Abstract

In this paper 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 bibliometric indicators. These indicators 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 skew. 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 bibliometric indicators, particularly the journal-normalized and the field-normalized indicators are approaching normal distributions. The results of our study underline the importance of the idea of ‘group oeuvre’, i.e., the role of sets of related publications as a unit of analysis.

1. Introduction

An appropriate application of bibliometric indicators in the evaluation of scientific research requires the consideration of the statistical properties of these indicators. 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 work of Lotka (1926), Price (1965), Naranan (1971), Haitun (1982), Redner (1998), and to Simon (1955), and to Laherrère and Sornette (1998) for a recent, general discussion 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 from 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 to the ISI1 journal impact factor.

1The 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 paper we use the term ‘CI’ (Citation Index) for the above set of databases.
only, and did not consider other types of journal impact indicators. The careful analysis of the relations between specific indicators, such as article citedness and journal impact, is a crucial part of the bibliometric methodology. In this study on statistical properties of bibliometric indicators we focus on distribution functions of the individual bibliometric 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 above mentioned skewness of distributions. For instance, only a small fraction of articles in a journal is cited around 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 bibliometric 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 bibliometric 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) implying a non-linear increase of impact with increasing size, demonstrated by the finding that the number of citations as a function of number 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 will probably exist across smaller entities such as research groups. In this paper we will show that this is indeed the case (Section 3.1). Katz argues that the ‘conventional’ bibliometric indicators may fail to account for this non-linearity between size -measured by number of publications- and impact -measured by number of citations- and this could result in an over- or under-estimation of research performance. In other words, smaller groups have a smaller propensity than larger groups to be cited. Therefore, ‘conventional’ bibliometric indicators should be supplemented with measures that are corrected for the above scaling relationships (Katz 2004).

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 will grow. At the same time, they can develop a larger and more differentiated research programme with strongly related themes, which very well may reinforce further the scientific performance and influence of a group. Thus one could also argue that a larger impact as measured on the bases of citations cannot be simply waved aside as purely a scale-dependent effect. In this way groups are ‘punished’ for having reached a considerable size as the number of citations received by them should be corrected for size. Furthermore, groups with 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.

An interesting question, however, is whether indicators that are more complex than simple citation counts -such as our field-specific normalized indicator ‘crown indicator’ \(\text{CPP/FCS}\)\(^m\).

\(^2\) 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.
still exhibit a ‘cumulative advantage’ scaling behaviour. We will show in this paper that this is not the case.

Next to the application as a tool for research performance evaluation, bibliometric indicators can also be applied as an instrument for investigating characteristics of science as a knowledge-generating and communication system. Thus, further important questions are: 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 such as: how do scientists disseminate their findings as optimal as possible; how will these knowledge flows go given general network properties (e.g., co-author 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-dimensional characteristic, ranging from recognition to ‘utility’ or persuasion. An interesting recent finding is that scientists refer less to ‘authoritative’ papers in the case of smaller total numbers of references (Moed and Garfield 2004). From this observation these authors conclude that citing such ‘authoritative’ papers is not a major motivation of an author. As these ‘authoritative’ papers are often highly cited, the above citation behaviour may cause a mitigation of the accumulation of citations to already highly cited papers (Matthew effect) and thus it will influence citation distribution functions.

In this paper 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 R&D laboratories of companies. We use the results of two large evaluation studies. First, all university chemistry groups in the Netherlands, covering in the 10-year period 1991-2000 in total 157 research groups, about 700 senior researchers with about 18,000 publications (CI-based) and 175,000 citations (excluding self-citations) to these publications. Second, all 65 research groups of the Faculty of Medicine, Leiden University (Leiden University Medical Center, LUMC) covering a twelve years period (1990-2001) with about 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 as compared to the Seglen study which involved 16 researchers and 907 publications (in journals to which an impact factor could be assigned).

In this paper we will apply several quantitative methods (Egghe and Rousseau 1990; Newbold 1995) to unravel the statistical properties of bibliometric indicators. The structure of this paper is as follows. In Section 2 we discuss the data material for two sets of research groups, the application of the bibliometric method and the calculation of the indicators. Section 3 addresses the statistical analysis and discusses the results of the analysis. Finally, in Section 4 we summarize the main outcomes of this study.

