Average-based versus high- and low-impact indicators for the evaluation of scientific distributions

Pedro Albarrán, Ignacio Ortuño and Javier Ruiz-Castillo

Albarrán *et al* (2009a) introduced a novel methodology for the evaluation of citation distributions using a pair of high- and low-impact measures defined over the set of articles with citations below or above a critical citation level. Albarrán *et al* (2009b) presented the first empirical applications to a situation in which the world citation distribution in 22 scientific fields is partitioned into three geographical areas: the USA, the European Union, and the rest of the world. In this paper, we find that cardinal differences between the results obtained with our high-impact index and the mean citation rate are of a large order of magnitude. When, in addition, the percentage in the top 5% of most cited articles or the percentage of uncited articles are used, there are still important quantitative differences with respect to the high- and low-impact indicators advocated in our approach.

P EDRO ALBARRÁN *et al* (2011a) introduced a novel methodology for the evaluation of research units of a certain size working in the same homogeneous field, namely, a scientific field where the number of citations received by any two papers is comparable independently of the journal where they have been published. The methodology is based on three points.

First, since citation distributions are highly skewed, their upper and lower parts are typically very different.[1] Consequently, it seems useful to describe a distribution by means of two real valued functions defined over the subsets of articles with citations above or below a *critical citation line*. These will be referred to as a *high-* and a *low-impact indicator*, respectively.

Second, a citation distribution can be identified with an income distribution: instead of individuals we have scientific articles, and instead of dollars we have citations. Once this step is taken, it can be seen that the measurement of low-impact coincides with the measurement of economic poverty, where the poor are defined as those individuals whose incomes are below a certain poverty line (which plays an analogous role to the critical citation line for a citation distribution). In turn, it is equally natural to identify the measurement of high-impact with the measurement of a certain notion of economic affluence.

Third, the question of which low-impact indicators might be used in practice can be answered in terms of a family of indices — originally suggested by Foster *et al* (1984) — that satisfies a number of desirable properties, and has been widely used for the measurement of economic poverty in the last 25 years. These same properties lead to the selection of an equally convenient class of high-impact measures.

For this methodology to be of interest in bibliometrics, we must demonstrate that it is useful in practice. Albarrán *et al* (2011b) presented the first empirical application of this approach using a large dataset acquired from Thomson Scientific for the evaluation of the citation impact in three geographical areas:

Pedro Albarrán is at the Departamento de Fundamentos del Análisis Económico, Universidad de Alicante, Campus de San Vicente, 03080 Alicante, Spain. Ignacio Ortuño and Javier Ruiz-Castillo (corresponding author) are at the Departamento de Economía, Universidad Carlos III, Madrid 126, Getafe 28903, Spain; Email: jrc@eco.uc3m.es.

For acknowledgements see page 338.

- The USA;
- The EU, namely, the 15 countries forming the European Union before the 2004 accession; and
- The remaining countries grouped in an area referred to as the rest of the world (RW hereafter).

In this paper we complete the illustrative nature of Albarrán *et al* (2011b) with a comparison of some of its main results with those that can be obtained using average-based and other indicators of scientific performance for the same dataset and the same partition of the world.[2]

The only information required from research units in the approach advocated in Albarrán *et al* (2011a,b) is the homogeneous field to which each unit's publications belong and the number of citations each paper receives. In this scenario, the mean citation rate (MCR hereafter) is a good overall indicator of scientific performance. Consequently, the first empirical exercise in this paper compares the results obtained in Albarrán *et al* (2011b) with those obtained exclusively using the MCR. Ordinal and cardinal issues are typically raised. From an ordinal point of view, two questions are asked:

1. Within each homogeneous field, does it always follow that when a geographical area has a greater MCR than another, the former has also a greater high-impact and a smaller low-impact level than the latter?
2. Within a given geographical area, does the ranking of fields induced by a normalized MCR coincide with the ranking induced by a high- or a low-impact indicator?

From a cardinal point of view, the typical issue refers to the extent to which one geographical area dominates another in a given homogeneous field. For example, the question is: by how much does the USA dominate the EU in physics according to a high-impact indicator and according to the MCR?

It must be recognized that a single statistic of centrality — such as the MCR or the median — may not adequately summarize the asymmetries presented by a typically skewed citation distribution (see *inter alia* Bornmann *et al*, 2008, and, in a different context, Glänzel, 2002). In particular, authors from the Leiden group are very aware of the need to include in their battery of indicators some that capture what takes place at the tails of any citation distribution. Thus, together with average-based indicators, since its inception this group has always taken into account the papers published by a research unit that has received no citations at all (see, *inter alia*, Moed *et al*, 1985, 1988, 1995, and Van Raan, 2004).

More recently, the Leiden group has turned its attention to the upper tail of the distribution, and has introduced the percentage in the top 5% of the most highly cited papers as an indicator of scientific excellence (see Tijssen *et al*, 2002, and Van Leeuwen *et al*, 2003, as well as Aksnes and Sivertsen, 2004).

In this paper we will refer to the MCR and these two indicators as the Leiden triad. In the second empirical exercise of this paper the results in Albarrán *et al* (2011b) for high- and low-impact indicators are compared with the percentage of articles in the top 5% of highly cited papers, and the percentage of articles with no citations at all, respectively. As before, ordinal and cardinal comparisons are performed.[3]

Our choice of a homogeneous field should be clarified at the outset. Naturally, the smaller the set of closely linked journals used to define a given research field, the greater the homogeneity of citation patterns among the articles included must be. Therefore, ideally one should always work at the lowest aggregation level that the data allows. In our case, this may mean the 219 Web of Science categories distinguished by Thomson Scientific. However, articles are assigned to Web of Science categories through the assignment of the journals where they have been published. Many journals are unambiguously assigned to one specific category, but many other typically receive a multiple assignment. As a result, only about 58% of the total number of articles published in 1998–2007 is assigned to a single Web of Science category (see Albarrán *et al*, 2011c).

On the other hand, Thomson Scientific distinguishes between 20 broad fields for the natural sciences and two for the social sciences. Although this firm does not provide a link between the 219 Web of Science categories and the 22 broad fields, Thomson Scientific assigns each article in our dataset to a single broad field. Therefore, given the illustrative nature of our work at this point, both in Albarrán *et al* (2011b) and in this paper a homogeneous field is identified with one of the 22 Thomson Scientific broad fields. In this way, the thorny problems raised by the multiple assignments of articles to Web of Science categories, as well as the difficulties involved in the aggregation from the Web of Science to the broad field level, are provisionally avoided.

The rest of this paper is organized into three sections. Section 2 introduces the indicators and the methods that will be used in the empirical part of the paper. Section 3 presents the data and the empirical findings, while Section 4 discusses the results and offers some conclusions and extensions.

2. Notation, definitions and methods

2.1. Notation

A discrete citation distribution of papers published in a given year is an ordered, non-negative vector $x = (x_1, \ldots, x_i, \ldots, x_n)$, where $x_1 \leq x_2 \leq \ldots \leq x_n$, and $x_i \geq 0$ is the number of citations received by the i-th article over a certain number of years since its publication date — a period known as the citation window. Given a distribution x and a critical citation line, $z \geq 0$, classify as low- or high-impact articles all papers with citation $x_i \leq z$, or $x_i > z$. Denote by $n(x)$ the total

number of articles in the distribution, and by $l(\boldsymbol{x}; z)$ and $h(\boldsymbol{x}; z) = n(\boldsymbol{x}) - l(\boldsymbol{x}; z)$ the number of low- and high-impact articles. A *low-impact index* is a real valued function L whose typical value $L(\boldsymbol{x}; z)$ indicates the low-impact level associated with the distribution $\boldsymbol{x}$ and the critical citation line z, while a *high-impact index* is a real valued function H whose typical value $H(\boldsymbol{x}; z)$ indicates the high-impact level associated with the distribution $\boldsymbol{x}$ and the critical citation line z. We say that a high-impact (low-impact) measure is *monotonic* if one more citation increases (decreases) its value.

2.2. The Leiden triad of indicators

Given a citation distribution $\boldsymbol{x}$, and a critical citation line $z = $ 95th percentile of the world citation distribution in that field, the high-impact index

$$H(\boldsymbol{x}; z = \text{95th}) = h(\boldsymbol{x}; z = \text{95th}^{\text{th}})/n(\boldsymbol{x}) \qquad (1)$$

is the percentage of articles in distribution x belonging to the top 5% of highly cited articles in the world. Similarly, given a citation distribution $\boldsymbol{x}$, and a critical citation line $z = 0$, the low-impact index

$$L(\boldsymbol{x}; z = 0) = h(\boldsymbol{x}; z = 0)/n(\boldsymbol{x}) \qquad (2)$$

is the percentage of articles without citations in that distribution.

