometric indicators. Here only "external" citations, i.e., citations corrected for self-citations, are taken into account. An overview of these indicators in given in the list in this section. In particular, we call attention to the definition of our journal impact indicator, the journal citation score, *JCS*. For a detailed discussion we refer to Van Raan (1996, 2004).

The following are the CWTS Standard Bibliometric Indicators:

- Number of publications (*P*) in CI-covered journals of the research group in the entire period.
- Number of citations received by *P* during the entire period, with and without self-citations (*Ci* and *C*).
- Average number of citations per publication, again with and without self-citations (*CPPi* and *CPP*).
- Percentage of publications not cited (in the specified period), *Pnc*;
- Journal-based worldwide average impact as an international reference level for the research group (*JCS*, journal citation score), without self-citations (on this worldwide scale!); in the case of more than one journal we use the average *JCSm*; for the calculation of *JCSm* the same publication and citation counting procedure, time windows, and article types are used as in the case of *CPP*.
- Field-based worldwide average impact as an international reference level for the research group (*FCS*, field citation score), without self-citations (on this worldwide scale); in the case of more than one field (as almost always) we use the average *FCSm*; for the calculation of *FCSm* the same publication and citation counting procedure, time windows, and article types are used as in the case of *CPP*.
- Comparison of the actual international impact of the research group with the worldwide average based on *JCSm* as a standard, without self-citations, indicator *CPP*/*JCSm*.
- Comparison of the actual international impact of the research group with the worldwide average based on *FCSm* as a standard, without self-citations, indicator *CPP*/*FCSm*.
- Ratio *JCSm*/*FCSm* as journal-level indicator: i.e., is the research group publishing in top or in subtop (in terms of "citedness") journals?
- Percentage of self-citations of the research group, *SelfCit*.

The indicators are calculated on the basis of the *total block analysis*, which means that publications are counted for the entire 10-year period 1991–2000 and citations are counted up to and including 2000 (e.g., for publications from 1991, citations are counted in the period 1991–2000, and for publications from 2000, citations are counted only in 2000). The universities covered by this evaluation study are Leiden, Utrecht, Groningen, Amsterdam UvA, Amsterdam VU, Nijmegen, Delft, Eindhoven, Enschede (Twente), and Wageningen. All fields of chemistry were covered by this set of university groups; the main fields were analytical chemistry, spectroscopy and microscopy, computational and theoretical chemistry, physical chemistry, catalysis, inorganic chemistry, organic and bioorganic chemistry, biochemistry, microbiology, biochemical engineering, polymer science and technology, materials science, and chemical engineering.

Medical Research Over a Period of 12 Years

The second set concerns all publications (again in journals covered by the Citation Index, "CI publications") of all research groups in the Leiden University Medical Center (LUMC). Also in the case of medical research groups, the focus on articles published in CI-covered journals generally

DOI: 10.1002/asi

provides a very good representation of the scientific output. These publications were collected as part of an internal Leiden evaluation study. In the framework of this evaluation study, we performed a detailed bibliometric analysis to support the LUMC research committee. Details of the bibliometric results are available from the author of this article. The period covered is 1990–2001 for both publications and citations received by these publications. The citation counting procedure is the same as for the chemistry groups. In total, the analysis covers about 10,000 publications and about 185,000 citations of 65 medical groups. The LUMC is a large clinical and basic research organization with a high international reputation. Practically all fields of medical research are represented, ranging from molecular cell biology to oncological surgery, and from organ transplantation to T-cell immune response research.

Basic Results of the Bibliometric Analyses

In Table 1 we show as an example the results of our bibliometric analysis for the most important indicators for all 12 chemistry research groups of one of the 10 universities (Univ A). The quality judgment of the international peer committee is also indicated. The peers used a 3-point scale to judge the research quality of a group: grade 5 is excellent, grade 4 is good, and grade 3 is satisfactory (VSNU, 2002).

Table 1 makes clear that our indicator calculations allow a statistical analysis of these indicators for the entire set of research groups (i.e., the groups of all 10 universities in the Netherlands covered by the VSNU evaluation study). In a follow-up study (van Raan, in press) we will address the correlation between the bibliometric indicators and the quality judgments of peers.

We applied the same CWTS standard bibliometric indicators used for the chemistry research to the medical research groups. An example of the results (first 10 groups) is presented in Table 2. Thus, the results of both cases are based on a strictly consistent methodology and are directly comparable. The LUMC (medical) case, however, was based on a 12-year period, whereas the chemistry case was based on a 10-year period. In this medical case no peer review committee was involved. We added two further standard indicators, the percentage of not-cited publications and the percentage of self-citations.

First we focus on the results for chemistry groups, then on the results for the medical research groups in comparison with those of the chemistry groups. Our analysis deals with eight bibliometric indicators: *P*, *C*, *CPP*, *JCSm*, *FCSm*, *CPP/JCSm*, *CPP/FCSm*, and *JCSm/FCSm*.

The set of chemistry groups and the set of medical groups differ in some important aspects. The chemistry groups are from 10 different universities, they have grown more or less "naturally," and they are not subject to one specific research policy strategy, as all 10 universities have their own priorities. The medical groups, however, are all within one large institution. They are subject to one and the same research policy and

TABLE 1. Example of the results of the bibliometric analysis for the chemistry groups.

Research group	*P*	*C*	*CPP*	*JCSm*	*FCSm*	*CPP/JCSm*	*CPP/FCSm*	*JCSm/FCSm*	Quality
Univ A, 01	92	554	6.02	5.76	4.33	1.05	1.39	1.33	5
Univ A, 02	69	536	7.77	5.12	2.98	1.52	2.61	1.72	4
Univ A, 03	129	3780	29.30	17.20	11.86	1.70	2.47	1.45	5
Univ A, 04	80	725	9.06	8.06	6.25	1.12	1.45	1.29	4
Univ A, 05	188	1488	7.91	8.76	5.31	0.90	1.49	1.65	5
Univ A, 06	52	424	8.15	6.27	3.56	1.30	2.29	1.76	4
Univ A, 07	52	362	6.96	4.51	5.01	1.54	1.39	0.90	3
Univ A, 08	171	1646	9.63	6.45	4.36	1.49	2.21	1.48	5
Univ A, 09	132	2581	19.55	15.22	11.71	1.28	1.67	1.30	4
Univ A, 10	119	2815	23.66	22.23	14.25	1.06	1.66	1.56	4
Univ A, 11	141	1630	11.56	17.83	12.30	0.65	0.94	1.45	4
Univ A, 12	102	1025	10.05	10.48	7.18	0.96	1.40	1.46	5

TABLE 2. Example of the results of the bibliometric analysis for the medical groups.

Research group	*P*	*C*	*CPP*	*JCSm*	*FCSm*	*CPP/JCSm*	*CPP/FCSm*	*JCSm/FCSm*	*Pnc*	*Scit*
LU 01	117	1,836	15.69	12.20	11.08	1.29	1.42	1.10	11%	20%
LU 02	197	3,587	18.21	14.28	14.75	1.28	1.23	0.97	11%	21%
LU 03	46	449	9.76	14.55	8.78	0.67	1.11	1.66	20%	23%
LU 04	560	16,906	30.19	25.22	15.29	1.20	1.97	1.65	10%	19%
LU 05	423	17,144	40.53	29.60	16.85	1.37	2.41	1.76	6%	21%
LU 06	369	13,454	36.46	30.34	17.54	1.20	2.08	1.73	6%	19%
LU 07	91	1,036	11.38	11.91	7.72	0.96	1.47	1.54	15%	22%
LU 08	95	554	5.83	6.52	5.80	0.89	1.01	1.13	22%	33%
LU 09	52	334	6.42	6.98	8.00	0.92	0.80	0.87	23%	33%
LU 10	512	5,729	11.19	8.70	6.44	1.29	1.74	1.35	22%	17%

DOI: 10.1002/asi

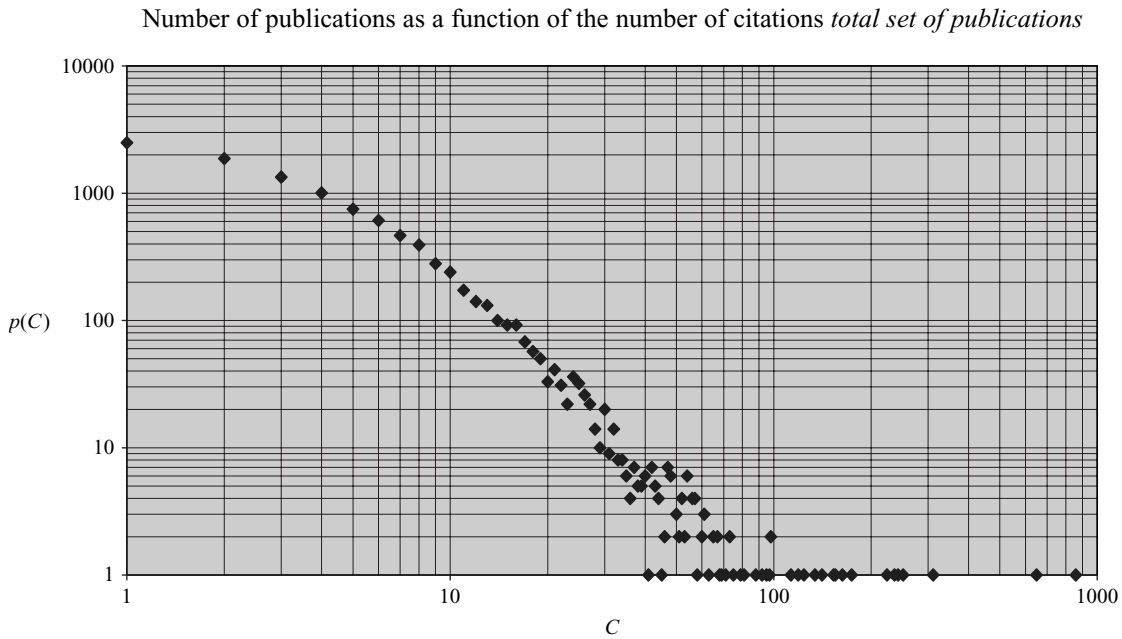


FIG. 1. Distribution function $p(C)$: number of publications as a function of number of citations for the total set of around 14,000 chemistry publications.

research management strategy; thus they are entities that can be considered as having a "natural" basis as a research group around one or two full professors, but at the same time they are "reconstructed" to a certain extent by the policy of the LUMC as a whole. We explain the observed differences of similar distribution functions for both sets in this context.

Statistical Analysis of Bibliometric Indicators

Fundamental Distribution Functions and Application of the Central Limit Theorem

First we present in Figure 1, as a starting point, the most basic distribution function: the number of publications as a function of the number of citations (regardless of research group, i.e., the distribution function of all "individual" publications). For this basic distribution function we use a *fixed citation window*: We take all publications of the 157 chemistry research groups together for the years 1991–1998 (p, about 14,000) and count the citations for a time window of 3 years starting with the publication year (i.e., for publications from 1991, citations are counted in the period 1991–1993, and for publications from 1998, citations are counted in the period 1998–2000). We use the symbol P for the total number of publications per group and p for the distribution functions based on the number of "individual" publications. We immediately observe that distribution function $p(C)$ follows a power law

$$p(C) = \alpha C^{s} \tag{1}$$

only for the higher-C tail of the distribution. The slope s of this part appears to be around -2.4 and thus we find a simple expression for the distribution function:

$$p(C) = \alpha C^{-2.4} \tag{2}$$

where α is a constant factor that can be established empirically from the data. Notice that in this logarithmic presentation we left out publications that had 0 citation; this number is 3,251, which is 23% of the total number of publications. As we can easily see with help of the figure, the number of publications with 0 citation would smoothly fit in a curve extrapolated from the low-C part of the distribution. Because of the very small number of publications with a very high number of citations, we also observe a "scattering" of data points on the horizontal axis.

It is very important to stress that from the empirical results presented in Figure 1 follows that the *non–power law* part of the distribution function covers the citation characteristics of nearly 12,000 publications (including publications with 0 citation). So in fact the power law only describes a small minority of the publications, namely, the more highly cited articles, about 2,000, i.e., 16%. We also refer for similar findings to Redner (1998), Laherrère and Sornette (1998), and Tsallis and de Albuquerque (2000).

Publications act as nodes in a citation network (Albert & Barabási, 2002), in which the links are the citations from other articles, so that the distribution of the number of citations represents the "in-degree" distribution of the network. In these network terms, the "scale-free" part (power law part of the distribution) of the citation network clearly accounts only for the "hubs." The appearance of a power law is at most an asymptotic behavior of the complex system of citation relations between publications. And even that is doubtful, as we showed with an *ab initio* citation distribution calculation based on a two-step competition model (van Raan, 2001a, 2001b) by which the number of publications with 0 citation is also predicted very well. Thus, in contrast to popular belief, basic bibliometric distribution functions, such as the number of publications as a function of number of citations, are not power-law functions for the entire data range and certainly not in the part that covers most entities (in this case, publications).