2. Data material, the bibliometric method and calculation of the indicators

2.1 The two datasets

As discussed in the foregoing 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-years period, and the second concerning all research groups in a large medical 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 are used for statistical analysis of the characteristics of bibliometric indicators at the research group level. We stress again that the research level is the most important ‘work floor entity’ in science. However, data at the research group level are by far a trivial matter as ‘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 word, there are no data on actual research groups available externally like the availability of data on the level of the individual scientist. The only possibility 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 above, the data used in this study are the results of evaluation studies and are therefore based on data acquirement with strict verification procedures.

2.2 Chemistry research over a period of ten years

The first set concerns all publications (as far as covered by the Citation Index, ‘CI publications’) of all university research groups in chemistry and chemical engineering in the Netherlands. These 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 we refer to the evaluation report (VSNU 2002). In the framework of this evaluation study, we performed an extensive bibliometric analysis as a support to the international peer committee (van Leeuwen et al 2002). An executive summary of the bibliometric results is included in the evaluation report. The time 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 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 textbox in this section, for a detailed discussion we refer to Van Raan (1996, 2004).

**CWTS Standard Bibliometric Indicators:**

- Number of publications \( (P) \) in international journals of the research group/institute/etc. in the entire period;
- Number of citations received by \( P \) during the entire period, with and without self-citations \( (Ci \text{ and } C) \);
- Average number of citations per publication, again with and without self-citations \( (CPPi \text{ and } CPP) \);
- Percentage of publications not cited (in the give time period), \( Pnc \);
- Journal-based worldwide average impact as an international reference level for the research group/institute/etc. \( (JCS, \text{journal citation score}, \text{without self-citations (on this world-wide scale!)} \), 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 are used as in the case of \( CPP \);
- Field-based worldwide average impact as an international reference level for the research group/institute/etc. \( (FCS, \text{field citation score}, \text{without self-citations (on this world-wide scale!)} \) 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 are used as in the case of \( CPP \);
- Comparison of the actually received international impact of the research group/institute/etc. with the worldwide average based on \( JCSm \) as a standard, without self-citations, indicator \( CPP/JCSm \).
• Comparison of the actually received international impact of the research group/institute/etc. with the world-wide 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/institute/etc. publishing in top or in sub-top (in terms of ‘citedness’) journals?  
• Percentage of self-citations, \( SelfCit. \)

The indicators are calculated on the basis of the ‘total block analysis’, which means that publications are counted for the entire 10-year period from 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 within the chemistry were covered by this set of university groups, the main fields being analytical chemistry, spectroscopy and microscopy; computational and theoretical chemistry, physical chemistry; catalysis; inorganic chemistry; organic and bio-organic chemistry; biochemistry, microbiology and biochemical engineering; polymer science and technology; materials science; chemical engineering.

2.3 Medical research over a period of twelve years

The second set concerns all publications (as far as covered by the Citation Index, ‘CI publications’) of all research groups in the Leiden University Medical Center (LUMC). 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 as a support to the LUMC research commission. Details of the bibliometric results are available from the author of this paper. The time period covered is 1990-2001 for both publications and the citation 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 of high international reputation. Practically all fields of medical research are present, ranging from molecular cell biology to oncological surgery, and from organ transplantation to T-cell immune response research.

2.4 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 ten universities (‘Univ A’). Also the quality judgement of the international peer committee is indicated. The peers used a three-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**: Example of the results of the bibliometric analysis for the chemistry groups