It is important to understand that, not being monotonic, the mere percentage of articles satisfying some interesting condition –such as the indices defined in (1) and (2) — only captures what can be referred to as the *incidence* aspect of the phenomenon in question.[4]

To appreciate this limitation, consider a numerical example where z is four, and there are two equal-sized citation distributions with the same number of high-impact articles: (5, 6, 7) for distribution $\boldsymbol{A}$ and (9, 10, 15) for distribution $\boldsymbol{B}$. A measure of high-impact that only takes into account the incidence aspect of that phenomenon will not distinguish between the two research units. For example, in the situation just described $H(\boldsymbol{A}; z = 4) = H(\boldsymbol{B}; z = 4)$ according to the index defined in equation 1. However, a monotonic high-impact measure that takes into account not only the incidence but also the *intensity* aspect of this phenomenon would rank $\boldsymbol{A}$ below $\boldsymbol{B}$. A similar example would show the analogous limitations of the low-impact index $L(\boldsymbol{x}; z = 0)$ introduced in equation 2.

Consider instead the MCR of the distribution $\boldsymbol{x}$, $\mu(\boldsymbol{x})$, defined as

$$\mu(\boldsymbol{x}) = (\Sigma_i x_i)/n(\boldsymbol{x}). \qquad (3)$$

Since the MCR increases with the percentage and the total citations received by high-impact articles, it is sensitive to the incidence and the intensity aspects of a citation distribution. However, to see the limitations of indicators that only capture these two dimensions consider two research units C and D that are equally ranked in terms of both aspects of the high-impact phenomenon, and assume that most of the highly cited papers of unit C have a similar number of citations above the critical citation line, while a large part of the citations received by unit D are concentrated in a few articles of, say, Nobel-prize quality. In such situations it might be desirable that a high-impact measure takes into account the citation *inequality* among the set of high-impact papers, in which case unit D will exhibit a greater high-impact level than unit C. Of course, the MCR fails to respond to distributional changes at either tail of a distribution that maintains the MCR constant.[5]

2.3. The Foster, Greer and Thorbecke family of low- and high-impact indicators

Given a citation distribution $\boldsymbol{x}$ and a critical citation line z, the Foster, Greer and Thorbecke (FGT hereafter) family of low-impact indicators, originally introduced in Foster *et al* (1984) for the measurement of economic poverty, is defined by:

$$L_\beta(\boldsymbol{x}; z) = [1/n(\boldsymbol{x})] \Sigma_i = 1^{l(\boldsymbol{x}; z)} (\Gamma_i)^\beta, 0 \le \beta,$$

where β is a parameter identifying the members of the family, $\Gamma_i = (z - x_i)/z$ is the *normalized low-impact gap* for any article with x_i citations. Observe that $\Gamma_i \ge 0$ for low-impact articles, while $\Gamma_i = 0$ for high-impact articles. The class of FGT high-impact indicators is defined by

$$H_\beta(\boldsymbol{x}; z) = [1/n(\boldsymbol{x})] \Sigma_{i = l(\boldsymbol{x}; z)} + 1^{n(\boldsymbol{x})} (\Gamma^*_i)^\beta, 0 \le \beta,$$

where β is again a parameter identifying the members of the family, $\Gamma^*_i = (x_i - z)/z$ is the *normalized high-impact gap*. In this case, $\Gamma^*_i > 0$ for high-impact articles, while $\Gamma^*_i = 0$ for low-impact articles.

It will be sufficient to understand the differences involved in the use of members of these two classes for parameter values $\beta = 0$, 1, and 2. First, note that the high- and low-impact indices obtained when $\beta = 0$ coincides with the proportion of high- or low-impact papers:

$$H_0(\boldsymbol{x}; z) = h(\boldsymbol{x}; z)/n(\boldsymbol{x}), \qquad (4)$$

and

$$L_0(\boldsymbol{x}; z) = l(\boldsymbol{x}; z)/n(\boldsymbol{x}). \qquad (5)$$

Of course, $H_0(\boldsymbol{x}; z) + L_0(\boldsymbol{x}; z) = 1$, so that if $H_0(\boldsymbol{x}; z)$ changes, then $L_0(\boldsymbol{x}; z)$ must change in the opposite direction. It should be noted that both the high- and low-impact indicators H_0 and L_0 only capture the incidence aspect of both phenomena. Second, consider the high-impact index corresponding to the parameter value $\beta = 1$, or the *per-article high-impact gap ratio*:

$$H_1(\mathbf{x}; z) = [1/n(\mathbf{x})] \; \Sigma_{i = l(\mathbf{x}; z)} + 1^{n(\mathbf{x})} \; \Gamma^*_i. \tag{6}$$

Similarly, the member of the FGT family of low-impact indicators for $\beta = 1$, or the *per-article low-impact gap ratio*, is equal to:

$$L_1(\mathbf{x}; z) = [1/n(\mathbf{x})] \; [\Sigma_{i = 1}^{l(\mathbf{x}; z)} \; \Gamma_i]. \tag{7}$$

Both H_1 and L_1 are seen to capture both the incidence and the intensity of these phenomena. Third, the high- and low-impact members of the FGT families obtained when $\beta = 2$ can be expressed as:

$$H_2(\mathbf{x}; z) = H_0(\mathbf{x}; z)\{[(H_1(\mathbf{x}; z)]^2 + [1 - H_1(\mathbf{x}; z)]^2 \\ (C_H)^2]\}, \tag{8}$$

$$L_2(\mathbf{x}; z) = L_0(\mathbf{x}; z)\{[(L_0(\mathbf{x}; z)]^2 + [1 - L_1(\mathbf{x}; z)]^2 \\ (C_L)^2]\}, \tag{9}$$

where $(C_H)^2$ and $(C_L)^2$ are the squared coefficient of variation (that is, the ratio of the standard deviation over the mean) among the high- and low-impact articles, respectively. In so far as the coefficients of variation C_H and C_L are two measures of citation inequality, the FGT indicators H_2 and L_2 defined in equations 8 and 9 simultaneously cover the incidence, the intensity, and the citation inequality aspects of the high- and low-impact phenomenon they measure.[6]

2.4. Methods

As indicated in the introduction, we distinguish between ordinal and cardinal comparisons. However, for reasons of space only a summary of ordinal comparisons will be included in the final Section 4, while the details are relegated to the Working Paper version of this paper (see Albarrán *et al*, 2010b). As far as cardinal comparisons are concerned, recall from Albarrán *et al* (2011a) that the FGT family of high-impact indicators is decomposable in the sense that, given a critical citation line *z,* a parameter value β, and any partition of a citation distribution $\mathbf{x} = (\mathbf{x^1}, \ldots, \mathbf{x^K})$ into K geographical areas, indexed by $k = 1, \ldots, K$, the overall high-impact measure for $\mathbf{x}$ can be expressed as:

$$H_\beta(\mathbf{x}; z) = \Sigma_k \; \omega_k \; H_\beta(\mathbf{x^k}; z),$$

where $H_\beta(\mathbf{x^k}; z)$ is the high-impact index value for geographical area k, and ω_k is the area's publication share in distribution $\mathbf{x}$. Similarly, the overall low-impact measure can be expressed as:

$$L_\beta(\mathbf{x}; z) = \Sigma_k \; \omega_k \; L_\beta(\mathbf{x^k}; z),$$

where $L_\beta(\mathbf{x^k}; z)$ is the low-impact index value for area k. To adequately interpret the results below, it is important to make it explicit that, from a normative point of view, for any geographical area k it is preferable to have a high $H_\beta(\mathbf{x^k}; z)$ and a low $L_\beta(\mathbf{x^k}; z)$.

In order to quantify the relative situation of any geographical area in a given field, it is convenient to refer to the ratio $\omega_k H_\beta(\mathbf{x^k}; z)/H_\beta(\mathbf{x}; z)$ as area k's observed contribution (OC hereafter) relative to the overall high-impact level for that β. We may ask: what is this area's relative expected contribution (EC hereafter) to that level? Clearly, the answer is its publication share ω_k. Thus, the ratio OC/EC = $H_\beta(\mathbf{x^k}; z)/H_\beta(\mathbf{x}; z)$ is greater than, equal to, or smaller than one as area k's OC is greater than, equal to, or smaller than this area's EC, namely its publication share ω_k. Similarly, the ratio $L_\beta(\mathbf{x^k}; z)/L_\beta(\mathbf{x}; z)$ is greater than, equal to, or smaller than one as area k's OC is greater than, equal to, or smaller than area k's EC, or ω_k.