DOI: 10.1002/asi

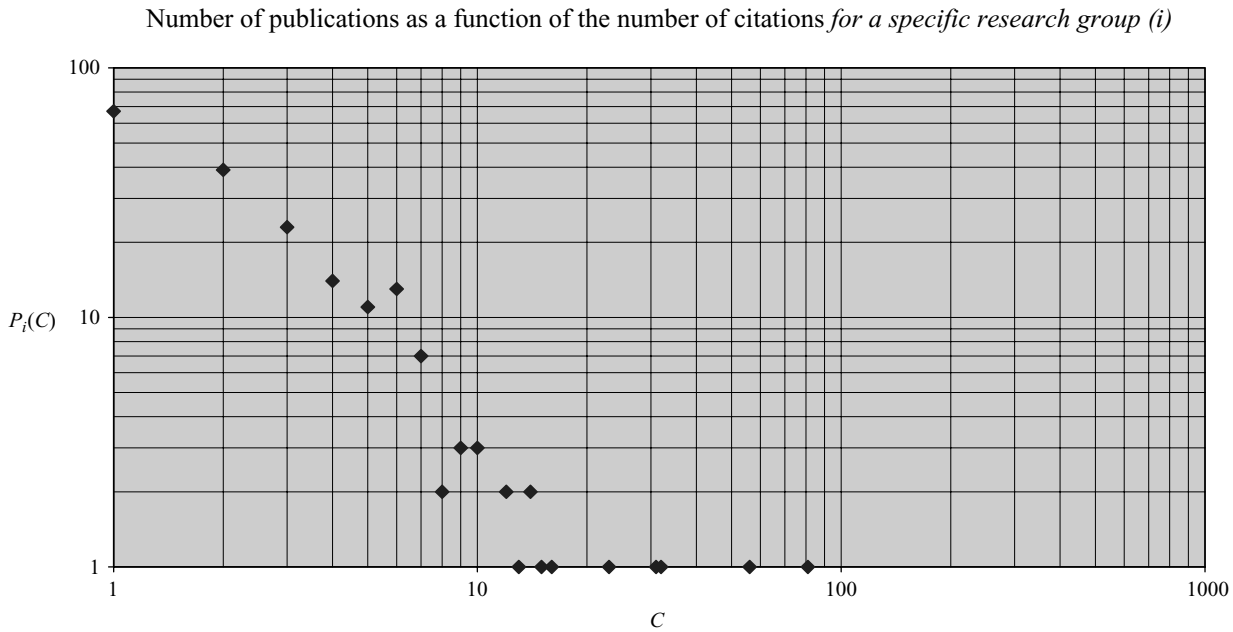


FIG. 2. Distribution function $P_i(C)$: number of publications within a research group (i) as a function of the number of citations.

The central limit theorem is particularly important for the practice of bibliometric analysis. Statistically, research groups can be considered as "samples" from the total population of publications. This total population of publications has a specific "income" distribution, i.e., the number of publications as a function of citations, $p(C)$, as presented in Figure 1. The "sample size" is given by the *total number* of publications of the group indicated by P (for a specific research group i, P_i is, for instance, 50), and thus the group can be characterized by its own distribution of the *individual publications within the group* given by an independent variable P_i (C). For N research groups we thus have a series of N independent variables:

$$P_1(C), P_2(C), \ldots, P_N(C).$$

An example of the distribution function for one of the 157 chemistry research groups (i) is given in Figure 2.

The mean number of citations per publication for each research group distribution is our indicator *CPP*. By virtue of the central limit theorem, the distribution of the *CPP* values of all research groups will tend to a normal distribution. In reality, research groups are certainly not ideal "samples" of the same size in terms of numbers of publications (P); their sizes show a skewed distribution. See Figure 3 for the number of groups as a function of number of publications, $G(P)$. Thus the distribution of the *CPP* indicators will be less skewed as compared to the basic $P(C)$ distribution, but it will still not resemble a normal distribution, as is clearly confirmed by the measurement of the *CPP* distribution for all chemistry research groups, $G(CPP)$; see Figure 4.

For comparison we also show the much more skewed distribution of the total number of citations (C) for all groups, $G(C)$; see Figure 5.

For further understanding of the statistical behavior of bibliometric indicators it is important to know how a specific collection of publications (namely, a research group) is characterized in terms of the *relation* between size (the total number of publications P of a specific research group[3]) and the total number of citations received by this group in a given period, C. This relation for all 157 chemistry research groups is presented in Figure 6. This figure shows us that this relation *on the aggregation level of research groups* is described with reasonable significance (coefficient of determination of the fitted regression is $R^2 = 0.69$) by a power law:

$$C(P) = 2.31P^{1.25} \tag{3}$$

and we observe that the size of groups leads to a "cumulative advantage" (with exponent +1.28) for the number of citations received by these groups. Thus we confirm the assumption of Katz (1999, 2000) discussed earlier that the Matthew effect also occurs in a sufficiently large set of research groups. We have already discussed that size-dependent "corrections" (on the basis of number of publications) of measured impact (on the basis of citations) will lead to a misleading leveling off of the impact indicators at the level of research groups as size has to be regarded as an intrinsic characteristic of performance.

We return to the application of the central limit theorem. Research groups are not samples from an entire population only in terms of citations. Publications are also characterized by the journal in which they appear, and particularly by the

[3]The number of publications is a valid measure of size in the statistical context described in this article. It is, however, a proxy for the "real size" of a research group in terms of number of staff full-time equivalents (FTEs) available for research. In the Appendix we present in Figure A1 the distribution function for research FTEs over the 157 chemistry research groups.

DOI: 10.1002/asi

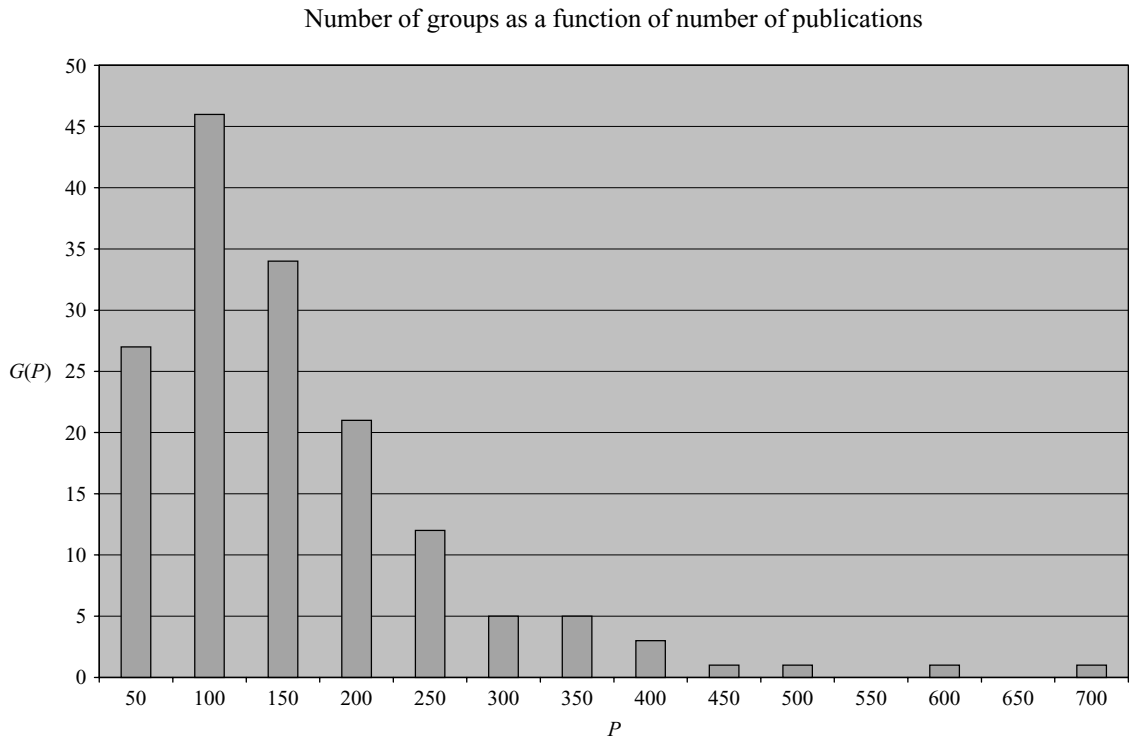


FIG. 3. Distribution function $G(P)$: number of chemistry groups as a function of number of publications (class width $\Delta P = 50$).

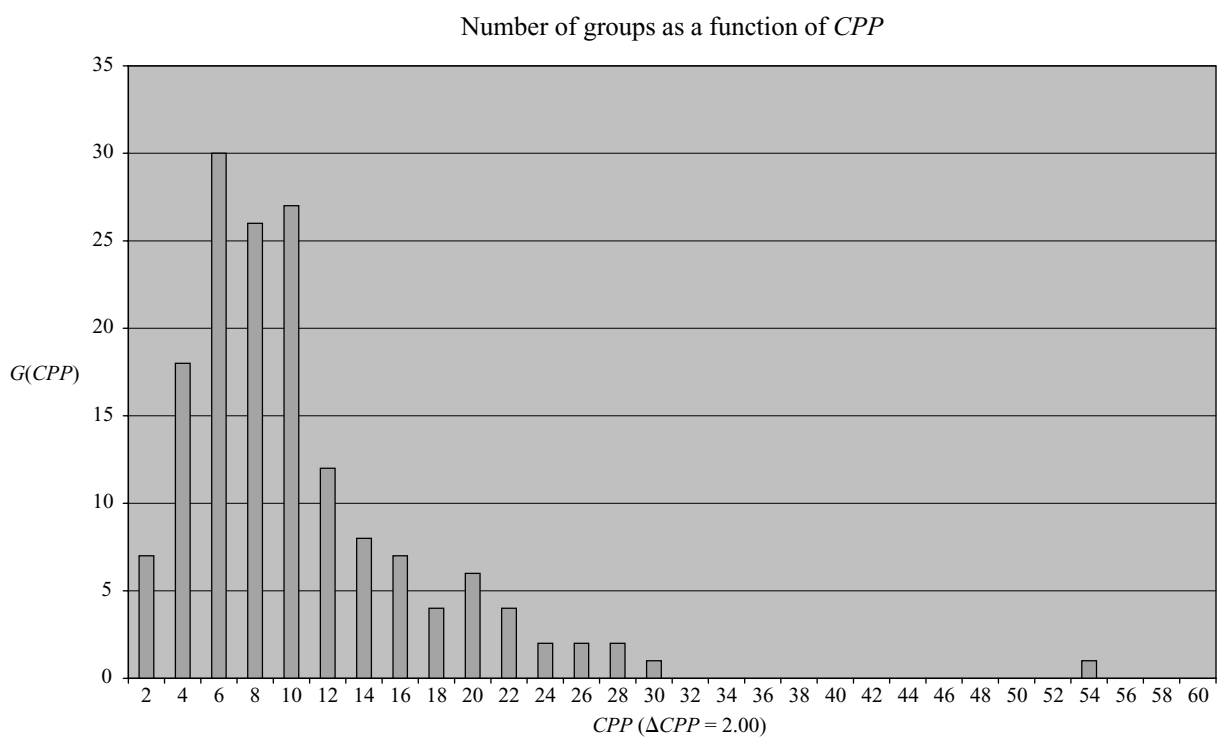


FIG. 4. Distribution function $G(CPP)$: number of chemistry groups as a function of CPP values (class width $\Delta CPP = 2.00$).

DOI: 10.1002/asi

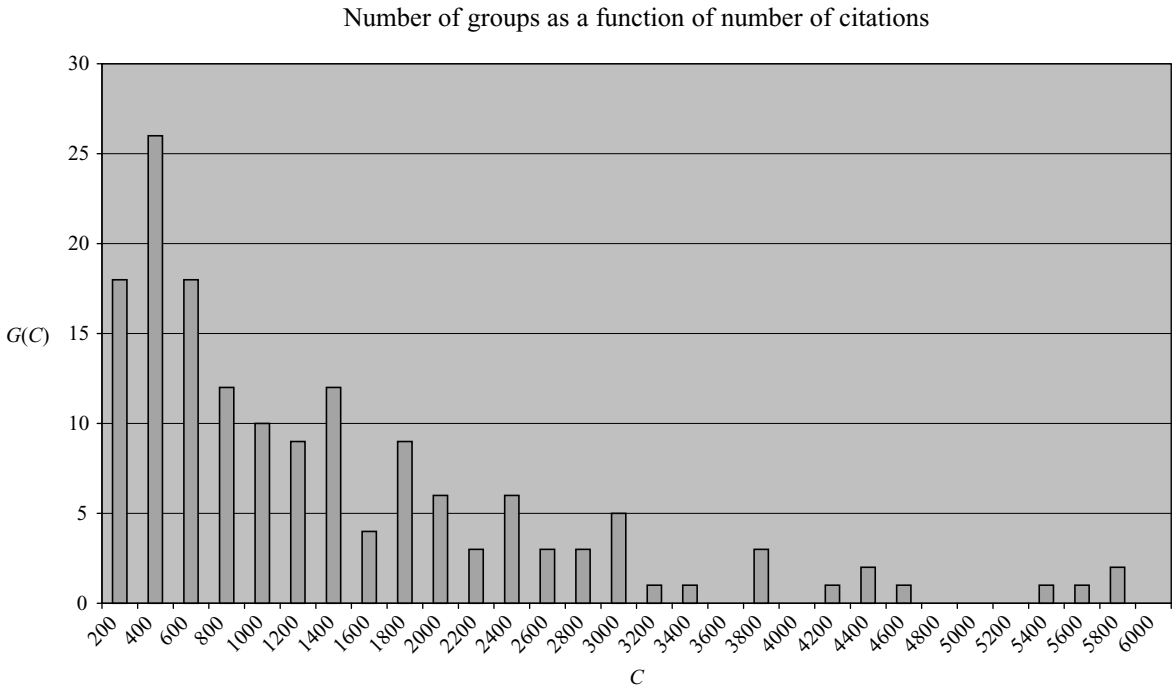


FIG. 5. Distribution function $G(C)$: number of chemistry groups as a function of total number of citations (C) (class width $\Delta C = 200$).