<table>
<thead>
<tr>
<th>Research group</th>
<th>P</th>
<th>C</th>
<th>CPP</th>
<th>JCSm</th>
<th>FCSm</th>
<th>CPP/JCSm</th>
<th>CPP/FCSm</th>
<th>JCSm/FCSm</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Univ A, 01</td>
<td>92</td>
<td>554</td>
<td>6.02</td>
<td>5.76</td>
<td>4.33</td>
<td>1.05</td>
<td>1.39</td>
<td>1.33</td>
<td>5</td>
</tr>
<tr>
<td>Univ A, 02</td>
<td>69</td>
<td>536</td>
<td>7.77</td>
<td>5.12</td>
<td>2.98</td>
<td>1.52</td>
<td>2.61</td>
<td>1.72</td>
<td>4</td>
</tr>
<tr>
<td>Univ A, 03</td>
<td>129</td>
<td>3780</td>
<td>29.3</td>
<td>17.2</td>
<td>11.86</td>
<td>1.7</td>
<td>2.47</td>
<td>1.45</td>
<td>5</td>
</tr>
<tr>
<td>Univ A, 04</td>
<td>80</td>
<td>725</td>
<td>9.06</td>
<td>8.06</td>
<td>6.25</td>
<td>1.12</td>
<td>1.45</td>
<td>1.29</td>
<td>4</td>
</tr>
<tr>
<td>Univ A, 05</td>
<td>188</td>
<td>1488</td>
<td>7.91</td>
<td>8.76</td>
<td>5.31</td>
<td>0.9</td>
<td>1.49</td>
<td>1.65</td>
<td>5</td>
</tr>
<tr>
<td>Univ A, 06</td>
<td>52</td>
<td>424</td>
<td>8.15</td>
<td>6.27</td>
<td>3.56</td>
<td>1.3</td>
<td>2.29</td>
<td>1.76</td>
<td>0.8</td>
</tr>
<tr>
<td>Univ A, 07</td>
<td>52</td>
<td>362</td>
<td>6.96</td>
<td>4.51</td>
<td>5.01</td>
<td>1.54</td>
<td>1.39</td>
<td>0.90</td>
<td>3</td>
</tr>
<tr>
<td>Univ A, 08</td>
<td>171</td>
<td>1646</td>
<td>9.63</td>
<td>6.45</td>
<td>4.36</td>
<td>1.49</td>
<td>2.21</td>
<td>1.48</td>
<td>4</td>
</tr>
<tr>
<td>Univ A, 09</td>
<td>132</td>
<td>2581</td>
<td>19.55</td>
<td>15.22</td>
<td>11.71</td>
<td>1.54</td>
<td>2.21</td>
<td>1.48</td>
<td>4</td>
</tr>
<tr>
<td>Univ A, 10</td>
<td>119</td>
<td>2815</td>
<td>23.66</td>
<td>22.23</td>
<td>14.25</td>
<td>1.06</td>
<td>1.66</td>
<td>1.66</td>
<td>4</td>
</tr>
<tr>
<td>Univ A, 11</td>
<td>141</td>
<td>1630</td>
<td>11.56</td>
<td>17.83</td>
<td>12.3</td>
<td>0.65</td>
<td>0.94</td>
<td>1.45</td>
<td>4</td>
</tr>
<tr>
<td>Univ A, 12</td>
<td>102</td>
<td>1025</td>
<td>10.05</td>
<td>10.48</td>
<td>7.18</td>
<td>0.96</td>
<td>1.40</td>
<td>1.46</td>
<td>5</td>
</tr>
</tbody>
</table>

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 2005c) we will address the correlation between the bibliometric indicators and the quality judgements by the peers.

As in the case of chemistry research, we applied the same CWTS standard bibliometric indicators 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 thus are directly comparable. Only in the LUMC (medical) case we a somewhat longer period (12 years) as compared to the chemistry case (10 years). 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.