All indicators in the Leiden triad defined in equations 1, 2, and 3 are also decomposable. For example, for the MCR we have:

$$\mu(\mathbf{x}) = \Sigma_k \; \omega_k \; \mu(\mathbf{x^k}),$$

where $\mu(\mathbf{x^k})$ is the MCR in geographical area k. Note that the observed contribution of any area k relative to the overall level, OC, is equal to the share of total citations, $s(\mathbf{x^k})$:

$$\text{OC} = \omega_k \; \mu(\mathbf{x^k})/\mu(\mathbf{x}) = (\Sigma_i \; x_i^k)/(\Sigma_k \; \Sigma_i \; x_i^k) = s(\mathbf{x^k}),$$

Thus, given that the expected contribution of any geographical area, EC, is again its publication share, ω_k, the ratio

$$\text{OC/EC} = \underline{\mu}(\mathbf{x^k})/\mu(\mathbf{x}) = s(\mathbf{x^k})/\omega_k$$

is greater than, equal to, or smaller than one as the citation share $s(\mathbf{x^k})$ is greater than, equal to, or smaller than this area's publication share ω_k. Similarly, for the indicators $H(\mathbf{x}; z = 95\text{th})$ [and $L(\mathbf{x}; z = 0)$], we have that for any geographical area the ratio of the observed to the expected contribution is greater than, equal to, or smaller than one as the area's share in the top 5% highly cited articles (or the set of uncited articles)] in the world is greater than, equal to, or smaller than the area's publication share ω_k.

3. Empirical results

3.1. The data

In this paper, only research articles or, simply, articles are studied. For reasons explained in the introduction, the key assumption that permits the link between theoretical concepts and the data is the identification of homogeneous fields with the 20 natural sciences and the two social sciences distinguished by Thomson Scientific. After the elimination of observations with missing values for some variables, this paper refers to 3,654,675 articles published in the period 1998–2002 with a five-year citation window. The total number of citations amount to 28,296,113.

Articles are assigned to geographical areas according to the institutional affiliation of their authors as recorded in the Thomson Scientific database on the basis of what has been indicated in the by-line of the publications. In any field, an article might be written by one or more scientists working in only one of the three geographical areas, or it might be co-authored by scientists working in two or three of them. In every internationally co-authored article a whole count is credited to each contributing area.[7]

3.2. A comparison of Foster, Greer and Thorbecke indicators and mean citation rates

As indicated in the introduction, the empirical part of this paper consists of two exercises. The results obtained in Albarrán *et al* (2011b) will be first compared with those obtained using only the MCR. Taking into account the discussion in Section 2.4, we will assess the high-impact level achieved by different geographical areas in any field according to the two approaches by comparing the ratios $\mu(x^k)/\mu(x)$ and $H_\beta(x^k; z)/H_\beta(x; z)$ when $\beta = 0, 1$, and 2, and the critical citation line is equal to the 80th percentile of the world distribution.

The numerical information about $\mu(x^k)/\mu(x)$ is in the first three columns of Table 1, while the information about $H_\beta(x^k; z = 80\text{th})/H_\beta(x; z = 80\text{th})$ for all β is in Table 4A in Albarrán *et al* (2011b), and will not be repeated here. Instead, the comparison between the two approaches is illustrated in Figure 1: in each field and each geographical area, the three vertical bars reflect the ratios $H_\beta(x^k; z = 80\text{th})/H_\beta(x; z = 80\text{th})$ for $\beta = 0, 1$, and 2; the left, center, and right bars are for the USA, the EU, and the RW, respectively; finally, the ratios $\mu(x^k)/\mu(x)$ appear as a horizontal line for each of the three geographical areas.

Let us first observe that, as indicated in Albarrán *et al* (2011b), as β takes on the values 0, 1, and 2, the relative situation of all geographical areas is clearly affected. Specifically, the relative scientific performance of the USA is essentially reinforced for all sciences as β increases. The opposite is the case for the RW, while for the EU the results are more mixed. Nevertheless, except for some important exceptions, the high-impact performance of the EU leaves much to be desired as β increases.

In this scenario, Figure 1 clearly illustrates that judging the USA relative situation in terms of the MCR would seriously alter the results obtained

Table 1. Contributions to world mean citation rates, the top 5% of highly cited articles, and articles without citations by geographical area

	(A)			(B)			(C)		
	USA	EU	RW	USA	EU	RW	USA	EU	RW
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Life sciences									
Clinical medicine	1.27	0.94	0.79	1.44	0.89	0.67	0.75	1.13	1.21
Biology and biochemistry	1.33	0.97	0.74	1.58	0.88	0.60	0.56	0.87	1.57
Neuroscience and behavioral sc.	1.24	0.95	0.77	1.48	0.86	0.58	0.60	1.11	1.46
Molecular biology and genetics	1.25	0.96	0.73	1.40	0.88	0.62	0.69	0.78	1.76
Psychiatry and psychology	1.11	0.93	0.82	1.20	0.89	0.66	0.86	1.11	1.27
Pharmacology and toxicology	1.28	1.04	0.79	1.61	1.03	0.59	0.92	0.95	1.12
Microbiology	1.30	1.02	0.73	1.61	0.95	0.55	0.53	0.80	1.73
Immunology	1.19	0.94	0.84	1.34	0.88	0.73	0.66	1.12	1.39
Physical sciences									
Chemistry	1.51	1.10	0.75	2.04	1.04	0.59	0.59	0.69	1.38
Physics	1.43	1.09	0.76	1.69	1.11	0.63	0.74	0.90	1.26
Computer science	1.34	0.93	0.77	1.54	0.86	0.65	0.87	1.01	1.14
Mathematics	1.25	1.06	0.78	1.45	1.06	0.67	0.87	0.90	1.20
Space science	1.27	0.97	0.75	1.42	0.95	0.63	0.69	1.10	1.38
Other natural sciences									
Engineering	1.19	1.09	0.83	1.39	1.09	0.71	0.95	0.90	1.13
Plant and animal science	1.18	1.13	0.80	1.32	1.21	0.66	0.83	0.85	1.28
Materials science	1.41	1.07	0.83	1.76	1.06	0.72	0.80	0.89	1.16
Geoscience	1.29	1.05	0.76	1.54	0.97	0.64	0.73	0.75	1.49
Environment and ecology	1.15	1.05	0.83	1.35	0.99	0.71	0.92	0.86	1.26
Agricultural sciences	1.26	1.15	0.75	1.45	1.21	0.60	0.74	0.78	1.37
Multidisciplinary	1.91	1.24	0.54	2.33	1.28	0.36	0.63	0.92	1.19
Social sciences									
Social sciences, general	1.11	0.95	0.78	1.24	0.84	0.59	0.94	1.01	1.17
Economics and business	1.25	0.83	0.70	1.46	0.66	0.51	0.86	1.06	1.26