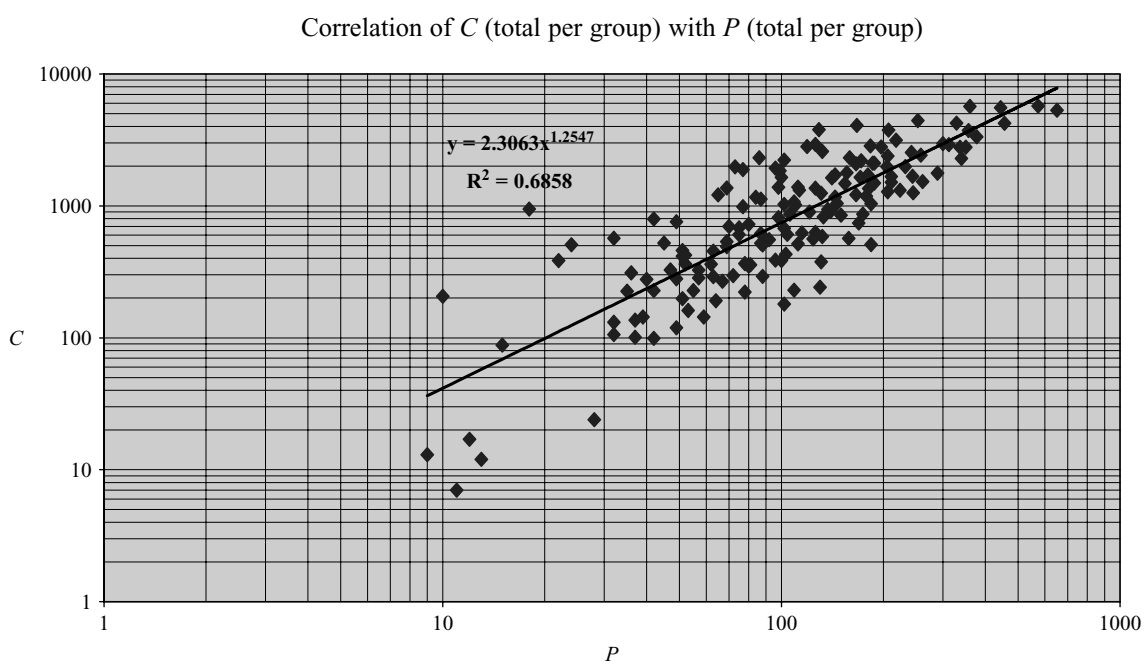
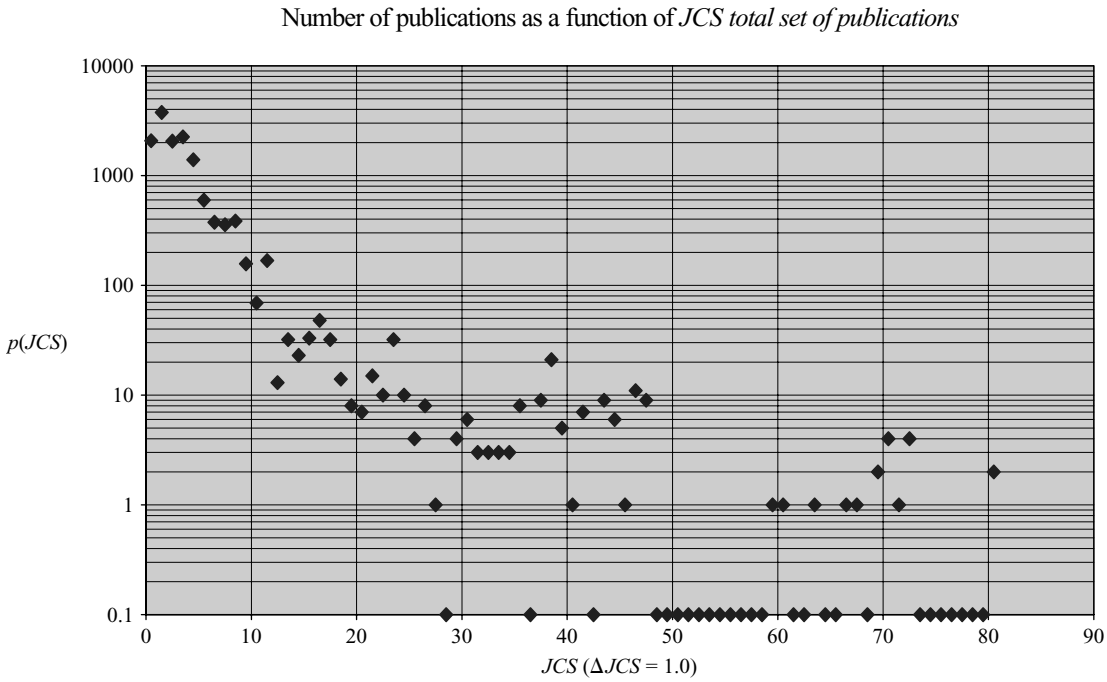


FIG. 6. Correlation of the number of citations (C) received per research group with the number of publications (P) of these groups, for all chemistry groups.

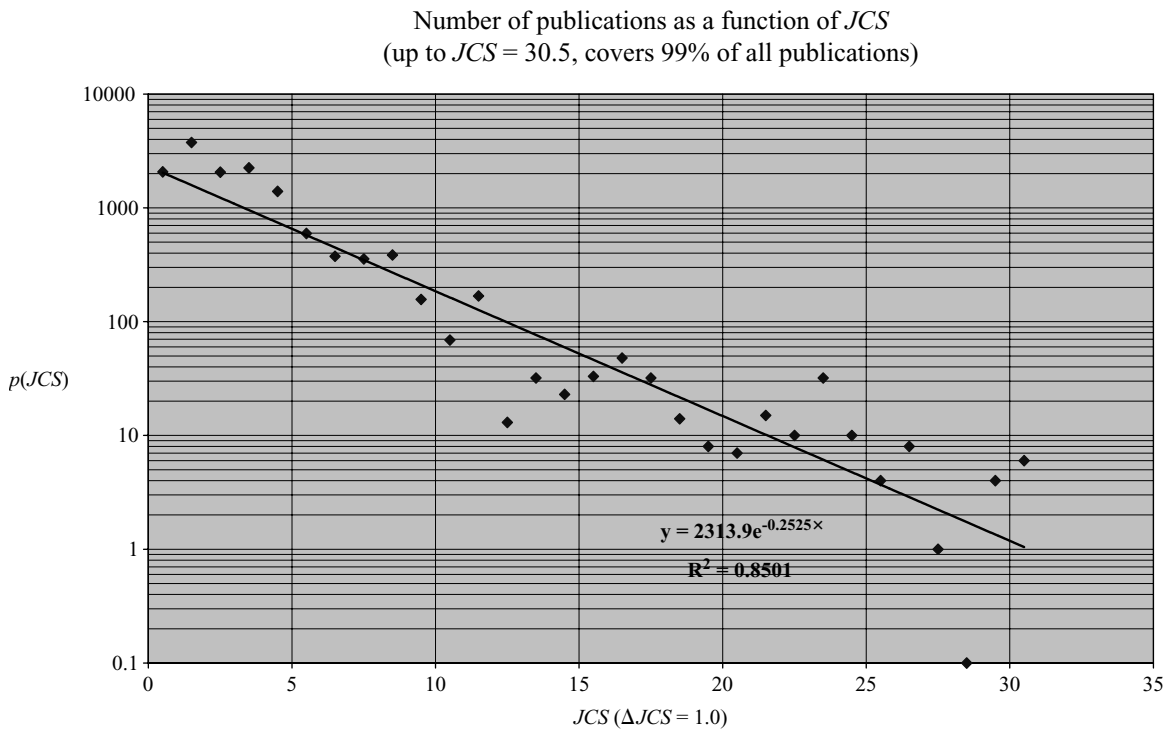
journal impact, our indicator *JCS*. In Figure 7a we present the distribution function for this variable (entire population): the number of publications as a function of (classes of) *JCS* values.

Given the logarithmic scale of the ordinate, we use for simplicity $p(JCS) = 0.1$ in order to include *JCS* values with 0 publication. Notice that in this case we have a semilogarithmic plot, with a linear abscissa. Only a smaller part of the entire publication population belongs to the very-high-value *JCS* classes (i.e., $JCS > 30.0$), making the distribution for this high-*JCS* part very noisy. If we restrict the analysis to the publications with values of $JCS \leq 30.5$, we cover 99% of the approximately 14,000 publications. The distribution function for these publications is shown in

DOI: 10.1002/asi



(a)



(b)

FIG. 7. (a) Distribution function $p(JCS)$: number of publications as a function of JCS (class width $\Delta\ JCS = 1.0$). (b) Same distribution function $p(JCS)$ as in Fig. 7a, now without the highest JCS values (class width $\Delta JCS = 1.0$).

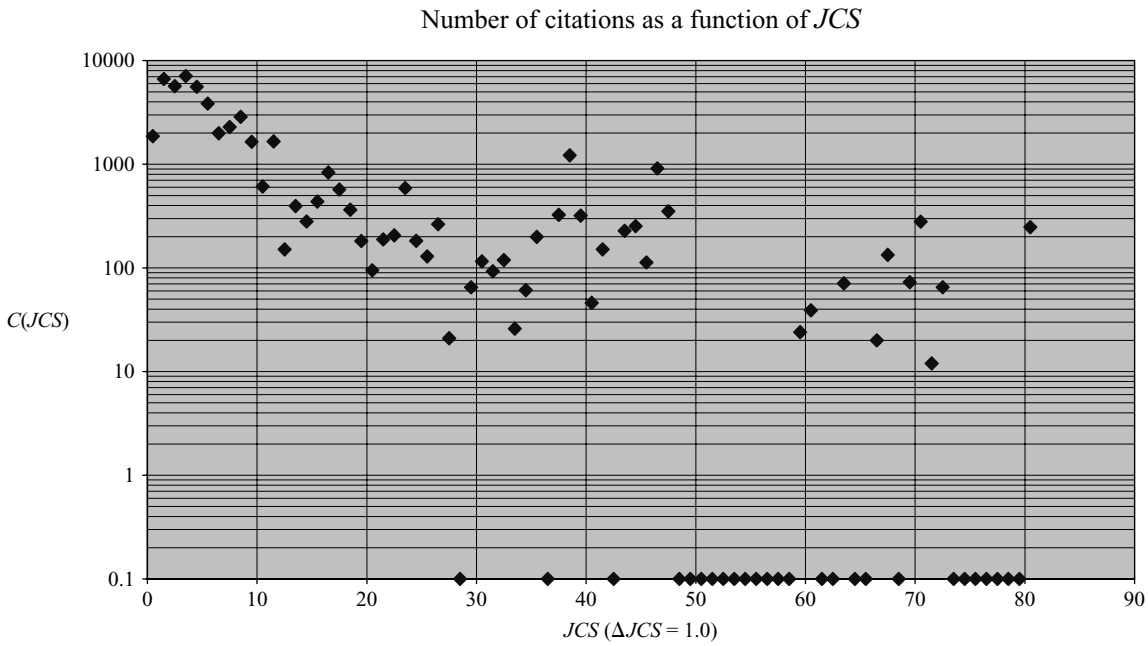
Figure 7b. We immediately observe a significant ($R^2 = 0.85$) exponential (non–power law) relation given by the simple equation (using the parameter x for JCS):

$$p(x) = \beta \exp[-0.25x] \tag{4}$$

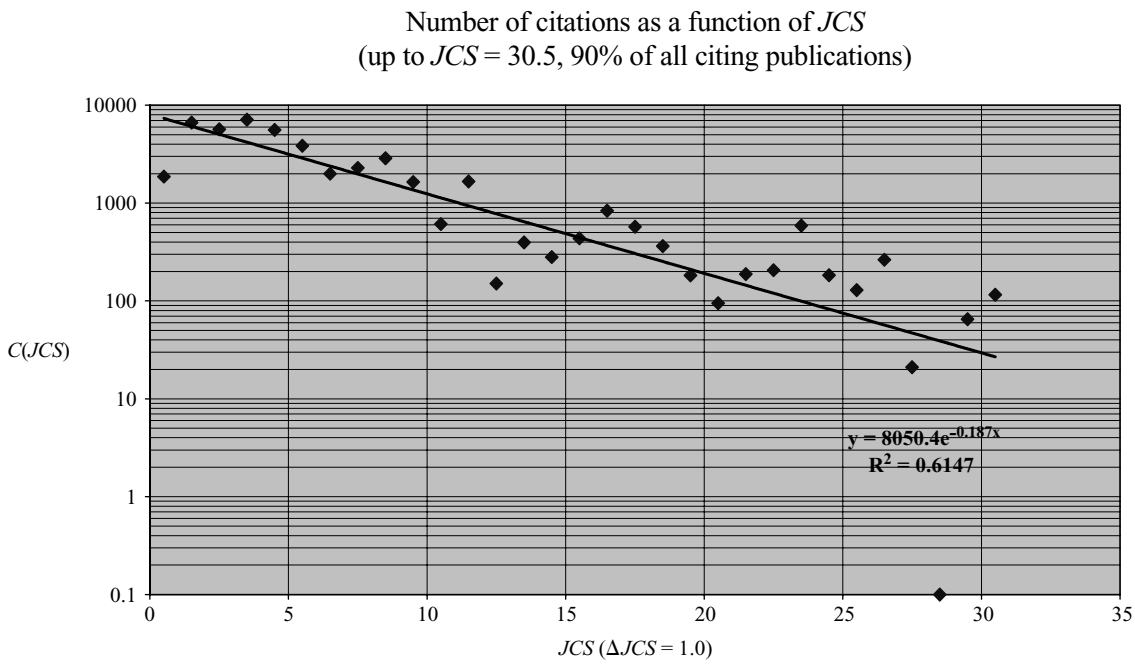
where β is a constant factor that can be determined empirically from the plot ($\beta = 2131.9$).

The next step is to analyze the relation between the number of citations and journal impact. One has to be very careful in defining what precisely is measured and which

DOI: 10.1002/asi



(a)



(b)

FIG. 8. (a) Number of citations received by NL (Netherlands) chemistry publications as a function of JCS (class width $\Delta JCS = 1.0$), similarly to Fig. 7a, we use for simplicity $C = 0.1$ in order to include *JCS* values with 0 citation. (b) Same distribution function as in Fig. 8a, now without the highest *JCS* values (class width $\Delta JCS = 1.0$).

indicators are used. Here our measurement concerns the number of citations (i.e., the number of citing publications) received by publications as a function of the *JCS* values of these cited publications. For example, if a publication in *Physical Review Letters* is cited 10 times (in the given period), then the *JCS* of *Physical Review Letters* gets a score of 10.

The results of this analysis for the entire population of chemistry publications are given in Figure 8a. In this figure we present the number of citations (C) as a function of (classes of) *JCS* values (again class width $\Delta JCS = 1.0$). Again we have a semilogarithmic plot, with a linear abscissa. And again only a smaller part of the distribution belongs to the very noisy high-value *JCS* classes. Similarly

DOI: 10.1002/asi

to the distribution function for the publications, we restrict the analysis to the publications with values of $JCS \leq 30.5$. We cover 90% of the approximately 52,000 citations. This part of the distribution function is given in Figure 8b. We again observe a reasonably significant ($R^2 = 0.61$) exponential (non–power law) relation given by the equation (using the parameter x for *JCS*):

$$C(x) = \gamma \exp[-0.19x] \tag{5}$$

where γ is a constant factor that can be determined empirically from the plot ($\gamma = 8050.4$).