**Table 2: Example of the results of the bibliometric analysis for the medical groups**

<table>
<thead>
<tr>
<th>Research group</th>
<th>P</th>
<th>C</th>
<th>CPP</th>
<th>JCSm</th>
<th>FCSm</th>
<th>CPP/JCSm</th>
<th>CPP/FCSm</th>
<th>JCSm/FCSm</th>
<th>Pnc</th>
<th>Scit</th>
</tr>
</thead>
<tbody>
<tr>
<td>LU 01</td>
<td>117</td>
<td>1,836</td>
<td>15.69</td>
<td>12.2</td>
<td>11.08</td>
<td>1.29</td>
<td>1.42</td>
<td>1.1</td>
<td>11%</td>
<td>20%</td>
</tr>
<tr>
<td>LU 02</td>
<td>197</td>
<td>3,587</td>
<td>18.21</td>
<td>14.28</td>
<td>14.75</td>
<td>1.28</td>
<td>1.23</td>
<td>0.97</td>
<td>11%</td>
<td>21%</td>
</tr>
<tr>
<td>LU 03</td>
<td>46</td>
<td>449</td>
<td>9.76</td>
<td>14.55</td>
<td>8.78</td>
<td>0.67</td>
<td>1.11</td>
<td>1.66</td>
<td>20%</td>
<td>23%</td>
</tr>
<tr>
<td>LU 04</td>
<td>560</td>
<td>16,906</td>
<td>30.19</td>
<td>25.22</td>
<td>15.29</td>
<td>1.2</td>
<td>1.97</td>
<td>1.65</td>
<td>10%</td>
<td>19%</td>
</tr>
<tr>
<td>LU 05</td>
<td>423</td>
<td>17,144</td>
<td>40.53</td>
<td>29.6</td>
<td>16.85</td>
<td>1.37</td>
<td>2.41</td>
<td>1.76</td>
<td>6%</td>
<td>21%</td>
</tr>
<tr>
<td>LU 06</td>
<td>369</td>
<td>13,454</td>
<td>36.46</td>
<td>30.34</td>
<td>17.54</td>
<td>1.2</td>
<td>2.08</td>
<td>1.73</td>
<td>6%</td>
<td>19%</td>
</tr>
<tr>
<td>LU 07</td>
<td>91</td>
<td>1,036</td>
<td>11.38</td>
<td>11.91</td>
<td>7.72</td>
<td>0.96</td>
<td>1.47</td>
<td>1.54</td>
<td>15%</td>
<td>22%</td>
</tr>
<tr>
<td>LU 08</td>
<td>95</td>
<td>554</td>
<td>5.83</td>
<td>6.52</td>
<td>5.8</td>
<td>0.89</td>
<td>1.01</td>
<td>1.13</td>
<td>22%</td>
<td>33%</td>
</tr>
<tr>
<td>LU 09</td>
<td>52</td>
<td>334</td>
<td>6.42</td>
<td>6.98</td>
<td>8</td>
<td>0.92</td>
<td>0.8</td>
<td>0.87</td>
<td>23%</td>
<td>33%</td>
</tr>
<tr>
<td>LU 10</td>
<td>512</td>
<td>5,729</td>
<td>11.19</td>
<td>8.7</td>
<td>6.44</td>
<td>1.29</td>
<td>1.74</td>
<td>1.35</td>
<td>22%</td>
<td>17%</td>
</tr>
</tbody>
</table>

In this paper we will focus on the results for chemistry groups. Next, the results for the medical research groups will be discussed in comparison with 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 ten different universities, they have grown more or less ‘naturally’, and they are not subject to one specific research policy strategy as all these ten 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 research management strategy and 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 ‘reconstructed’ to a
certain extent by the policy of the LUMC as a whole. We will explain the observed

differences of similar distribution functions for both sets in this context.

3. Statistical analysis of bibliometric indicators

3.1 Fundamental distribution functions and application of the central limit theorem

First we present in Fig. 1, as a starting point, the most basic distribution function: the number
of publications as a function of the number of citations. 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 ($P_T$, about 14,000) and count the citations for a time window
of three 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 immediately observe that this distribution function $P_T(C)$ follows a

$$P_T(C) = \alpha C^s$$

(Eq. 3.1)

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:

$$P_T(C) = \alpha C^{-2.4}$$

(Eq. 3.2)

where \(\alpha\) is a constant factor which can be established empirically from the data.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{nlchemistry}
\caption{Distribution function $P_T(C)$: number of publications as a function of number of
citations, for the total set of around 14,000 chemistry publications}
\end{figure}
The lower-C part of the distribution, to which by far the most publications belong, clearly does not follow a power law. We already discussed this phenomenon on the basis of a two-step competition model (van Raan 2001a, b). Thus, in sharp 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 these case: publications).

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, \( P_T(C) \), as presented in Fig. 1. The ‘sample size’ is given by the 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 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 Fig. 2. Remarkably, this distribution for just one research group (which is one of the largest, with around 300 publications) approaches more a ‘complete’ power law dependence than the entire set of publications as presented in Fig. 1.

![Number of publication in a research group (i) as a function of the number of citations](image)

**Fig. 2:** Distribution function \( P_i(C) \): number of publications as a function of the number of citations for one specific research group (\( i \))

The mean 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 skew 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. This is clearly confirmed by the measurement of the CPP distribution for all chemistry research groups, $G(CPP)$, see Figure 4.

**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 $\Delta CPP = 2.00$)
For comparison we also show the much skewer distribution of the total number of citations \( C \) for all groups, \( G(C) \), see Figure 5.