Notes: (A) Contributions to MCRs in area *k*, $MCR(x^k)/MCR(x)$, *k* = USA, EU, RW
(B) Contributions to the top 5% of highly cited articles in area *k*, $H(x^k, z = 95\text{th})/H(x, z = 95\text{th})$, *k* = USA, EU, RW
(C) Contributions to the articles with zero citations in area *k*, $L(x^k, z = 0)/L(x, z = 0)$, *k* = USA, EU, RW
* In any field and any column, a cell value is greater than, equal to, or smaller than one when the geographical area's observed contribution to the world level is greater than, equal to, or smaller than the area's publication share (or expected contribution)
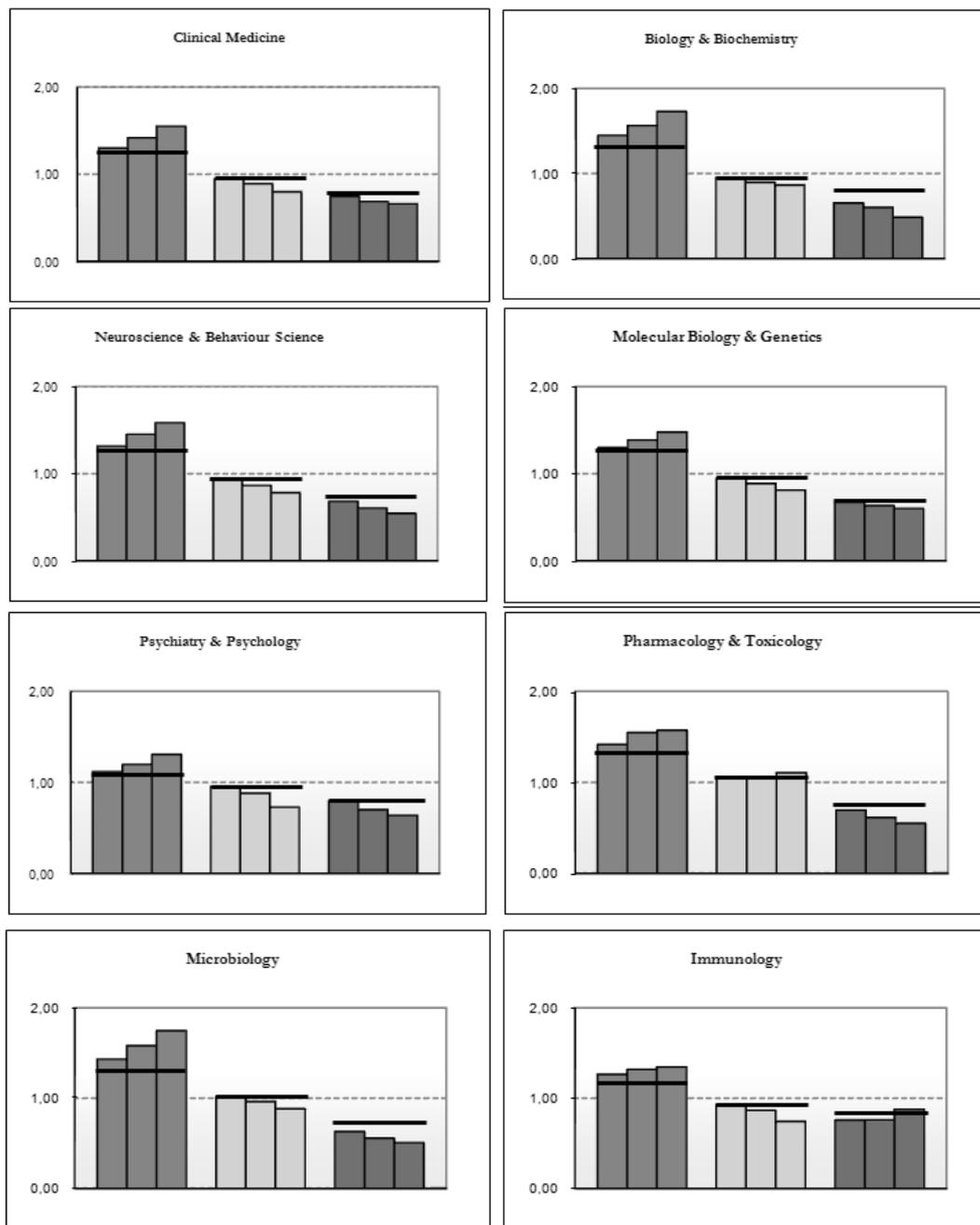
Figure 1. Relative contributions to world high-impact levels by the USA (left bars), the EU (center bars), and the RW (right bars) according to the H_0, H_1, and H_2 members of the Foster, Greer and Thorbecke family of high-impact indicators (*continued*)

Note: Relative mean citation rates appear as horizontal lines

using the high-impact indicators, especially in the case of H_2. For concreteness, columns 1 to 3 in Table 2 present the numerical differences between $H_2(x^k; z = 80\text{th})/H_2(x; z = 80\text{th})$ and $\mu(x^k)/\mu(x)$. The discrepancies in the USA (column 1) amount to a percentage between 8% and 20% in ten fields, between 20% and 30% in nine fields, and between 30% and 40% in the remaining three fields. The difference is of a smaller order of magnitude for four fields in the EU and two fields — immunology and computer science — in the RW. However, the relative situation of the EU (column 2) would be more favorable according to the MCR in 11 fields by a relatively small percentage below 20%, and by a rather large margin above 22% in seven fields.

Finally, the ratio $\mu(x^k)/\mu(x)$ would present the RW (column 3) in a much more favorable position than the ratio $H_2(x^k; z = 80\text{th})/H_2(x; z = 80\text{th})$ in as many as 20 fields.

By way of example, the consequences of these differences in the measurement of the gap between the USA and the EU in every field are explored. Columns 1 and 2 in Table 3 present the ratios $H_2(x^k; z = 80\text{th})/H_2(x^l; z = 80\text{th})$ and $\mu(x^k)/\mu(x^l)$ for $k = $ USA, $l = $ EU, while column 3 presents the difference between these two columns in percentages. The fact that all entries in columns 1 and 2 are greater than one indicates that both measures agree that the USA has a greater high-impact than the EU. It is observed that the USA/EU gap according to our high-impact
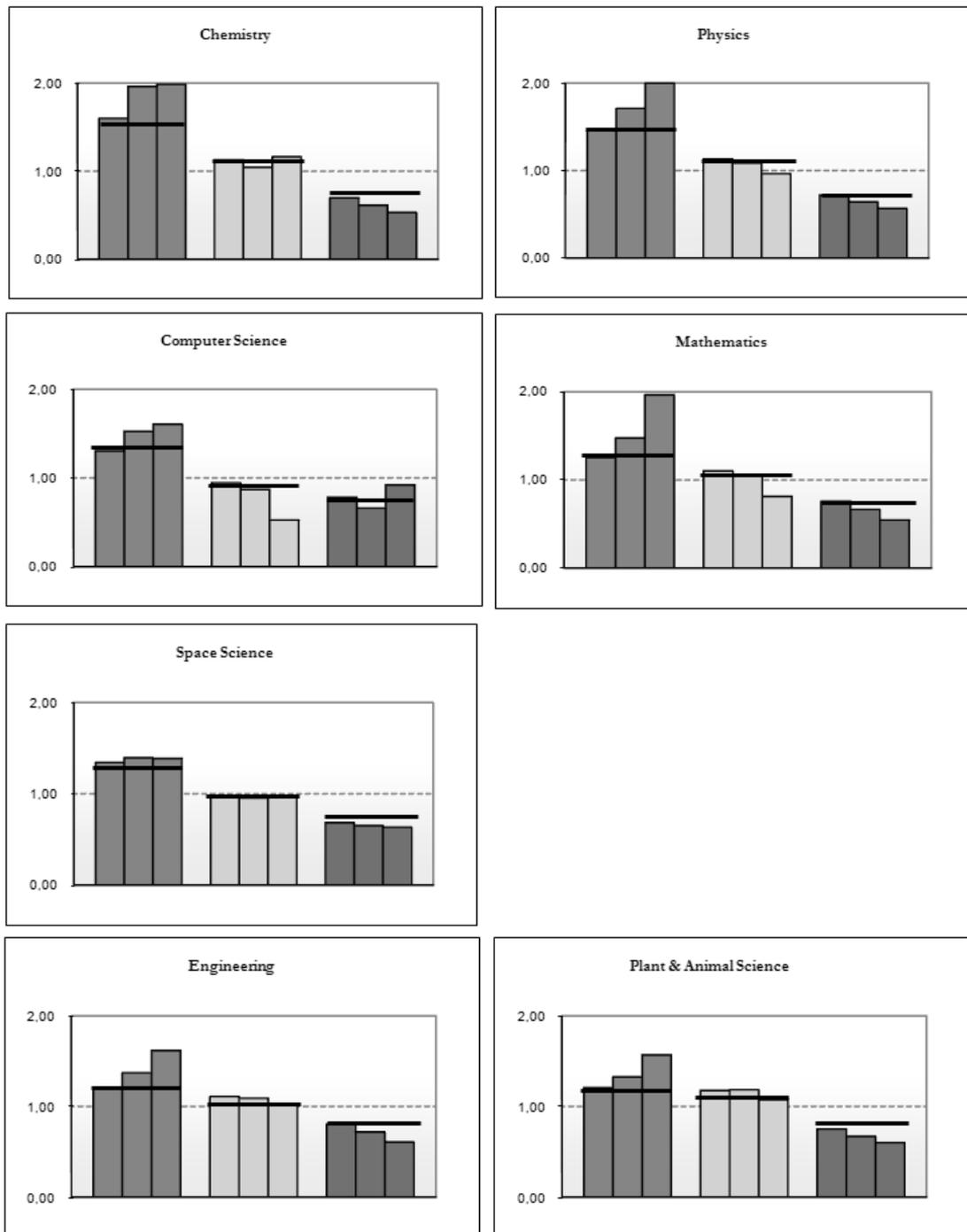
Figure 1 (*continued*)

indicator H_2 is larger in every field, and in 14 out of 22 fields the difference is greater than 30%.

From a cardinal point of view, the conclusion is that using only the MCR generates a very different solution from our approach to the evaluation problem. In the next sub-section, we investigate the consequences of using the remainder of the Leiden indicators.