We refer to Van Raan (2001a) for a thorough discussion of an *ab initio* model ("two-step competition") to relate the distribution functions on the basis of publications (p), citations (C), and journal impact (*JCS*) and to explain their remarkable differences, on the one hand, and similarities, on the other.

The preceding discussion makes clear that *JCS* is an important additional variable to characterize a population of publications. Research groups make their own choices for journals. Some try to publish as much as possible in top journals, and others, because of their specialization and often smaller target audience, mainly publish in journals with a considerably lower impact level. Examples of both are, on the one hand, breakthrough work in biochemistry and molecular biology and, on the other, application-oriented work in chemical engineering.

The preceding discussion indicates that in the context of the central limit theorem, the entire population is characterized not only by citations (C) but also by the variable *JCS*. In the previous section we have demonstrated that there is—with an exception of the high *JCS* values—a reasonably strong relation between these two variables. But this relation is not a simple proportionality, but an exponential relation. A good comparison is the relation between income and education level. We further illustrate this comparison in the following list:

- Families in a country: P, Publications in a large discipline
- Income of a family: C, Citations received by a publication
- Samples of families: Research groups
- Sample size of number of families in sample: Group size of number of publications in the group
- Average income of families in a sample: *CPP* of a research group
- Education level of families: *JCS* of publications
- Average education level of families in a sample: *JCSm* of a research group

Thus, in precisely the same way as in the case of *CPP*, we can again apply the central limit theorem to the distribution of *JCS*. In the entire population of publications, each of the publications is characterized by its own *JCS* value, and the research groups can be considered as samples of publications with an average *JCS* value, in our terminology *JCSm*. The distribution of *JCSm* over all research groups will tend to a normal distribution. As in the case of the *CPP* distribution, the *JCSm* distribution will be much less skewed than the *JCS* distribution of the entire population but not a perfect normal distribution given the size difference of the research groups. However, as the *JCS* values in the entire population are already a mean (namely, the mean value of the number of citations per article in a journal), the *JCSm* distribution must be closer to normal than in the case of the *CPP* distribution. These considerations are confirmed by the measurement of the *JCSm* distribution function; see Figure 9a.

Similar arguments also apply to the distribution function of the research group average *FCSm* of the *field-normalized* indicator *FCS*. As the *FCS* values in the entire population are even more an average than the *JCS* values (namely, the mean value of the number of citations per article in a journal, and then averaged over all journals in a field), the *FCSm* distribution will again be closer to normal than in the case of the *JCSm* distribution. The measured distribution function clearly confirms these arguments; see Figure 9b.

Distribution Functions for the Normalized Indicators

How can we find the distribution function in which "income" is normalized to "education level"? In order to tackle this problem, we analyzed the correlation between the average number of citations per publication (*CPP*) and the average *JCSm* values for all chemistry research groups. The results of this analysis are presented in Figure 10. We find that this relation *on the aggregation level of research groups* is described with reasonable significance ($R^2 = 0.75$) by the equation

$$CPP = 1.13\ JCSm^{0.97} \tag{6a}$$

which indicates that *CPP* at the aggregation level of a research group is related in a very simple, almost proportional manner to *JCSm*. This observation has interesting consequences. At the research group level we can expect for the distribution function $G(CPP/JCSm)$ of the value of the *journal*-normalized number of citations per publication, on the basis of Equation 6a and the coefficient $R^2 = 0.75$, in good approximation a normal distribution around a value given by

$$CPP/JCSm \sim 1.1 \tag{6b}$$

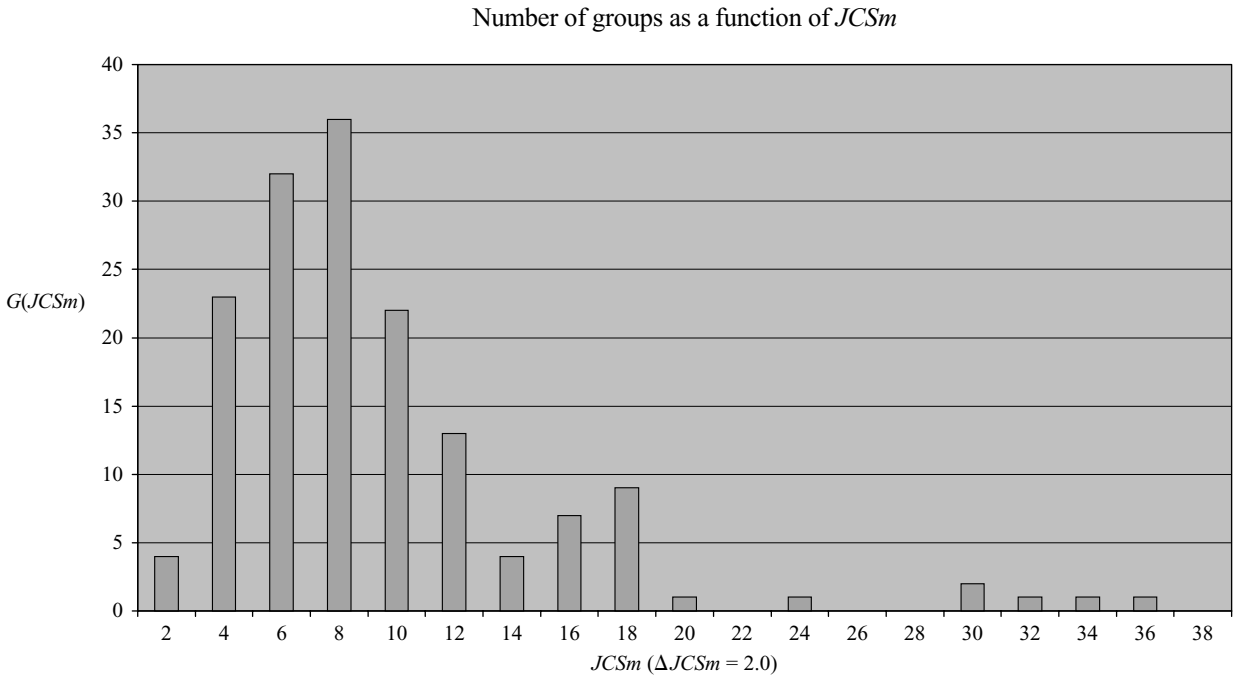
This finding is nicely confirmed by Figure 11. Here we show the distribution function of $CPP/JCSm$ for all chemistry research groups.

In the application of bibliometric indicators, in most cases we regard the *field*-normalized indicator $CPP/FCSm$ as our crown indicator, so we are particularly interested in the statistical properties of this indicator. We follow a similar approach by analyzing the correlation between the average number of citations per publication (*CPP*) and the average *FCSm* values for all chemistry research groups. The results of this analysis are presented in Figure 12.

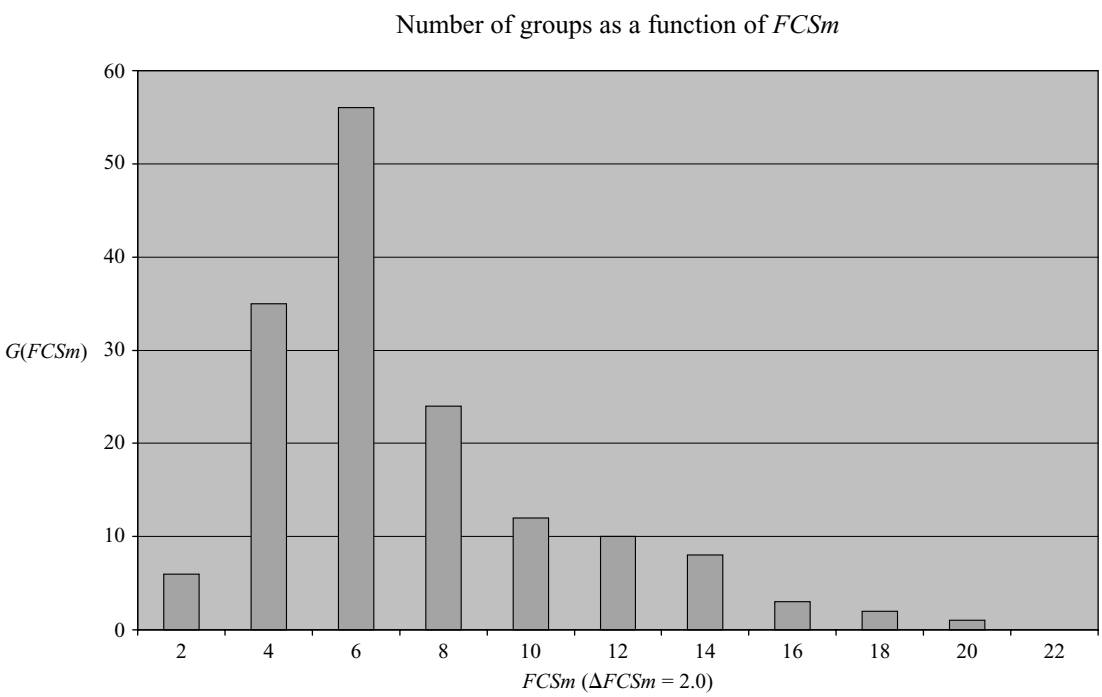
We find that this relation *on the aggregation level of research groups* is described (with less significance [$R^2 = 0.51$] than in the case of *JCSm*) by the equation

$$CPP = 1.53\ FCSm^{0.94} \tag{7a}$$

DOI: 10.1002/asi



(a)



(b)

FIG. 9. (a) Distribution function $G(JCSm)$: number of chemistry groups as a function of $JCSm$ values (class width $\Delta JCSm = 2.0$). (b) Distribution function $G(FCSm)$: number of chemistry groups as a function of $FCSm$ values (class width $\Delta FCSm = 2.0$).

We observe for *CPP* at the aggregation level of a research group a possible slight "cumulative disadvantage" (exponent +0.94) with *FCSm*. In a forthcoming article (van Raan, 2005b) we will address these particular statistical properties in more detail. As a first explanation we suggest that it is increasingly difficult for research groups to reach an impact substantially above the field average as this field average becomes higher and higher. In other words, the higher the crossbar, the more difficult it is to jump over it comfortably.

DOI: 10.1002/asi

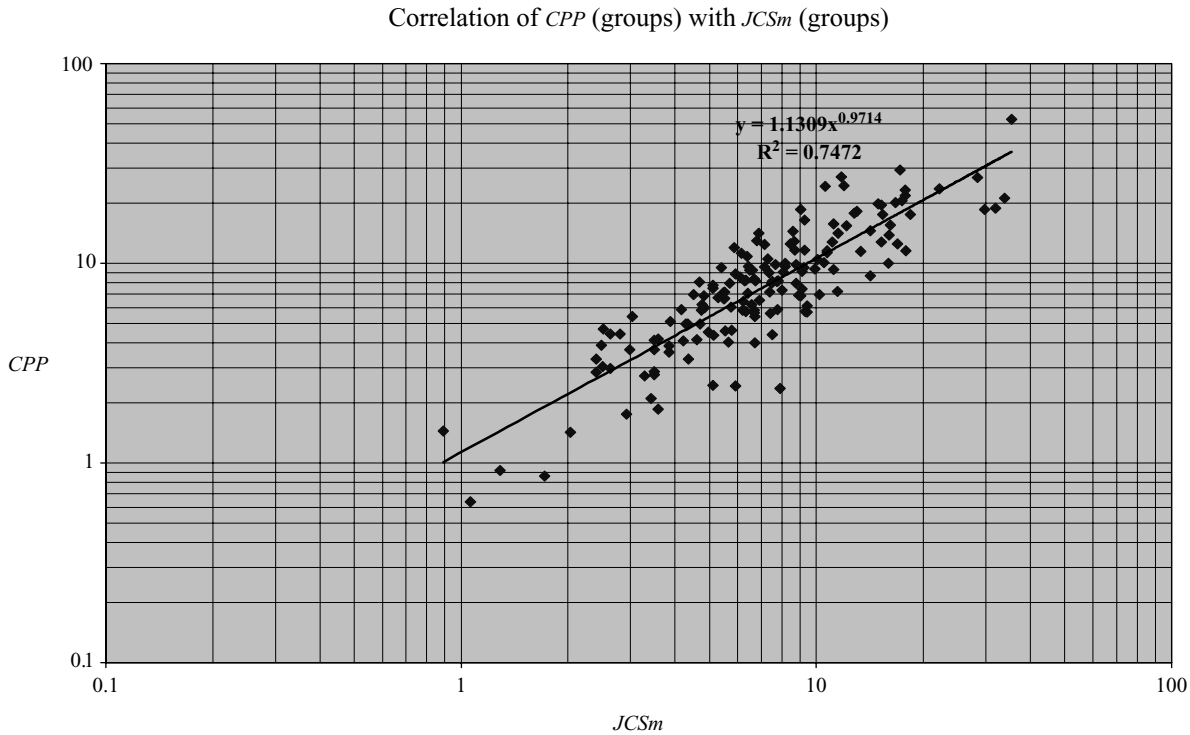


FIG. 10. Correlation of *CPP* with the *JCSm* values for all chemistry groups.