![Number of NL chemistry groups as a function of number of citations](image)

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

For a further understanding of the statistical behaviour 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 number of publications \( P_i \) of a specific research group (i) \(^3\)) and the total number of citation received by this group in a given period of time, \( C_i \). 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:

\[
C(P) = 2.04 P^{1.28}
\] 

(Eq. 3.3)

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 earlier discussed assumption of Katz (1999, 2000) that the ‘Matthew effect’ also works in a sufficiently large set of research groups. We already discussed that size-dependent ‘corrections’ (on the basis of number of publications) of measured impact (on the basis of citations) will lead to an unreasonable levelling off of the impact indicators at the level of research groups as size has to be regarded as an intrinsic characteristic of performance.

\(^3\) The number of publications is a valid measure of size in the statistical context described in this paper. It is, however, a proxy for the ‘real size’ of a research group in terms number of *staff full time equivalents* (fte) available for research. In Appendix 1 we present in Fig. A1.1 the distribution function for research fte’s over the 157 chemistry research groups.
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.

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

Fig. 7a: Distribution function $P(JCS)$: number of publications as a function of $JCS$ (class width $\Delta JCS = 1.0$)
Given the logarithmic scale of the ordinate, we use for simplicity \( N = 0.1 \) in order to include JCS values with zero publications. Notice that in this case we have a semi-logarithmic plot, with a linear abscissa. Only a smaller part of the entire publication population belongs to the very high-value JCS classes (i.e., \( \text{JCS} > 30.0 \)), making the distribution for this high JCS part very noisy. If we restrict the analysis to the publications with values of \( \text{JCS} \leq 30.5 \), we cover 99% of the around 14,000 publications. The distribution function for these publications is shown in Figure 7b. We immediately observe a significant \( (R^2 = 0.85) \) exponential (not power law) relation given by the simple equation (using the parameter \( x \) for \( \text{JCS} \)):

\[
P(x) = \beta \exp[-0.25x]
\]

(Eq. 3.4)

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

A next step is to analyse the relation between the number of citations and journal impact. One has to be very careful in defining what precisely is measured, what 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 the Physical Review Letters is cited ten times (in the given period of time), 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 \( \Delta \text{JCS} = 1.0 \)). Again we have a semi-logarithmic plot, with a linear abscissa. And again only a smaller part of the distribution belongs to the very noisy high-value JCS classes. Similar to the distribution function for the publications, we restrict the analysis to the publications with values of \( \text{JCS} \leq 30.5 \). We cover 90% of the around 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 (not power law) relation given by the equation (using the parameter $x$ for JCS):

$$C(x) = \gamma \exp[-0.19 \times]$$

(Eq. 3.5)

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

**Fig. 8a:** Number of citations as a function of JCS (class width $\Delta \text{JCS} = 1.0$), similar to Fig. 7a, we use for simplicity $C = 0.1$ in order to include JCS values with zero citations.

**Fig. 8b:** Same distribution function as in Fig. 5a, now without the highest JCS values (class width $\Delta \text{JCS} = 1.0$)
We refer to Van Raan (2001a, b) for a thorough discussion of an *ab initio* model (‘two-step competition’) to relate the distribution functions based on publication (*P*), citations (*C*), and journal impact (*JCS*), and to explain their remarkable differences on the one hand and similarities on the other.

The above makes clear that *JCS* is an important additional variable to characterize a population of publications. Research groups make their own choices for journals. Some will try to publish as much as possible in top-journals and other, because of their specialization and often smaller target audience, will mainly publish in journals with a considerable 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 above means 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 foregoing section we have proven 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. In the box below we further illustrate this comparison.

<table>
<thead>
<tr>
<th>Families in a country</th>
<th><em>P</em>, Publications in a large discipline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Income of a family</td>
<td><em>C</em>, Citations received by a publication</td>
</tr>
<tr>
<td>Samples of families</td>
<td>Research groups</td>
</tr>
<tr>
<td>Sample size is number of families in sample</td>
<td>Group size is number of publications in the group</td>
</tr>
<tr>
<td>Average income of families in a sample</td>
<td><em>CPP</em> of a research group</td>
</tr>
<tr>
<td>Education level of families</td>
<td><em>JCS</em> of publications</td>
</tr>
<tr>
<td>Average education level of families in a sample</td>
<td><em>JCSm</em> of a research group</td>
</tr>
</tbody>
</table>

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 *JCS* value, and the research groups can be considered as samples of publications with an average *JCS* value, in our terminology *JSCm*. The distribution of *JSCm* over all research groups will tend to a normal distribution. As in the case of the *CPP* distribution, the *JSCm* 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 paper in a journal), the *JSCm* distribution must be closer to normal than in the case of the *CPP* distribution. These considerations are confirmed by the measurement of the *JSCm* distribution function, see Figure 9.