3.3. A comparison of Foster, Greer and Thorbecke indicators and other Leiden indicators

It remains to explore the possibility of completing the MCR, as is done in the Leiden triad, with the percentage of articles in the top 5% of the world distribution, $H(x^k; z = 95\text{th})$, and the percentage of

articles with no citations at all, $L(x^k; z = 0)$, defined in equations 1 and 2, respectively. We begin with a discussion of high-impact indicators. Since $H(x^k; z = 95\text{th})$ coincides with $H_0(x^k; z = 95\text{th})$, and the impact of changing the critical citation line from the 95th to the 80th percentile is small (see Albarrán *et al*, 2011b, as well as Figure 2 in Albarrán *et al*, 2010b, where the comparison of the two approaches for the two critical citation lines is illustrated). We proceed to compare $H(x^k; z = 95\text{th})/H(x; z = 95\text{th})$ only with $H_2(x^k; z = 95\text{th})/H_2(x; z = 95\text{th})$. Any discrepancy between the two measures can be unambiguously attributed to the fact that the first one does not take into account the intensity and the inequality aspects of the high-impact phenomenon.
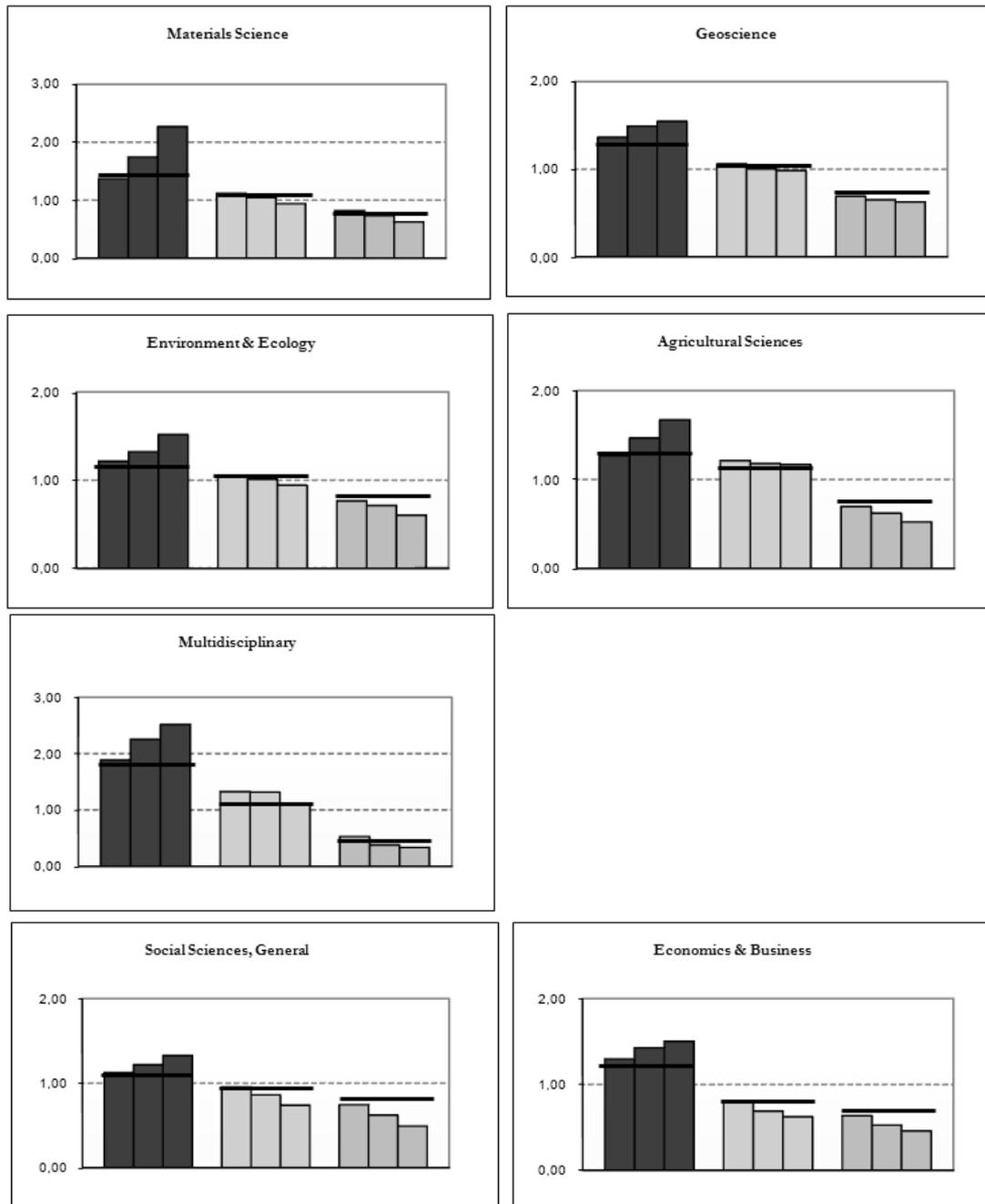
Figure 1 (*continued*)

The main result is that, as with MCRs, in a majority of cases the USA appears worse off according to $H(x^k; z = 95\text{th})/H(x; z = 95\text{th})$. However, the order of magnitude of this phenomenon is rather moderate. To see this, columns 4 to 6 in Table 2 present the differences between the two indicators in every geographical area and in every field. The worsening of the US situation is below 10% in seven fields, and above this percentage in only 12 cases (column 4). The relative position of the EU (column 5) would be improved upon in fewer fields than before, and by a smaller margin (in 11 cases below 20%, and in seven above that percentage). In the RW, the relative situation of immunology and computer science is still better according to H_2, about the same in space science, and it is better

according to the ratio $H(x^k; z = 9\text{th})/H(x; z = 95\text{th})$ in the remaining 19 fields, but by a smaller margin than in column 3 (column 6).

As before, Table 3 explores the consequences of these differences for the USA/EU gap. Columns 4 and 5 present the ratios $H_2(x^k; z = 95\text{th})/H_2(x^l; z = 95\text{th})$ and $H(x^k; z = 95\text{th})/H(x^l; z = 95\text{th})$ for $k =$ USA and $l =$ EU, while column 6 presents the difference between these two columns in percentages. Both measures agree that the USA has a greater high-impact than the EU and — except in four fields, where column 6 is negative — the gap is larger according to our high-impact indicator H_2. This difference is greater than 30% in only six cases.

As far as the low-impact phenomenon is concerned, Figure 2 illustrates the comparison between

Table 2. Differences between the contributions to high- and low-impact levels in area k = USA, EU, and RW according to the two approaches

	(A)			(B)			(C)		
	USA	EU	RW	USA	EU	RW	USA	EU	RW
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Life sciences									
Clinical medicine	18.1	−17.2	−20.3	9.0	−15.4	−2.6	22.7	−8.3	−7.1
Biology and biochemistry	23.2	−12.7	−51.9	11.5	−2.4	−38.8	42.7	11.3	−26.5
Neuroscience and behavioral sc.	21.8	−21.4	−1.4	11.5	−18.1	−13.7	26.6	−8.6	−22.1
Molecular biology and genetics	15.6	−18.0	−22.0	8.7	−14.7	−7.7	16.0	20.8	−42.1
Psychiatry and psychology	14.7	−28.3	−28.0	12.7	−46.3	−8.8	7.5	−6.6	−13.3
Pharmacology and toxicology	18.8	6.2	−44.3	−3.7	11.5	−13.6	−9.4	0.5	0.5
Microbiology	25.5	−15.4	−44.5	14.2	−18.7	−14.0	29.9	14.7	−35.9
Immunology	11.7	−26.0	4.2	1.0	−32.8	26.0	20.0	−7.6	−19.9
Physical sciences									
Chemistry	24.2	5.9	−41.8	−5.6	15.7	−15.8	17.4	19.0	−16.1
Physics	31.2	−12.8	−34.3	22.1	−20.0	−14.8	7.4	2.5	−10.5
Computer science	16.5	−76.6	16.7	4.3	−69.1	30.6	1.8	0.1	−4.2
Mathematics	36.2	−31.2	−44.2	33.3	−50.9	−33.1	2.8	3.3	−6.5
Space science	8.0	0.4	−18.5	−3.6	3.9	0.2	10.4	−8.3	−13.7
Other natural sciences									
Engineering	26.4	−4.78	−36.8	20.7	−9.0	−29.1	−1.4	3.7	−4.7
Plant and animal science	24.9	−4.80	−33.6	22.7	−22.2	−15.1	7.2	5.2	−11.9
Materials science	37.9	−13.5	−33.6	29.9	−19.0	−26.6	5.2	3.8	−6.0
Geoscience	16.6	−6.6	−20.8	0.6	1.8	−3.2	7.6	16.3	−21.5
Environment and ecology	24.2	−11.0	−37.9	17.5	−9.8	−33.2	−0.1	7.4	−11.6
Agricultural sciences	24.3	1.6	−42.9	18.4	−2.9	−30.5	11.4	8.9	−13.9
Multidisciplinary	24.0	−10.7	−62.7	9.8	−22.3	−8.8	9.4	−1.3	−2.8
Social sciences									
Social sciences, general	16.0	−28.0	−58.7	10.6	−24.8	−40.6	0.7	−0.8	−5.0
Economics and business	16.7	−34.0	−52.8	4.2	−7.7	−17.7	3.7	0.0	−8.3