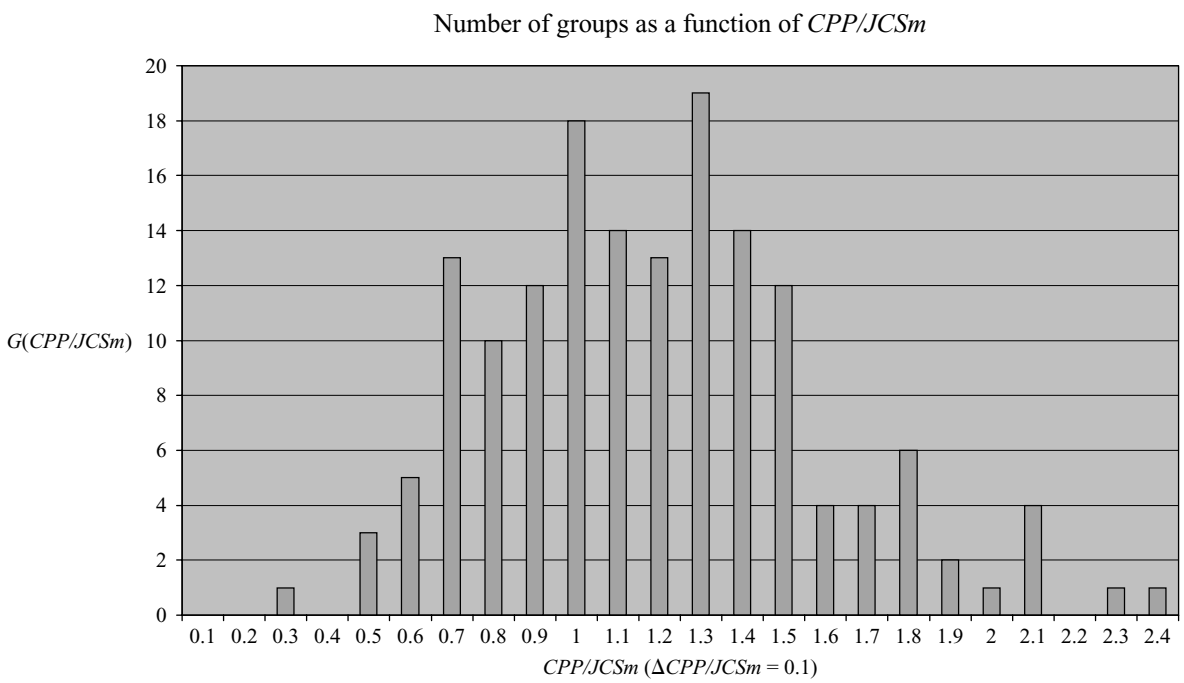


FIG. 11. Distribution function $G(CPP/JCSm)$: number of chemistry groups as a function of $CPP/JCSm$ values (class width $\Delta CPP/JCSm = 0.1$).

On the basis of Equation 7a and the coefficient $R^2 = 0.51$ we can expect that the distribution function $G(CPP/FCSm)$ of the value of the field-normalized number of citations per publication will approach a normal distribution around the value

$$CPP/FCSm \sim 1.5 \tag{7b}$$

Given the lower value of R^2 as compared to the case with $CPP/JCSm$, we have more variance and thus a broader distribution, particularly at the right-hand side: Figure 12 shows that the large *CPP* values correlate with relatively small *FCSm* values, thus extending the distribution more toward the higher-$CPP/FCSm$ side of the distribution. The measured distribution function of $CPP/FCSm$ for

DOI: 10.1002/asi

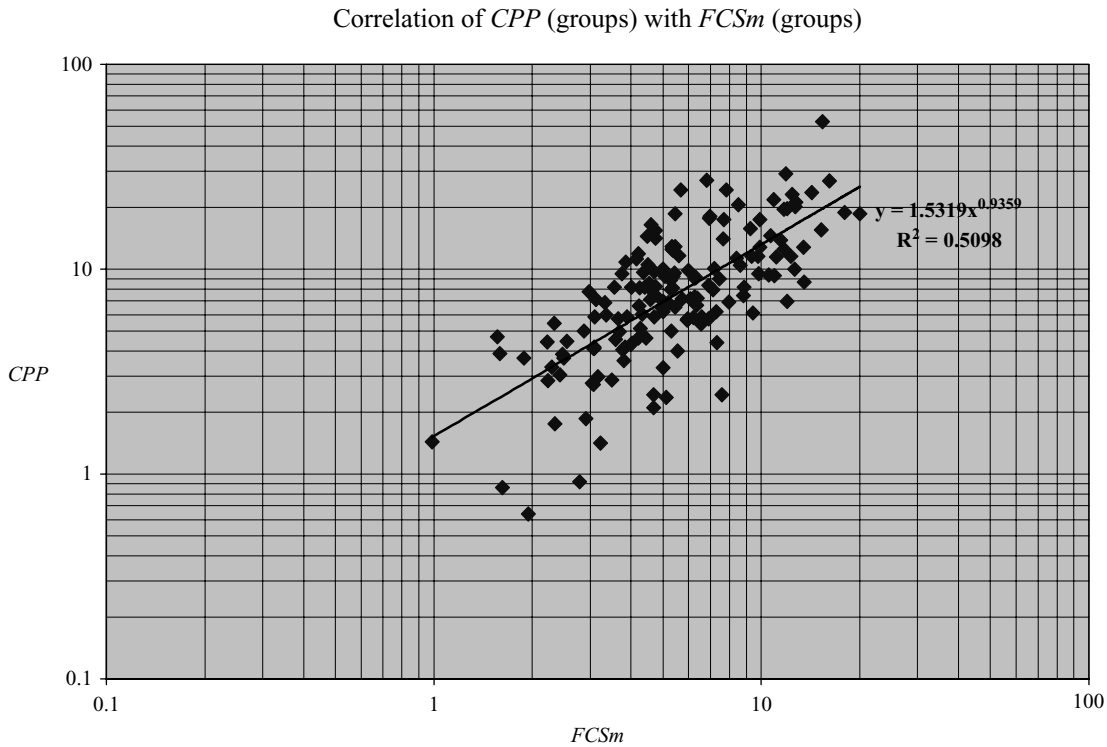


FIG. 12. Correlation of *CPP* with the *FCSm* values for all chemistry groups.

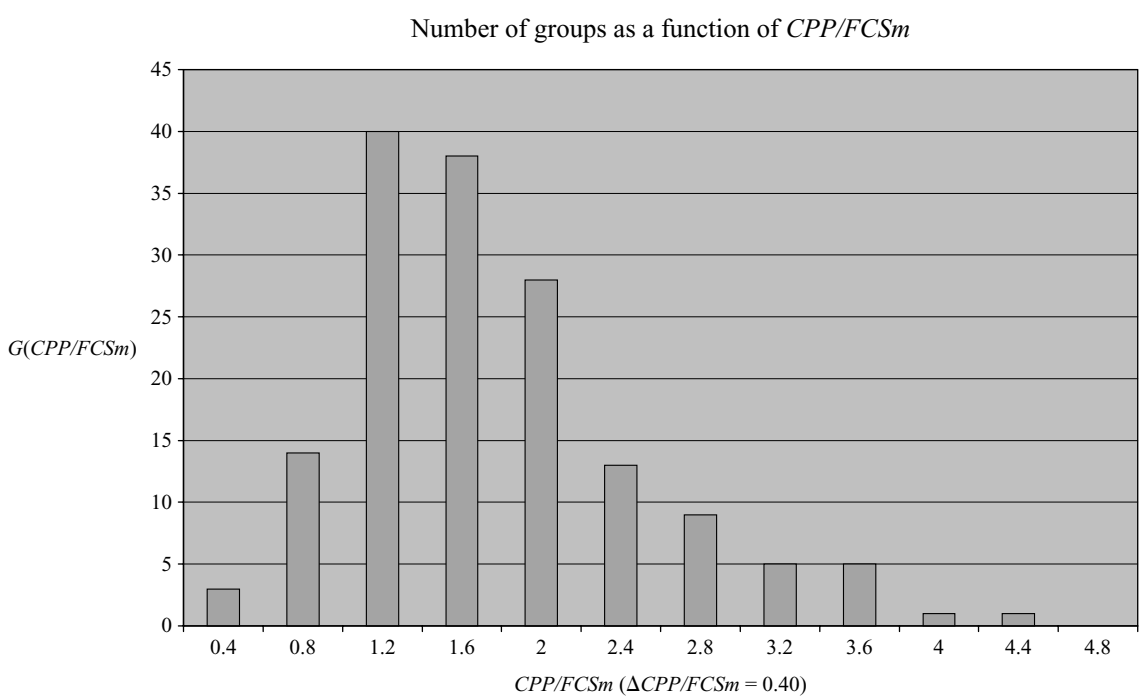


FIG. 13. Distribution function $G(CPP/FCSm)$: number of chemistry groups as a function of $CPP/FCSm$ values (class width $\Delta CPP/FCSm = 0.40$).

all chemistry research groups confirms these findings; see Figure 13.

Finally, the indicator $JCSm/FCSm$ is a normalization (by field) of an already average measure; thus its distribution will be very close to normal, as demonstrated by Figure 14. From Equations 6b and 7b it follows that this distribution will be around

$$JCSm/FCSm \sim 1.4 \tag{7c}$$

as is also confirmed by Figure 14.

DOI: 10.1002/asi

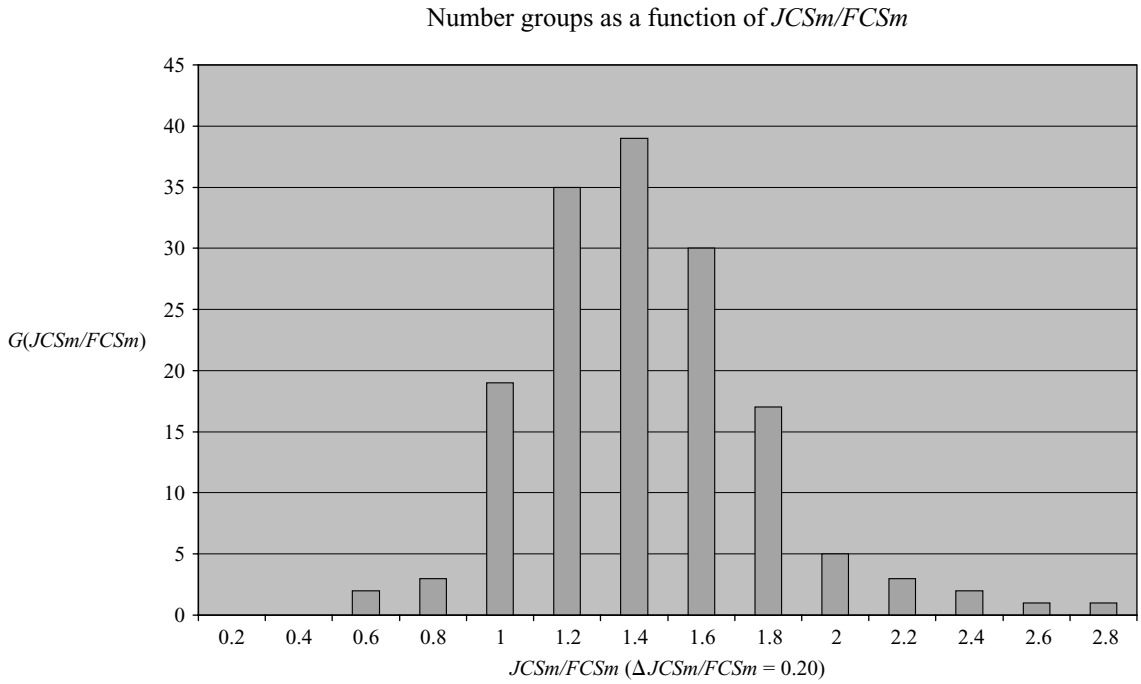


FIG. 14. Distribution function $G(JCSm/FCSm)$: number of chemistry groups as a function of $JCSm/FCSm$ values (class width $\Delta JCSm/FCSm = 0.20$).

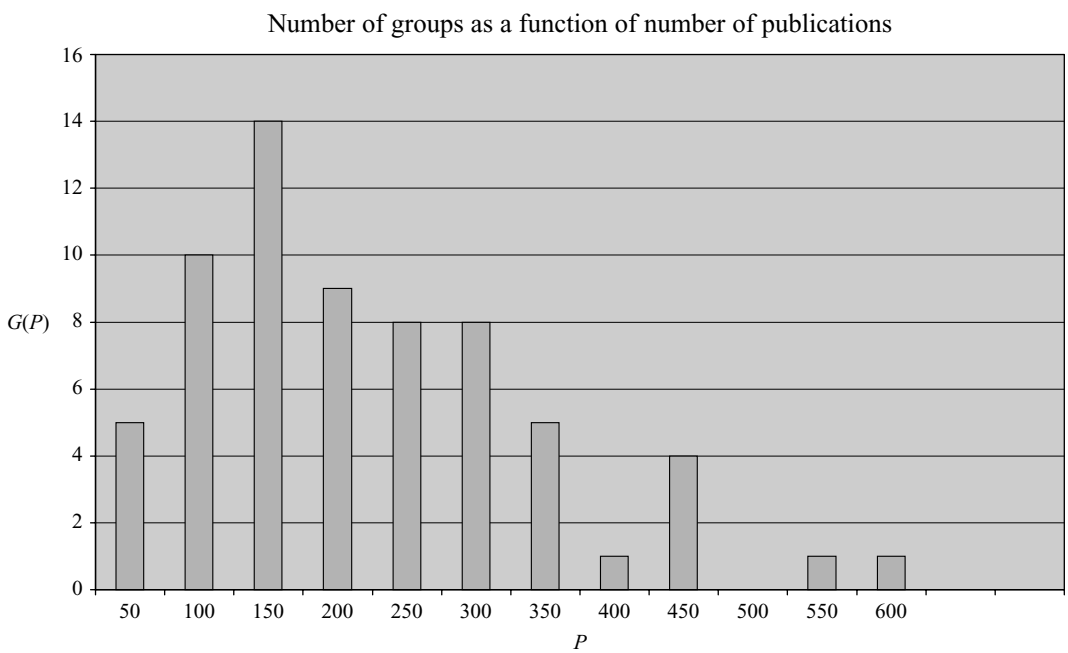


FIG. 15. Distribution function $G(P)$: number of medical research groups as a function of number of publications (class width $\Delta P = 50$).

In the next section we discuss the results for the medical research groups in comparison with the chemistry groups.

Overview of the Results for the Medical Research Groups

Similarly to the chemistry research groups, the 65 medical research groups are not ideal "samples": Their sizes in terms of numbers of publications (P) also show a skewed distribution; see Figure 15 for the distribution function of the number of publications, $G(P)$. Thus again the distribution of the *CPP* indicators will be less skewed as compared to the basic $P(C)$ distribution, but it will still not resemble a normal distribution; see the *CPP* distribution function for all medical research groups, $G(CPP)$, in Figure 16. The much more skewed distribution of the total number of citations (C) for all groups, $G(C)$, is presented in Figure 17.