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 paper in a journal, and then averaged over all journals in a field), the *FCSm* distribution will be again closer to normal than in the case of the *JSCm* distribution. The measured distribution functions clearly confirm these arguments, as we will see further on in this paper.
**Fig. 9:** Distribution function $G(JCSm)$: number of chemistry groups as a function of $JCSm$ values (class width $\Delta JCS = 2.0$)

### 3.2 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 analysed the correlation between the average number of citations per publication ($CPP$) with the average $JCSm$ values for all chemistry research groups. The results of this analysis are presented in Figure 10.

**Fig. 10:** Correlation of $CPP$ with the $JCSm$ values of these groups, for all chemistry groups
We find that this relation on the aggregation level of research groups is described with reasonable significance ($R^2 = 0.75$) by the equation:

$$CPP = 1.13 \cdot JCSm^{0.97} \quad \text{(Eq. 3.6a)}$$

which means 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 Eq. 3.6a and the coefficient $R^2 = 0.75$, in good approximation a normal distribution around a value given by

$$CPP/JCSm \sim 1.1 \quad \text{(Eq. 3.6b)}$$

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

![Number of NL chemistry groups as a function of CPP/JCSm](image)

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

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

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:

$$CPP = 1.53 \times FCSm^{0.94}$$  \hspace{1cm} (Eq. 3.7a)

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 paper (van Raan 2005c) 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.

On the basis of Eq. 3.7a 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

$$CPP/FCSm \sim 1.5$$  \hspace{1cm} (Eq. 3.7b)

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: Fig. 12 shows that the large CPP values correlate with relatively small FCSm values, thus extending the distribution more towards the higher CPP/FCSm side of the distribution. The measured distribution function of $CPP/FCSm$ for all chemistry research groups confirms these findings, see Figure 13.
Number of NL chemistry groups as a function of CPP/FCSm

![Graph showing distribution function G(CPP/FCSm) with class width ΔCPP/FCSm = 0.40.]

**Fig. 13:** Distribution function $G(CPP/FCSm)$: number of chemistry groups as a function of $CPP/FCSm$ values (class width $ΔCPP/FCSm = 0.40$)

Finally, the indicator $JCSm/FCSm$ is a normalization (by field) of an already average measure, thus its distribution will very close to normal, which is demonstrated by Fig. 14.

Number of NL chemistry groups as a function of JCSm/FCSm

![Graph showing distribution function G(JCSm/FCSm) with class width $ΔJCSm/FCSm = 0.20$.]

**Fig. 14:** Distribution function $G(JCSm/FCSm)$: number of chemistry groups as a function of $JCSm/FCSm$ values (class width $ΔJCSm/FCSm = 0.20$)

From Eqs. 3.6b and 3.7b it follows that this distribution will be around
which is also confirmed by Fig. 14.

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

### 3.3 Overview of the results for the medical research groups

Similar to the chemistry research groups, the 65 medical research groups are no ideal ‘samples’, their sizes in terms of numbers of publications \((P)\) also show a skew 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, \(G(CPP)\), Figure 16. The much skewer distribution of the total number of citations \((C)\) for all groups, \(G(C)\) is presented in Figure 17.

As compared to the chemistry research groups, the \(G(P)\) distribution function for the medical research groups is broader at the higher \(P\) side and also the mode of the distribution is more at the higher \(P\) side. For \(G(C)\) we find a similar situation, but even more pronounced. Also several medical research groups with very high numbers of citations are clearly visible. We explain these differences following the discussion at the end of Section 2: the medical research groups belong to one main organization (LUMC) with a policy aiming at reinforcing strong research programmes.

![Number of LUMC groups as a function of number of publications](image)

**Fig. 15:** Distribution function \(G(P): number of medical research groups as a function of number of publications (class width \(\Delta P = 50\))**
**Fig. 16:** Distribution function $G(\text{CPP})$: number of medical research groups as a function of CPP values (class width $\Delta \text{CPP} = 2.0$)

**Fig. 17:** Distribution function $G(\text{C})$: number of medical research groups as a function of total number of citations ($\text{C}$) (class width $\Delta \text{C} = 1,000$)
As discussed in the foregoing section, the distribution of \( \text{JCSm} \) over all research groups will tend to a normal distribution but is not a perfect normal distribution due to 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 \( \text{JCSm} \) distribution function, see Figure 18. We again observe in comparison to the chemical research groups a broader distribution and a mode at the higher \( \text{JCSm} \) side.