Notes: (A) Difference in % between $H_2(x^k; z = $ 80th$)/H_2(x; z = $ 80th$)$ and $\mu(x^k)/\mu(x)$
(B) Difference in % between $H_2(x^k; z = $ 95th$)/H_2(x; z = $ 95th$)$ and $H(x^k, z = $ 95th$)/H(x, z = $ 95th$)$
(C) Difference in % between $L_2(x^k; z = $ 80th$)/L_2(x; z = $ 80th$)$ and $L(x^k, z = 0)/L(x, z = 0)$

the percentage of uncited articles in each geographical area, $L(x^k; z = 0)/L(x; z = 0)$, and the ratio $L_2(x^k; z)/L_2(x; z)$ when the critical citation line in each field is fixed at the 80th and the 95th percentile of the world citation distribution. The ratios $L_2(x^k; z)/L_2(x; z)$ for the two critical citation lines are depicted as two bars in each area, while the ratio $L(x^k; z = 0)/L(x; z = 0)$ appears as a horizontal line in each case. Note that the change in the critical citation line has again a minimal impact on the relative positions of all areas in every field: only the RW tends to improve somewhat, while the USA tends to slightly worsen as the critical citation line is raised.

From a cardinal point of view the low-impact situations of areas k and l in any field can be compared by means of the ratios $L(x^k; z = 0)/L(x^l; z = 0)$ and $L_2(x^k; z = $ 80th$)/L_2(x^l; z = $ 80th$)$. The more remarkable fact is that according to the percentage of articles without citations the RW is worse off than according to the low-impact indicator. The EU and, above all, the USA are correspondingly better off according to the $L(x^k; z = 0)/L(x^l; z = 0)$ ratio. The differences in every area between $L_2(x^k; z = $ 80th$)/L_2(x; z = $ 80th$)$, and $L(x^k; z = 0)/L(x; z = 0)$ are in columns 7 to 9 in Table 2. They are greater than 20%, approximately, in six fields for the RW

(column 9), four fields for the USA (column 7), and two fields for the EU (column 8). The implications of such differences for the USA/EU gap can be seen in Table 3. Columns 7 and 8 present the ratios $L_2(x^k; z = $ 80th$)/L_2(x^l; z = $ 80th$)$ and $L(x^k; z = 0)/L(x^l; z = 0)$ for k = USA, l = EU and z = 80th percentile, while column 9 presents the difference between these two columns in percentages. Except in engineering, the fact that all entries in columns 7 and 8 are less than one indicates that the USA has a lower low-impact level than the EU. Using the percentage of articles without citations enhances the relative situation of the USA in 16 out of 22 cases, but except in two fields (neurosciences and behavioral sciences, and immunology) the differences between the two approaches are less than 25%.

In brief, in this particular dataset the percentage of articles in the top 5% worsens the relative situation of the USA and boosts that of the RW and, on most occasions, that of the European Union, while the percentage of articles without citations worsens the relative situation of the RW and improves that of the USA In many fields, the two approaches drastically alter the relative situation of different areas.

Table 3. The USA/EU gap according to different indicators

	H_2, $z = 80^{th}$ vs. MCR			H_2, $z = $ 95th vs. H (x, $z = $ 95th)			L_2, $z = 80^{th}$ vs. L (x, $z = 0$)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Life sciences									
Clinical medicine	1.94	1.36	30.2%	2.04	1.62	20.8%	0.81	0.67	17.5%
Biology and Biochemistry	2.00	1.36	31.8%	2.09	1.80	13.7%	0.76	0.65	15.4%
Neuroscience and behav. sc.	2.03	1.31	35.6%	2.30	1.71	25.3%	0.81	0.54	32.7%
Molecular biology and genetics	1.82	1.30	28.5%	2.00	1.58	20.9%	0.83	0.88	−6.2%
Psychiatry and psychology	1.79	1.19	33.5%	2.26	1.35	40.2%	0.89	0.78	13.2%
Pharmacology and toxicology	1.42	1.23	13.4%	1.34	1.57	−17.3%	0.88	0.96	−9.5%
Microbiology	1.98	1.28	35.5%	2.34	1.70	27.5%	0.80	0.66	18.1%
Immunology	1.81	1.27	29.9%	2.05	1.52	25.6%	0.80	0.59	25.7%
Physical sciences									
Chemistry	1.71	1.37	19.4%	1.57	1.96	−25.4%	0.84	0.86	−2.6%
Physics	2.15	1.32	39.0%	2.32	1.52	34.7%	0.87	0.83	5.2%
Computer science	3.06	1.44	52.7%	3.17	1.79	43.6%	0.88	0.87	1.5%
Mathematics	2.42	1.18	51.4%	3.11	1.38	55.7%	0.96	0.96	−0.6%
Space science	1.42	1.31	7.6%	1.38	1.49	−7.8%	0.76	0.62	18.0%
Other natural sciences									
Engineering	1.56	1.10	29.7%	1.75	1.27	27.1%	1.01	1.06	−5.2%
Plant and animal science	1.46	1.05	28.3%	1.73	1.09	36.8%	0.99	0.97	1.8%
Materials science	2.42	1.32	45.2%	2.81	1.66	41.0%	0.91	0.90	0.9%
Geoscience	1.57	1.23	21.8%	1.56	1.58	−1.4%	0.87	0.97	−10.8%
Environment and ecology	1.61	1.10	31.7%	1.83	1.37	25.1%	0.99	1.07	−7.6%
Agricultural sciences	1.43	1.10	23.1%	1.51	1.20	20.5%	0.97	0.94	2.9%
Multidisciplinary	2.24	1.54	31.3%	2.44	1.81	25.9%	0.76	0.68	10.8%
Social sciences									
Social sciences, general	1.79	1.18	34.3%	2.08	1.49	28.5%	0.95	0.94	1.8%
Economics and Business	2.41	1.50	37.9%	2.51	2.22	11.6%	0.84	0.81	3.8%

Notes:
(1) High-impact indicator: $H_2(x^{US}; z = 80th)/H_2(x^{EU}; z = 80th)$
(2) Mean citation ratio: $\mu(x^{US})/\mu(x^{EU})$
(3) = $100[(1) − (2)]/(2)$, i.e. difference in %
(4) High-impact indicator: $H_2(x^{US}; z = 95th)/H_2(x^{EU}; z = 95th)$
(5) Percentage in the top 5% of highly cited articles: $H_2(x^{US}; z = 95th)/H_2(x^{EU}; z = 95th)$
(6) = $100[(4) − (5)]/(5)$, i.e. difference in %
(7) Low-impact Indicator: $L_2(x^{US}; z = 80th)/L_2(x^{EU}; z = 80th)$
(8) Percentage of articles with zero citations: $L(x^{US}, z = 0)/H(x^{EU}, z = 0)$
(9) = $100[(7) − (8)]/(8)$, i.e. difference in %

4. Conclusions and extensions

4.1. Conclusions

Albarrán *et al* (2011a) introduced a novel methodology for the evaluation of citation distributions in terms of a pair of high- and low-impact indicators. Albarrán *et al* (2011b) applied this approach to a situation in which the world is partitioned into the USA, the EU, and the RW using a large sample that covered the 22 broad scientific fields distinguished by Thomson Scientific. This paper has compared some of the main results in the latter with those obtained with alternative methodologies.

We have first examined how far we can go using only the MCR. It turns out that, when the issue is ranking only three geographical areas in each field, except in two cases whenever one area has a greater MCR than another the first has a greater high-impact and a lower low-impact than the second according to the FGT indicators. However, the ranking of fields within a given geographical area according to a normalized MCR coincides with the ranking provided

by the new indicators only for high-impact levels in the USA.[8] For cardinal comparisons, we have confronted the MCR with high-impact indicators.