Compared to that of the chemistry research groups, the $G(P)$ distribution function for the medical research groups is

DOI: 10.1002/asi

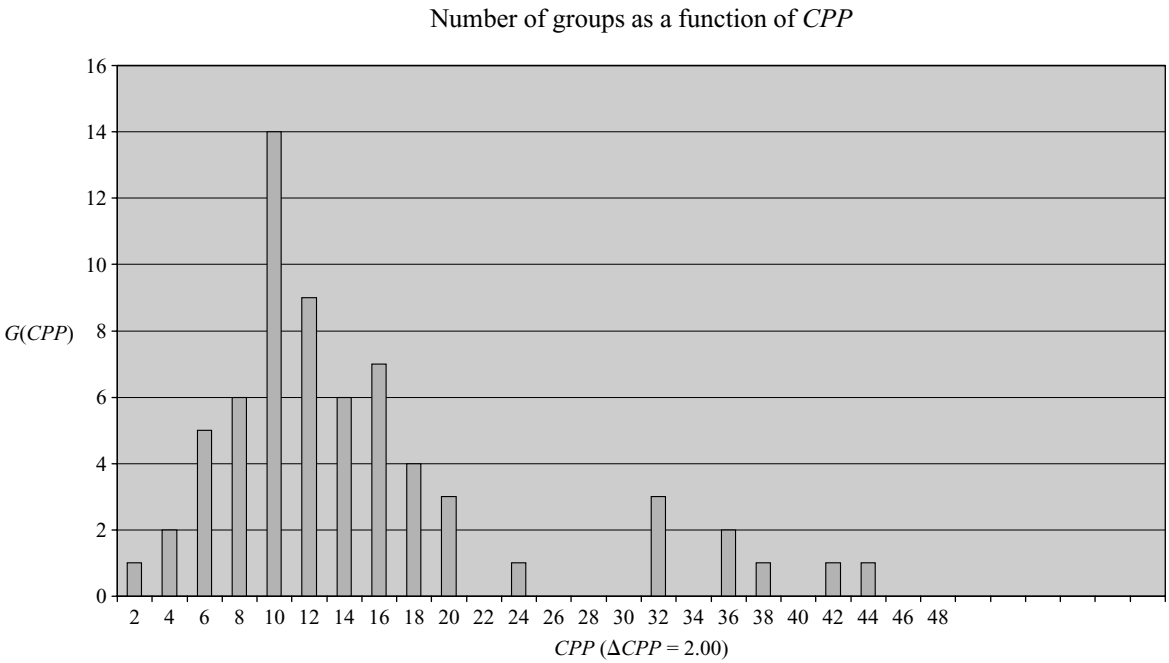


FIG. 16. Distribution function $G(CPP)$: number of medical research groups as a function of CPP values (class width $\Delta CPP = 2.0$).

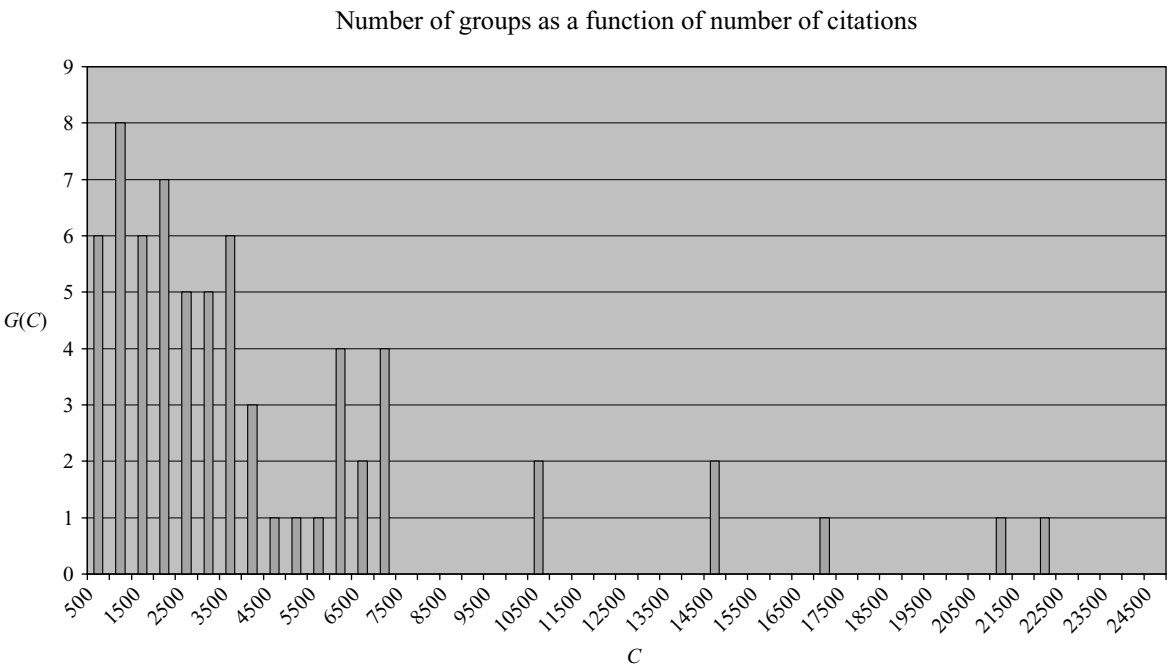


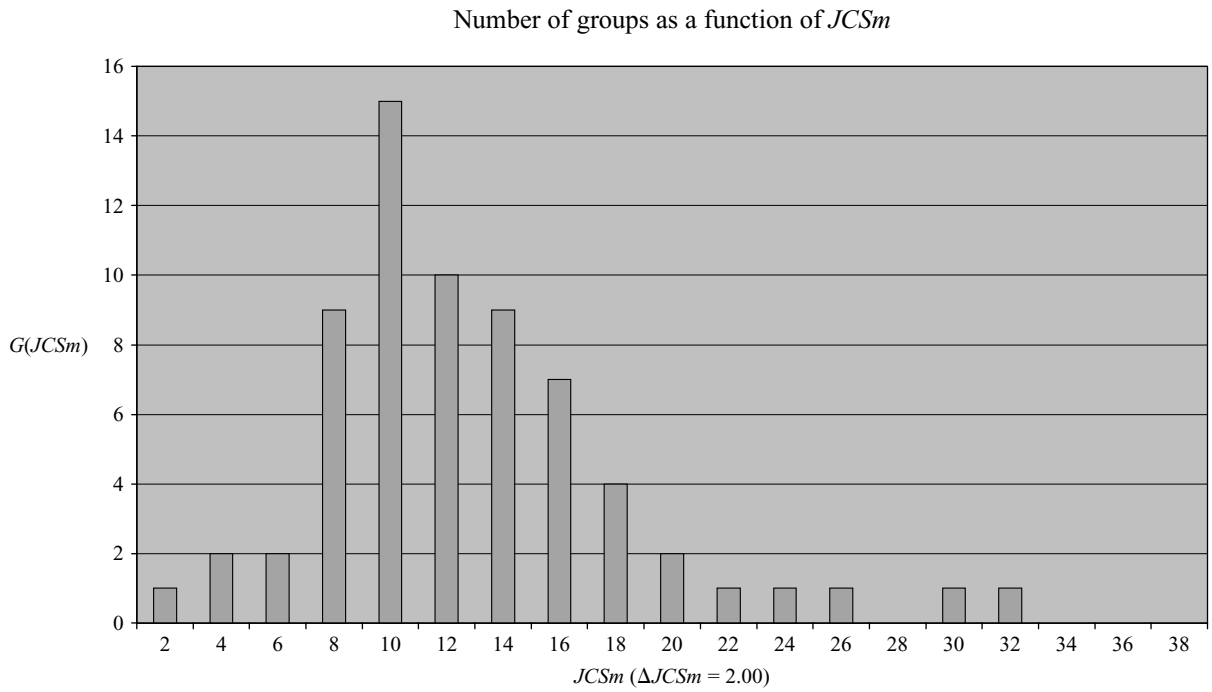
FIG. 17. Distribution function $G(C)$: number of medical research groups as a function of total number of citations (C) (class width $\Delta C = 1{,}000$).

broader at the higher-P side and the mode of the distribution is more at the higher-P side. For $G(C)$ we find a similar situation, but even more pronounced. In addition, several medical research groups with very high numbers of citations are clearly visible. Previously we explained these differences (after the discussion at the end of the second section): The medical research groups belong to one organization (LUMC), which has a policy of supporting strong research programs.

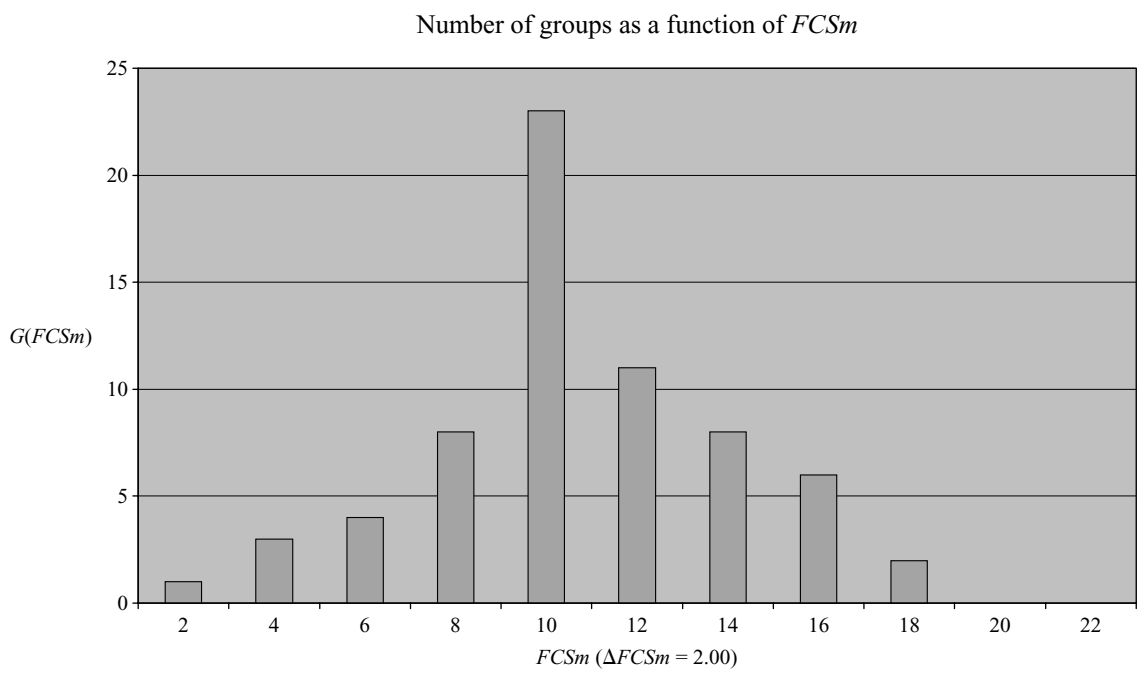
As discussed in the preceding section, the distribution of *JSCm* over all research groups will tend to a normal distribution but is not a perfect normal distribution because of the size difference of the research groups. Also, in the case of the medical research groups these considerations are confirmed by the measurement of the *JSCm* distribution function; see Figure 18a. We again observe a broader distribution and a mode at the higher-*JCSm* side in comparison to the chemical research groups.

As the *FCS* values in the entire population are even more an average than the *JCS* values (namely, the mean value of the number of citations per article in a journal, and then

DOI: 10.1002/asi



(a)



(b)

FIG. 18. (a) Distribution function $G(JCSm)$: number of medical research groups as a function of $JCSm$ values (class width $\Delta JCSm = 2.0$). (b) Distribution function $G(FCSm)$: number of medical research groups as a function of $FCSm$ values (class width $\Delta FCSm = 2.00$).

averaged over all journals in a field), the *FCSm* distribution will be again closer to normal than in the case of the *JCSm* distribution. In the case of the 65 medical research groups, the *FCSm* distribution is almost a perfect normal distribution; see Figure18b.

The *correlation* between size (the total number of publications P of a research group) and the total number of citations C received by a group in a given period for all medical research groups is presented in Figure 19. This figure shows us that this relation *on the aggregation level of research groups* is described with quite high significance ($R^2 = 0.87$) by the equation

$$C(P) = 0.89\, P^{1.55} \tag{8}$$

DOI: 10.1002/asi

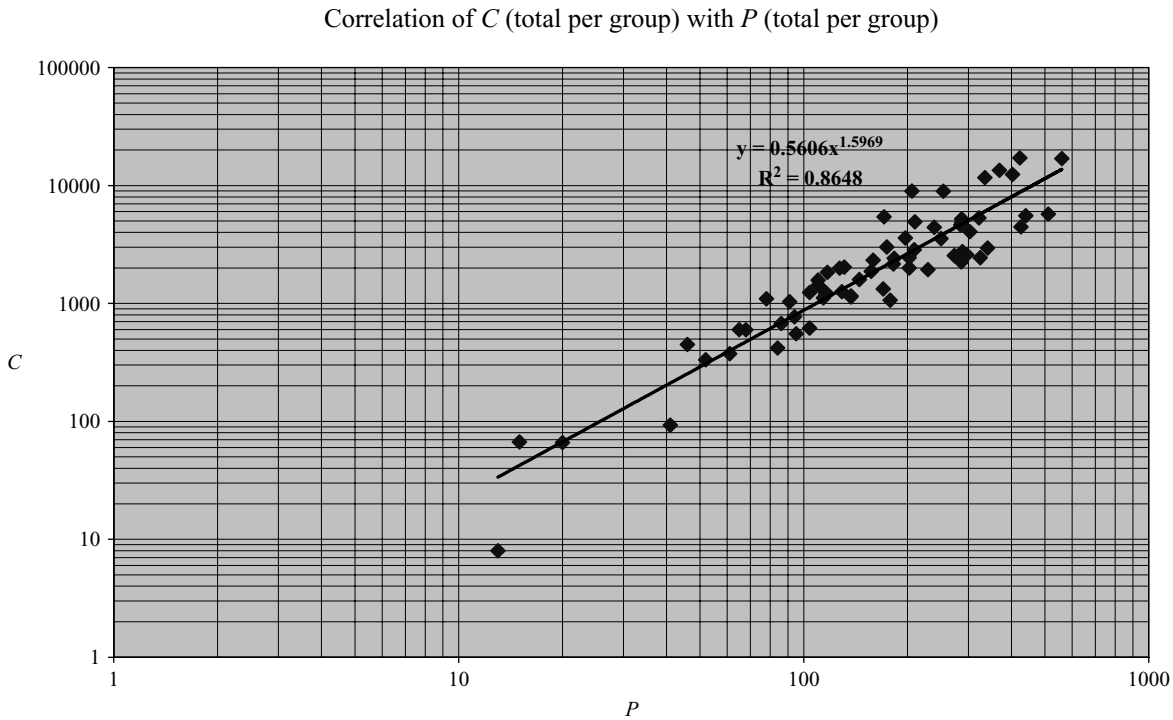


FIG. 19. Correlation of the number of citations (*C*) received per research group with the total number of publications (*P*) per group, for all medical research groups.