As the \( \text{FCS} \) values in the entire population are even more an average than the \( \text{JCS} \) values (namely, the mean value of the number of citations per paper in a journal, and then averaged over all journals in a field), the \( \text{FCSm} \) distribution will be again closer to normal than in the case of the \( \text{JCSm} \) distribution. In this case of the 65 medical research groups, the \( \text{FCSm} \) distribution is almost a perfect normal distribution, see Fig. A1.3 in Appendix 1.

The correlation between ‘size’ (the number of publications \( P_i \) of a research group) and the total number of citation received by a group in a given period of time, \( C_i \) 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:

\[
C(P) = 0.89 \ P^{1.55}
\]  

(Eq. 3.8)

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 Eq. 3.3 (exponent +1.28).
**Fig. 19:** Correlation of the number of citations (C) received per research group with the number of publications (P) of these groups, for all medical research groups.

For the statistical properties of the *normalized* indicators we follow the same approach as 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 ($R^2 = 0.87$):

$$\text{CPP} = 0.66 \text{JCSm}^{1.20} \quad \text{(Eq. 3.9)}$$

which means that also in this 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), which is in contrast to a practically non-existent 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 Eq. 3.9 we can expect for the distribution function $G(CPP/JCSm)$ in good approximation a transformed $JCSm$ distribution

$$CPP/JCSm \sim 0.7 \ JCSm^{0.20}$$  \hspace{1cm} (Eq. 3.10)

whereby the left hand side of the $JCSm$ distribution (see Fig. 18) 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 Eq. 3.10. The mode is even located at 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 skewer distribution for the medical research groups as compared to the chemistry groups.
The correlation between the average number of citations per publication ($CPP$) with the average $FCSm$ values for all medical research groups is presented in Fig. 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:

$$CPP = 0.65 FCSm^{1.30}$$  \hspace{1cm} (Eq. 3.11)

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 Eq. 3.11 we can expect for the distribution function $G(CPP/FCSm)$ in good approximation a transformed $FCSm$ distribution

$$CPP/FCSm \sim 0.7 FCSm^{0.30}$$  \hspace{1cm} (Eq. 3.12)

whereby just as in the foregoing case the left hand side of the $FCSm$ distribution (see Fig. A1.3) 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.
Fig. 22: Correlation of CPP with the FCSm values for all medical research groups

Fig. 23: Distribution function $G(\text{CPP/FCSm})$: number of medical research groups as a function of CPP/FCSm values (class width $\Delta \text{CPP/FCSm} = 0.40$)
As compared to the chemistry research groups, the $G(CPP/FCSm)$ distribution function for the medical research groups is more narrow, similar to what we have found for $G(CPP/JCSm)$.

![Number of LUMC groups as a function of JCSm/FCSm](image)

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

Finally, as an additional finding we present in Fig. 25 the correlation for all medical research groups of the percentage of not-cited publications ($Pnc$, see Table 2 in Section 2.4) and the average number of citations per paper $CPP$. With a reasonable significance ($R^2 = 0.65$) we observe that research groups that are cited more per paper also have more cited papers. This is an interesting indication of the consistency of the citation process to reveal the impact of research groups.
4. 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 a 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 148 chemistry groups covering the work of around 700 senior researchers over a period of ten 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 twelve 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 skew 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 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 behaviour of advanced bibliometric indicators is such...
that meaningful comparison (‘benchmarking’) between groups can be made in terms of reference values based on mean values and variances.

Acknowledgements

The author would like 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


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


van Raan, A.F.J. (2005c). What can we learn from bibliometric distributions and correlations? To be published


version 121104
Appendix 1

**Fig. A1.1:** Distribution function $G(s)$: number of chemistry groups as a function of size, in terms of number of staff full time equivalents (fte) available for research ($s$, in fte; class width $\Delta s = 2.0$)

**Fig. A1.2:** Distribution function $G(FCSm)$: number of chemistry groups as a function of $FCSm$ values (class width $\Delta FCSm = 2.0$)
Fig. A1.3: Distribution function $G(FCSm)$: number of medical research groups as a function of $FCSm$ values (class width $\Delta FCSm = 2.00$)