For a reasonable critical citation line fixed in every field at the 80th percentile of the world distribution, the differences between the results obtained with the two approaches are of a large order of magnitude: the discrepancies are greater than 20% half

Under the mean citation rate criterion the US situation systematically worsens, especially in physics, mathematics, and materials science, while the situation in the EU considerably improves in economics and business, and computer science
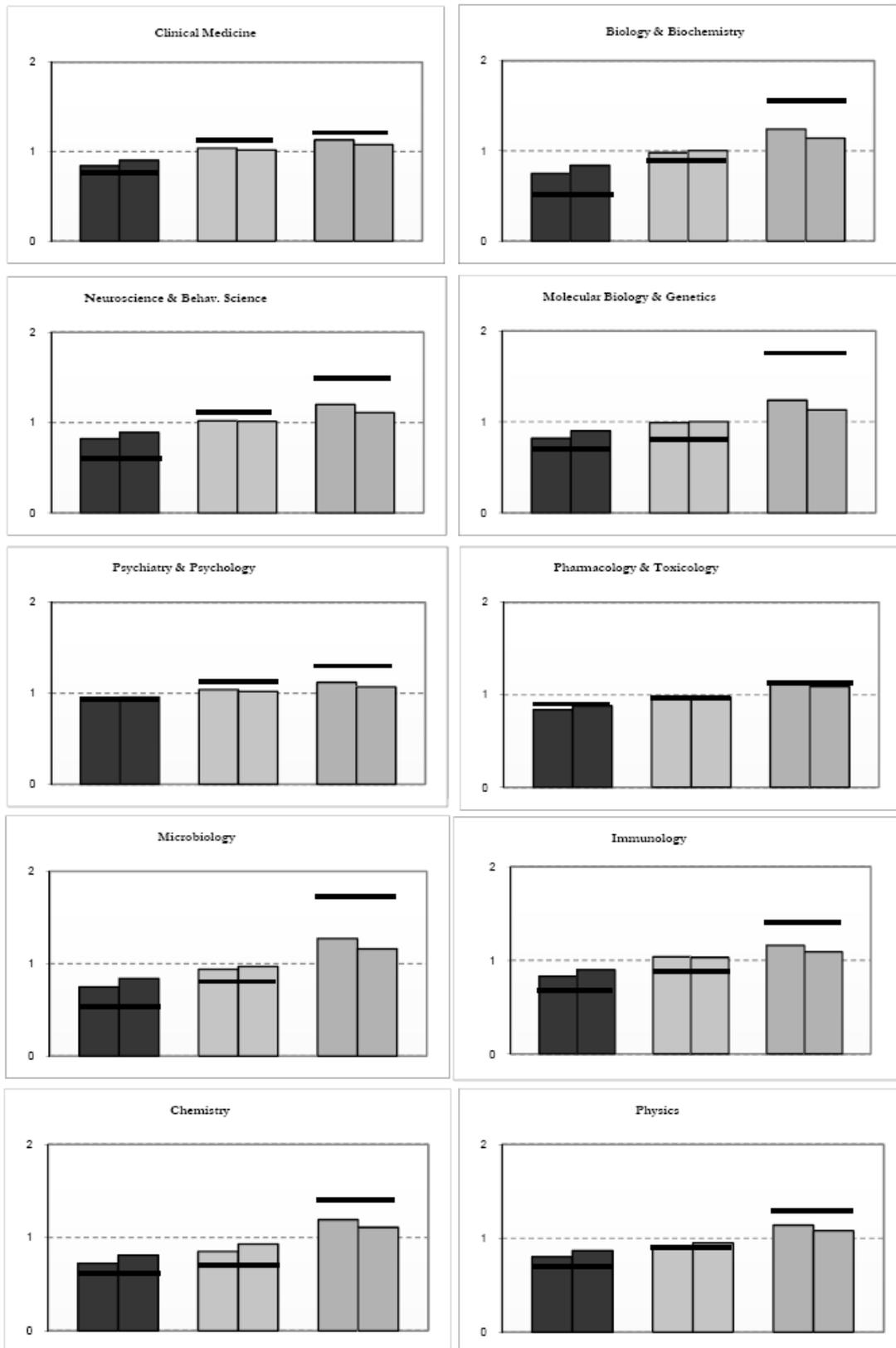
Figure 2. Relative contribution to world low-impact levels by the USA (left bars), the EU (center bars), and the RW (right bars) according to the L_2 low-impact indicator when the critical citation line is equal to the 80th and 95th percentiles (*continued*)

Note: Relative contributions to the percentage of uncited articles appear as horizontal lines

of the time. Consequently, in most fields the view according to both procedures is very different indeed. In particular, under the MCR criterion the US situation systematically worsens, especially in physics, mathematics, and materials science, while the situation in the EU considerably improves in economics and business, and computer science. Correspondingly, the USA/EU gap is greater according to our high-impact indicator in all fields, and by more than 38% in these five fields.
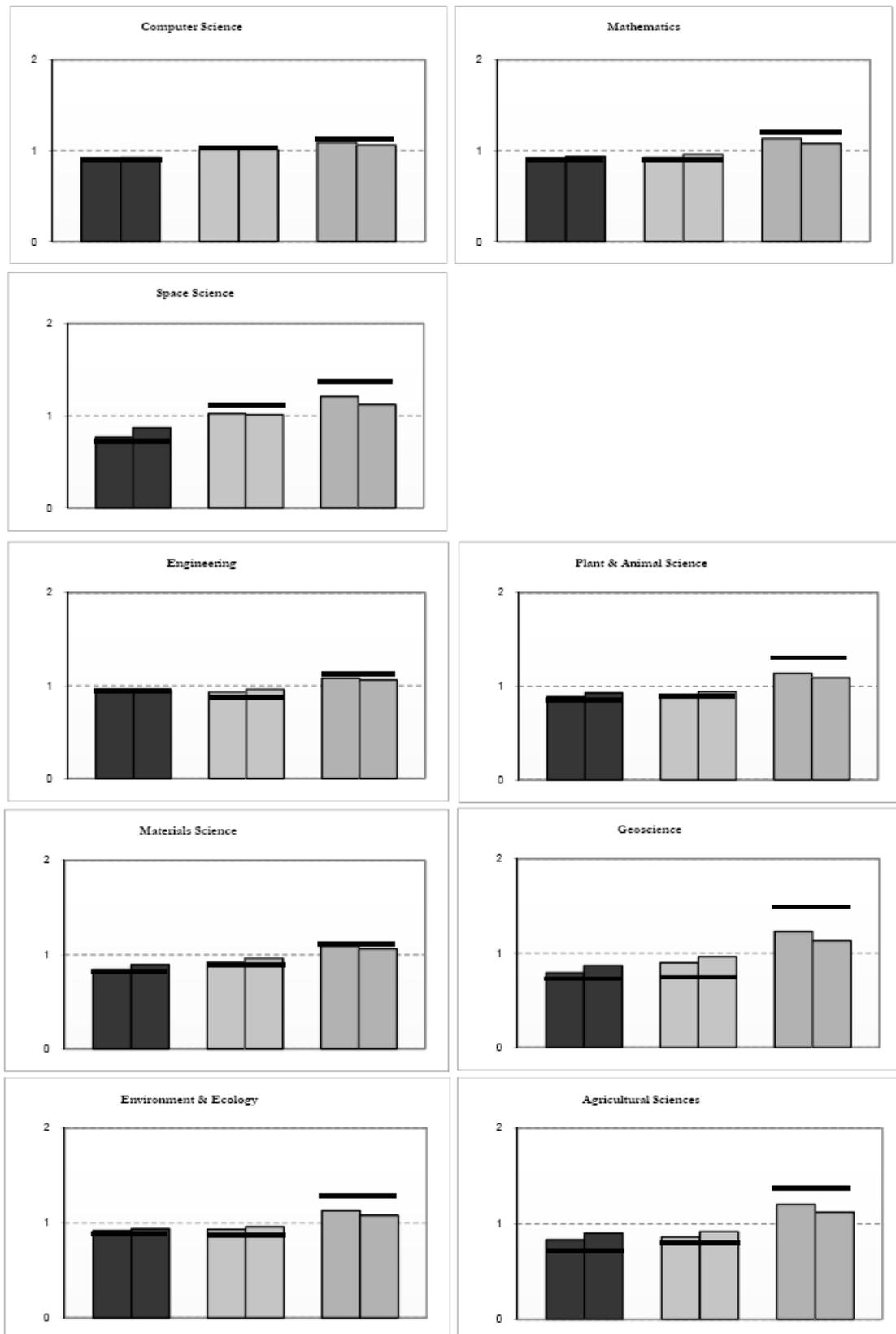
Figure 2 (*continued*)

These results are somewhat at variance with the impression offered in Moed *et al* (1995) for relatively small research units within a field:

Preliminary results suggest that the mean of the distribution correlates rather well to other statistics of the distribution, such as the median, the percentage of papers not cited, and the 90th percentile.