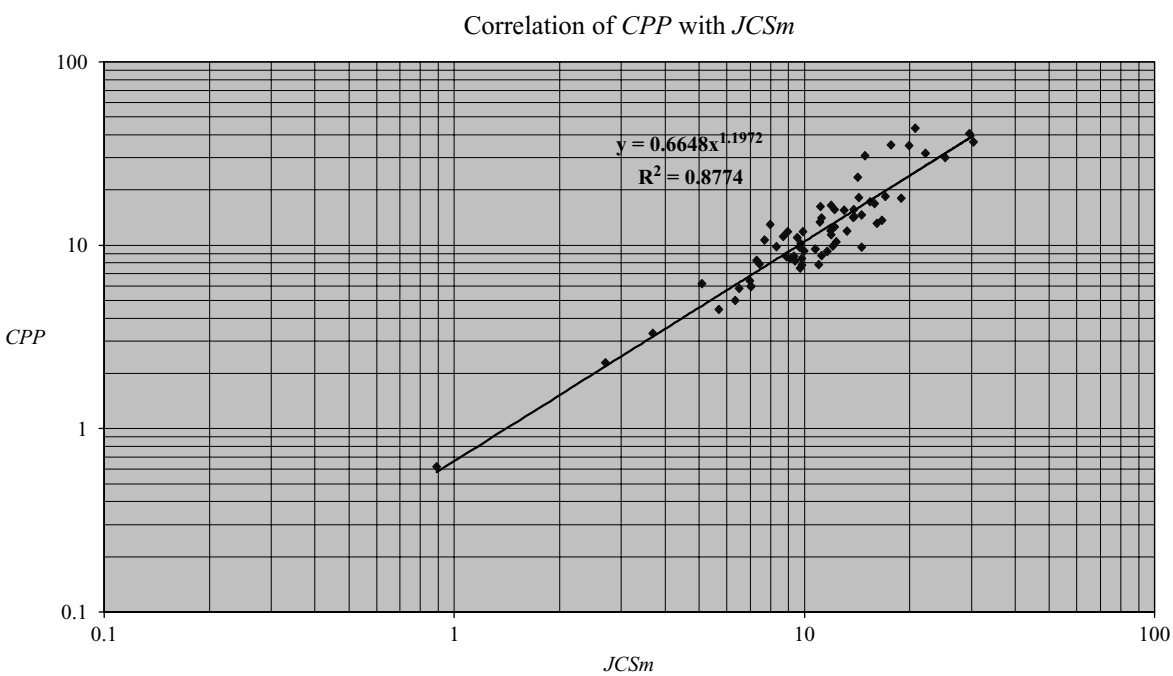


FIG. 20. Correlation of *CPP* with the *JCSm* values for all medical research groups.

and therefore it appears that the size of groups leads to a larger "cumulative advantage" (exponent +1.55) for the number of citations received by these groups than in the case of the chemistry groups; see Equation 3 (exponent +1.28).

For the statistical properties of the *normalized* indicators we follow the same approach used in the case of the chemistry research groups. Thus, we determine for the medical research groups the correlation between the average number of citations per publication (*CPP*) with the average *JCSm*. The results of this analysis are presented in Figure 20. We empirically find with a rather high significance ($R^2 = 0.88$):

$$CPP = 0.66\, JCSm^{1.20} \tag{9}$$

DOI: 10.1002/asi

which means that also in the case of the medical research groups, *CPP* at the aggregation level of a research group is related in a simple, almost proportional manner to *JCSm* with some "cumulative advantage" (exponent +1.20), in contrast to a practically nonexistent cumulative (dis)advantage (exponent +0.97) in the case of the chemistry research groups. An explanation could be that medical researchers are more eager than chemical researchers to publish their better work in higher-impact journals.

On the basis of Equation 9 we can expect for the distribution function $G(CPP/JCSm)$ in good approximation a transformed *JCSm* distribution

$$CPP/JCSm \sim 0.7\ JCSm^{0.20} \tag{10}$$

where by the left-hand side of the *JCSm* distribution (see Figure 18a) by virtue of the low exponent (0.20) will collapse to values near 1, and the right-hand side will collapse to values up to 2. Thus, the $CPP/JCSm$ will start around 0.7 with a near-normal distribution extending up to values around 2.

This finding is confirmed by Figure 21. We observe that for this journal-normalized impact indicator the medical research groups do not have a broader distribution on the right-hand side as compared to the chemical research groups, as can be expected on the basis of Equation 10. The mode is even located at a somewhat lower value of $CPP/JCSm$ for the medical research groups, but the left-hand side of the distribution is less "populated," resulting in a more skewed distribution for the medical research groups as compared to the chemistry groups.

The correlation between the average number of citations per publication (*CPP*) and the average *FCSm* values for all medical research groups is presented in Figure 22. We observe that this relation *on the aggregation level of research groups* is described (with reasonable significance, $R^2 = 0.75$) by the equation

$$CPP = 0.65\ FCSm^{1.30} \tag{11}$$

Also here we find for *CPP* at the aggregation level of a research group a stronger "cumulative advantage" (exponent +1.30) than for the chemistry research groups. On the basis of Equation 11 we can expect for the distribution function $G(CPP/FCSm)$ in good approximation a transformed *FCSm* distribution

$$CPP/FCSm \sim 0.7\ FCSm^{0.30} \tag{12}$$

where by the left-hand side of the transformed *FCSm* distribution (Figure 18b), by virtue of the low exponent (0.30), will collapse to values near 1, and the right-hand side will collapse to values around 2. Thus, the $CPP/FCSm$ will start around 0.7 with a very "sharp," near-normal distribution rapidly decreasing from around 2. This finding is confirmed by Figure 23.

As compared to the chemistry research groups, the $G(JCSm/FCSm)$ distribution function for the medical research group (see Figure 24) is more narrow, similar to what we have found for $G(CPP/JCSm)$.

Finally, as an additional finding we present in Figure 25 the correlation for all medical research groups of the percentage of not-cited publications (*Pnc*, see Table 2) and the average number of citations per article *CPP*. With a reasonable significance ($R^2 = 0.65$) we observe that research groups that are *cited more per article* also have *more cited articles*. This is an interesting indication of the consistency of the citation process in revealing the impact of research groups.

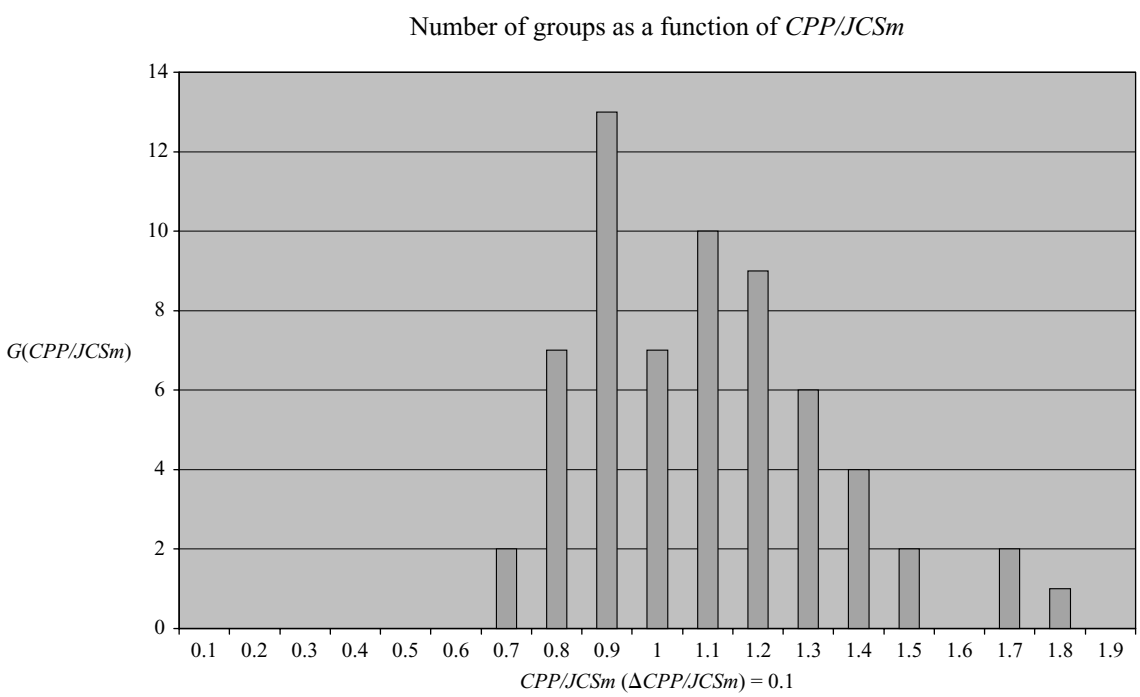


FIG. 21. Distribution function $G(CPP/JCSm)$: number of medical research groups as a function of $CPP/JCSm$ values (class width $\Delta CPP/JCSm = 0.1$).

DOI: 10.1002/asi

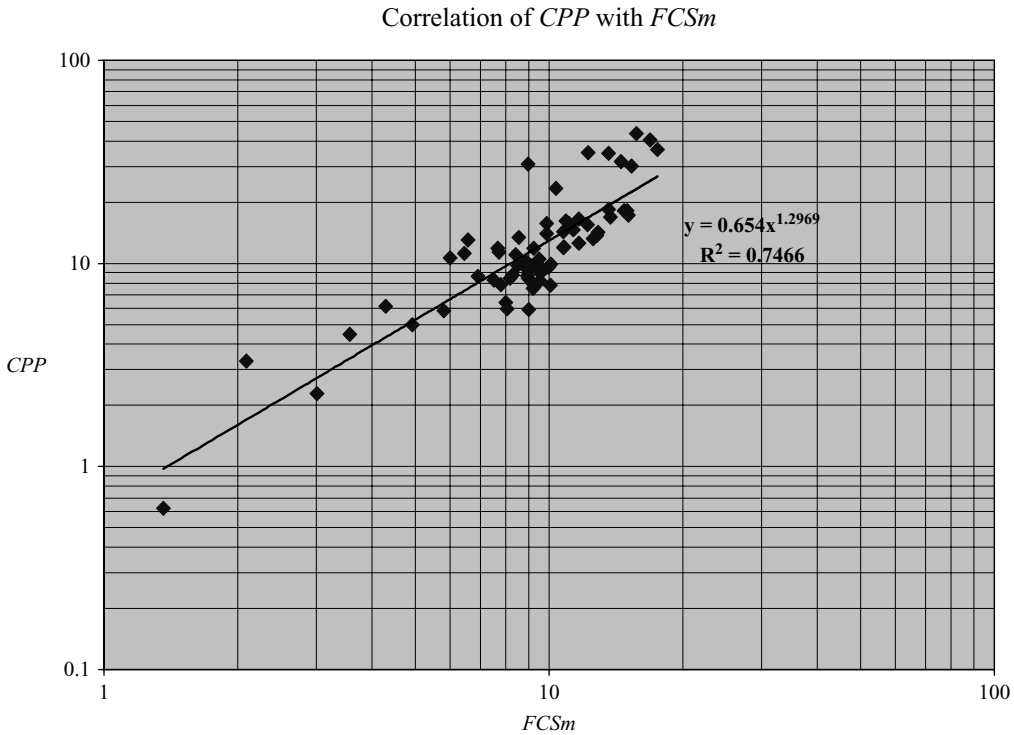


FIG. 22. Correlation of *CPP* with the *FCSm* values for all medical research groups.

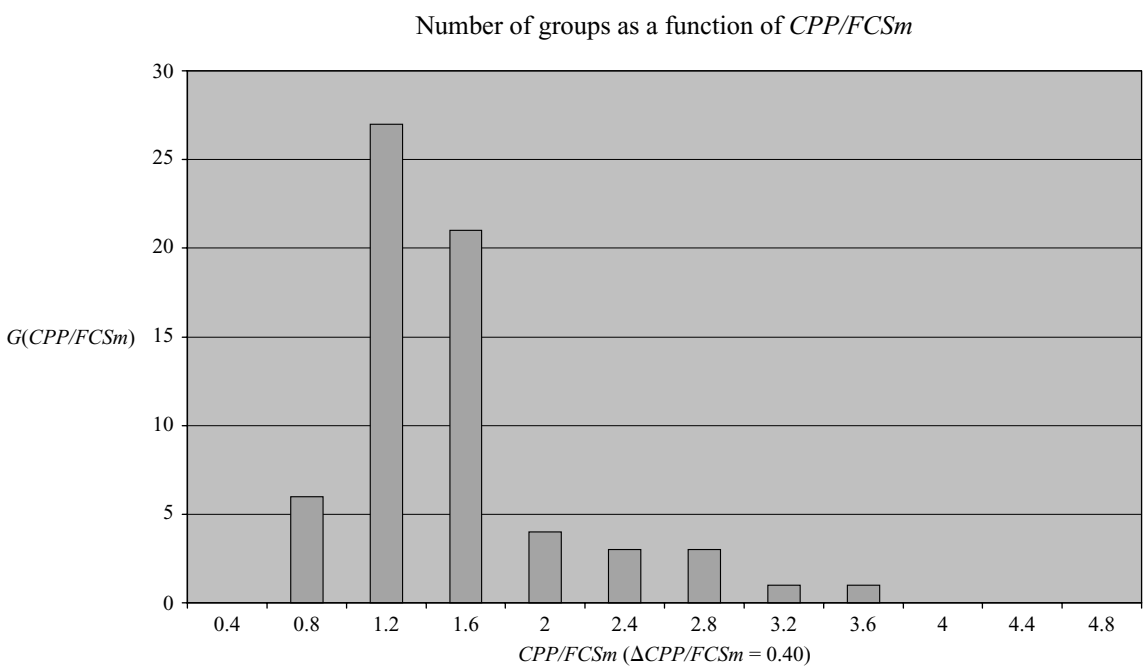


FIG. 23. Distribution function $G(CPP/FCSm)$: number of medical research groups as a function of $CPP/FCSm$ values (class width $\Delta CPP/FCSm = 0.40$).

Summary of the Main Findings and Concluding Remarks

We have presented an empirical approach to the study of the statistical properties of bibliometric indicators on a very relevant but not simply "available" aggregation level: the research group. The focus of our study is on the distribution functions of a coherent set of bibliometric indicators frequently used as a measuring instrument in the analysis of research performance in order to provide better insight into the statistical properties of the bibliometric instrument.

We have performed our analysis of the statistical properties of bibliometric indicators on the basis of the data of

DOI: 10.1002/asi

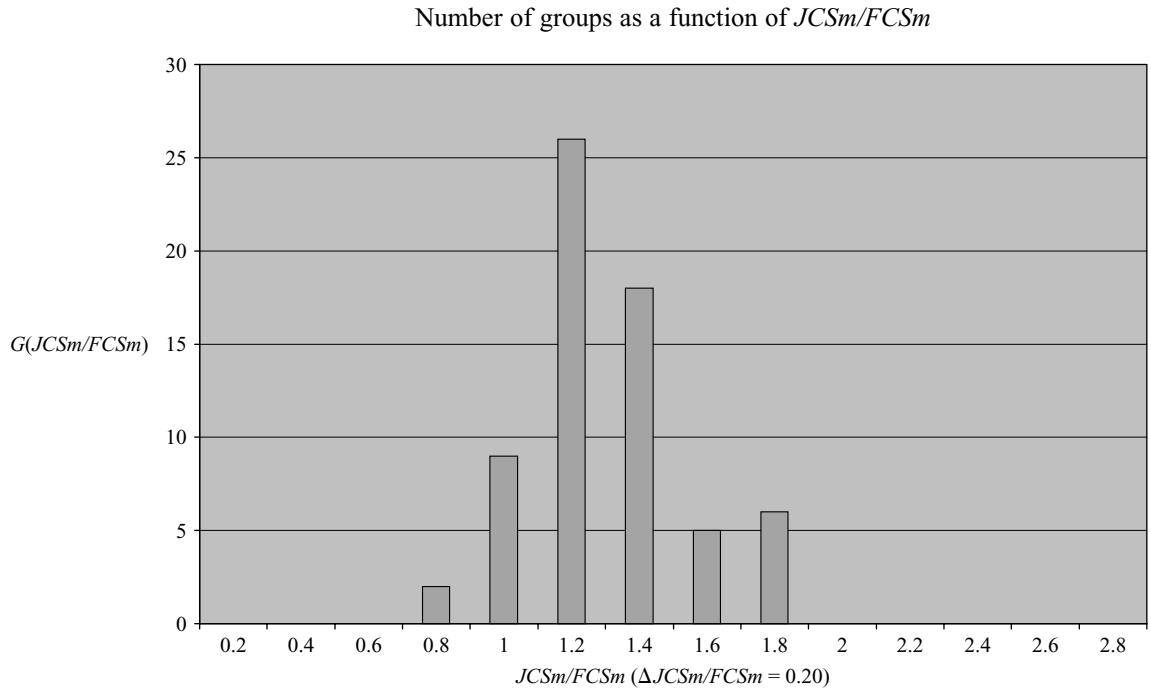


FIG. 24. Distribution function $G(JCSm/FCSm)$: number of medical research groups as a function of $JCSm/FCSm$ values (class width $\Delta JCSm/FCSm = 0.20$).

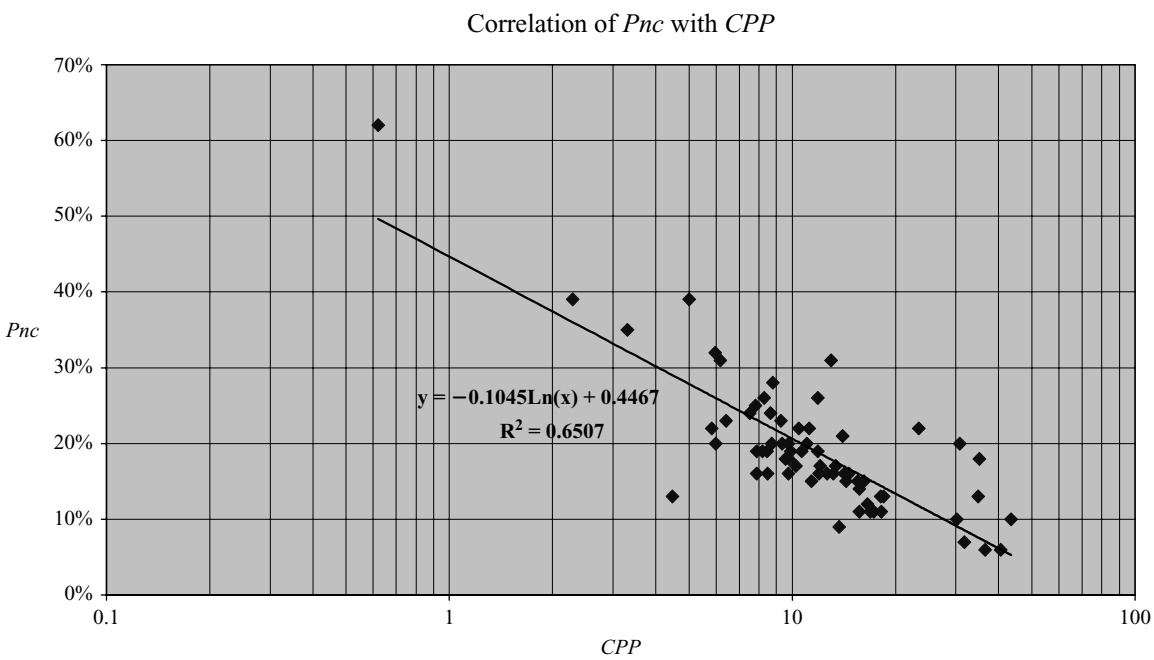


FIG. 25. Correlation of *Pnc* with *CPP* for all medical research groups).

157 chemistry groups covering the work of around 700 senior researchers over a period of 10 years, with about 18,000 publications and 175,000 citations to these publications, and the data of 65 medical research groups covering the work of around 300 senior researchers over a period of 12 years, with about 10,000 publications and 185,000 citations to these publications. Given the size of both data sets we are confident that our approach can be considered as a representative case study to investigate statistical properties of bibliometric indicators in general.

Starting with the most basic distribution in bibliometric analysis, the very skewed distribution of citations over publications, we clearly observe the working of the central limit theorem. We find that at the level of research groups the

DOI: 10.1002/asi

distribution functions of the main bibliometric indicators, particularly the journal-normalized and the field-normalized indicators, are approaching normal distributions. The important consequence of these findings is that on the level of research groups the statistical behavior of advanced bibliometric indicators is such that meaningful comparisons (benchmarking) between groups can be made in terms of reference values based on mean values and variances.

Finally, we remark that this study focuses on statistical characteristics of a large set of research groups. Currently we are investigating (van Raan, in press-b) how specific subsets within the total set, for instance, top-performance groups versus lower-performance groups, will exhibit different statistical properties.

Acknowledgments

The author would like to thank his CWTS colleague Thed van Leeuwen for the data collection, data analysis, and calculation of the bibliometric indicators for the two sets of research groups.

References

Albert, A., & Barabási, A.-L. (2002). Statistical mechanics of complex networks. Reviews of Modern Physics, 74, 47–97.

Egghe, L., & Rousseau, R. (1990). Introduction to informetrics: Quantitative methods in library, documentation and information science. Amsterdam: Elsevier Science.

Haitun, S.D. (1982). Stationary scientometric distributions: 1. Different approximations. Scientometrics, 4, 89–104.

Katz, J.S. (1999). The self-similar science system. Research Policy, 28, 501–517.

Katz, J.S. (2000). Scale independent indicators and research assessment. Science and Public Policy, 27, 1, 23–36.

Katz, J.S. (2005). Scale independent indicators. Measurement: Interdisciplinary Research and Perspectives, 3(1), 24–28.

Laherrère, J., & Sornette, D. (1998). Stretched exponential distributions in nature and economy: "Fat tails" with characteristic scales. European Physical Journal B, 2, 525–539.

Lotka, A.J. (1926). The frequency distribution of scientific productivity. Journal of the Washington Academy of Science, 16, 317–323.

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

Merton, R.K. (1988). The Matthew effect in Science: II. Cumulative advantage and the symbolism of intellectual property. Isis, 79, 606–623.

Moed, H.F., & Garfield, E. (2004). In basic science the percentage of "authoritative" references decreases as bibliographies become shorter. Scientometrics, 60(3), 295–303.

Moed, H.F., & van Leeuwen, T.N. (1995). Improving the accuracy of the Institute for Scientific Information's journal impact factors. Journal of the American Society for Information Science, 46, 461–467.

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

Naranan, S. (1971). Power law relations in science bibliography—A self-consistent interpretation. Journal of Documentation, 27, 83–97.

Newbold, P. (1995). Statistics for business and economics. London: Prentice-Hall International.

Price, D.J. de S. (1965). Networks of scientific papers. Science, 149, 510–515.

Redner, S. (1998). How popular is your paper? An empirical study of the citation distribution. European Physical Journal B, 4, 131–134.

Schubert A., & Glänzel, W. (1983). Statistical reliability of comparisons based on the citation impact of scientometric publications. Scientometrics, 5, 59–74.

Seglen, P.O. (1992). The skewness of science. Journal of the American Society for Information Science, 43, 628–638.

Seglen, P.O. (1994). Causal relationship between article citedness and journal impact. Journal of the American Society for Information Science, 45, 1–11.

Simon, H.A. (1955). On a class of skew distribution functions. Biometrika, 42, 425–440.

Tsallis, C., & de Albuquerque, M.P. (2000). Are citations of scientific papers a case of nonextensivity? European Physical Journal B, 13, 777–780.

van Leeuwen, T.N., Visser, M.S., Moed, H.F., & Nederhof, A.J. (2002). The third bibliometric study on chemistry research associated with the Council for Chemical Sciences of the Netherlands Organisation for Scientific Research (NWO-CW) 1991–2000 (Report CWTS 2002–01). Leiden: CWTS.

van Raan, A.F.J. (1996). Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises. Scientometrics, 36, 397–420.

van Raan, A.F.J. (2001a). Two-step competition process leads to quasi power-law income distributions: Application to scientific publication and citation distributions. Physica A, 298, 530–536.

van Raan, A.F.J. (2001b). Competition among scientists for publication status: Toward a model of scientific publication and citation distributions. Scientometrics, 51, 347–357.

van Raan, A.F.J. (2004). Measuring science: Capita selecta of current main issues. In H.F. Moed, W. Glänzel, & U. Schmoch (Eds.), Handbook of quantitative science and technology research (pp. 19–50). Dordrecht: Kluwer Academic.

van Raan, A.F.J. (2005a). Fatal attraction: Ranking of universities by bibliometric methods, Scientometrics, 62, 1, 133–143.

van Raan, A.F.J. (2005b). Reference-based publication networks with episodic memories. Scientometrics, 63(3), 549–566.

van Raan, A.F.J. (in press). Performance-related differences of bibliometric statistical properties of research groups: Cumulative advantages and hierarchically layered networks. Journal of the American Society for Information Science and Technology.

VSNU. (2002). Chemistry and chemical engineering (VSNU Assessment of Research Quality series). Utrecht: VSNU.

DOI: 10.1002/asi

Appendix

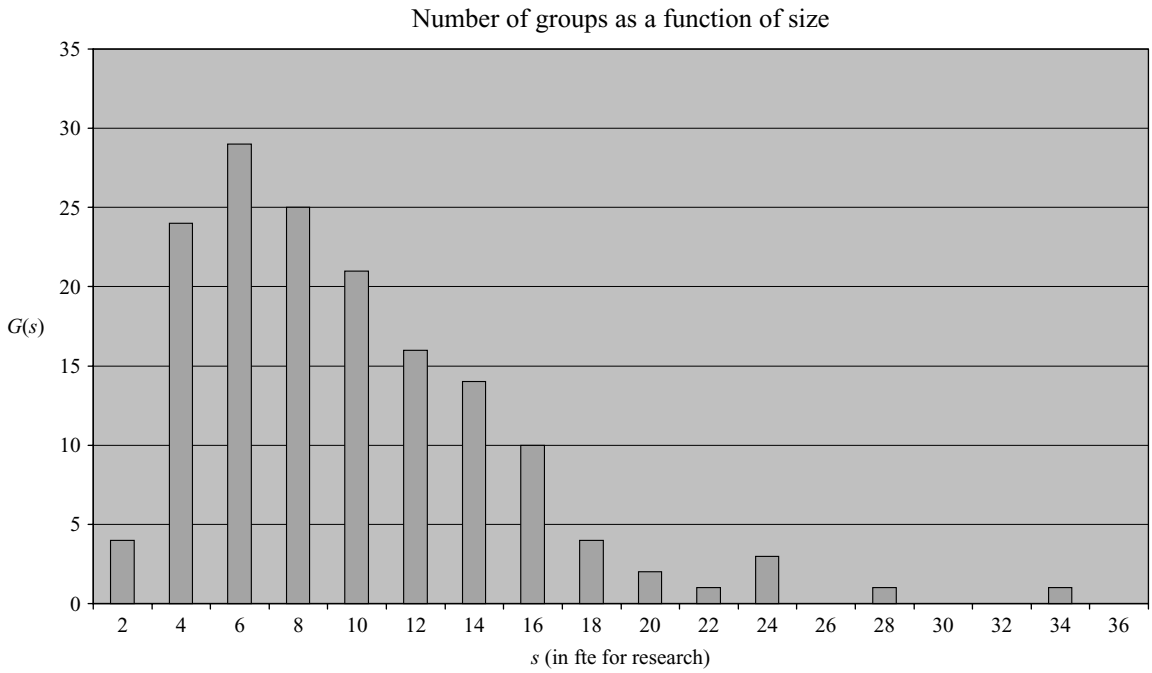


FIG. A1. Distribution function $G(s)$: number of chemistry groups as a function of size, in terms of number of staff full-time equivalents (FTEs) available for research (s, in FTE; class width $\Delta s = 2.0$).

DOI: 10.1002/asi