These important differences in the results are the consequence of the differences in the characteristics of the two approaches. In the first place, the MCR is defined over the entire distribution while the
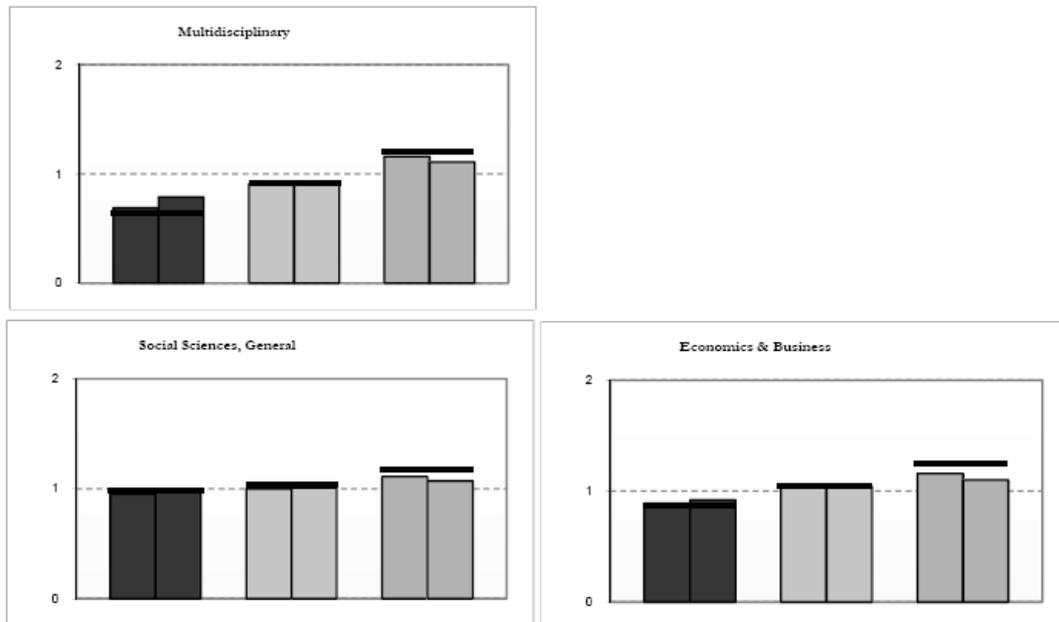
Figure 2 (*continued*)

high-impact indicators are defined only over high-impact articles. In the second place, contrary to the third member of the FGT family of high-impact indicators when $\beta = 2$, the MCR is not sensitive to distributional considerations. In our opinion, one plausible interpretation of the situation is that a mean-based indicator alone does not suffice to adequately represent what takes place in highly skewed citation distributions. This conclusion justifies the use of other, complementary indicators, such as those included in the Leiden triad.

The percentage of articles among the top 5% ranks the three areas in every field exactly as the new high-impact indicators do. Similarly, the percentage of articles without citations orders all areas in every field exactly as our low-impact indicators do. At least for the USA and the EU, the ranking of fields according to these two percentages coincides with the ranking provided by the new high- and low-impact indicators, respectively. However, due to the fact that these percentages do not capture the intensity and the citation inequality of the phenomena in question, there are still large differences between the geographical areas' relative situation according to both approaches.

First, as with the MCR, the relative situation of the USA appears as much weaker when the percentage of articles in the top 5% is used. Correspondingly, the relative situation of the EU and, above all, the RW appears reinforced. In almost one third of the cases the differences between the two approaches are greater than 20%, and there are distortions in about 13 of the 22 fields. The differences between the two approaches in the measurement of the USA/EU gap are greater than 30% in six cases, and between 20% and 30% in 11 additional fields.

Second, differences between the results obtained with the percentage of uncited articles or our preferred low-impact indicator are not that large: they are greater than 20% in only 16.5% of the cases, and there are serious distortions in only six fields. The main impact of using the Leiden rather than the low-impact indicator is to exaggerate the bad situation of the RW. Similarly, the differences between the two approaches in the measurement of the USA/EU gap are greater than 15% in only six fields.

In brief, for the ranking of only three areas in each field, or the ranking of fields within relatively homogeneous areas such as the USA or the EU, following the Leiden or the new approach produces very similar results. However, considerable differences arise when the aim is the cardinal comparison of each area's relative situation. As Moed *et al* (1995) eloquently point out (after evaluating research groups for many years using the MCR and the percentage of articles without citations):

> An important step would in fact be to develop indicators of the impact of a group's very best articles, and compare the results to those obtained by applying the citation per publication ratio.

This is indeed the step taken in important papers that have been already referred to, namely, Tijssen *et al* (2002) and van Leeuwen *et al* (2003). In our view, this is also the step that our methodology has attempted to take by introducing high-impact indicators. The difference, of course, is that we have provided an integrated framework in which the entire citation distribution can be conveniently described by a pair of high- and low-impact indices whose properties have been fully described and have proved to be useful in the empirical work. Furthermore, as indicated in Albarrán *et al* (2011b), the approach can be profitably extended to cover the analysis of inter-temporal trends, as well as the search for dominance results valid for any critical

citation line and any high- or low-impact indicator in a wide class of admissible indices.

4.2 Extensions

1. It should be clear that one of the novelties of the approach advocated here is that members of the FGT families when the parameter β is equal or greater than two vary directly with citation inequality. However, the sensitivity to citation inequality may result in drawbacks of its own. FGT and similar high-impact indices may become very sensitive to extreme observations, namely, to the presence of one or a few articles with a phenomenal number of citations (see Albarrán *et al*, 2011b). More research is needed to learn about the robustness of the results obtained with FGT indicators to the trimming of a handful of observations. At the same time, it might be interesting to use FGT high-impact measures in connection with other indicators that are robust to extreme observations.

2. In the first application of the new approach, it has been natural to exploit the additive decomposability of the FGT indicators to a small partition of the world into three geographical areas. Beyond the extension to the many countries case, the practical interest of the new approach should also be tested in the evaluation of smaller research units, such as the research groups or institutes that have occupied much of the attention of the Leiden group and other authors in bibliometrics.

3. Even more necessary is to address the comment from one referee that the main results of the paper could be translated as that the USA is disproportionally active in the highest-impact subject categories within the 22 broad fields selected. Further research should start by identifying homogeneous fields with Web of Science categories or similar sub-field notions in other datasets, and should also confront the aggregation problems and alternatives from the sub-field to the broad-field level in the presence of the multiple assignment of articles to sub-fields at the lowest aggregation level.

Acknowledgements

Notes

1. See *inter alia* Seglen (1992), Schubert *et al* (1987) for evidence concerning scientific articles published in the period 1981–1985 in 114 sub-fields; Glänzel (2007) for articles published in 1980 in 12 broad fields and 60 middle-sized disciplines; Albarrán and Ruiz-Castillo (2011) for articles published in the period 1998–2002 in 22 broad fields; and Albarrán *et al* (2011c) for these same articles classified in 219 Web of Science categories and three aggregation schemes consisting of a number of disciplines and broad fields.
2. The relationship between high- and low-impact levels and publication shares across geographical areas in each field, and publication efforts across fields in each area, as well as the impact of international co-authorship on each geographical area's scientific performance are topics tackled in Albarrán *et al* (2011b) that are not studied here at all. The impact of changes in critical citation levels is briefly mentioned in Section 3.
3. It should be emphasized that, contrary to what is assumed in this paper, members of the Leiden group usually define their measures relying on information about the journal where each paper is published. This information allows them to compare the observed behavior of relatively small research units, namely research groups, with the expected behavior of the set of journals where the research group is known to publish. The ratio of such expected behavior to the behavior of the journals in the entire field constitutes another interesting indicator in this case. Finally, the possibility of ordering the set of journals in a field in terms of their relative impact allows authors in the Budapest group to graphically represent relative impact indicators

in two-dimensional relational charts (Schubert and Braun, 1986, and Glänzel *et al*, 2002).
4. This is also the only aspect considered in Albarrán *et al* (2010a), where the US and EU performances were compared using a variety of high-impact indices $H(\boldsymbol{x}; z) = h(\boldsymbol{x}; z)/n(\boldsymbol{x})$ for a number of critical citation lines fixed at different percentiles of the world citation distribution for the 22 broad fields distinguished by Thomson Scientific.
5. For a numerical example, assume that the critical citation line is four, and consider two equal-sized citation distributions C and D with the same number of high-impact articles. Let (10, 10) and (5, 15) be the sets of high-impact articles with the same MCR equal to 10. A reasonable high-impact indicator sensitive to distributional considerations would rank unit D strictly above unit C.
6. Albarrán *et al* (2011a) contains a full discussion of all the properties possessed by a number of approaches to the evaluation of research units, including all indicators presented in Sections 2.2 and 2.3.
7. See Albarrán *et al* (2011b) for some descriptive statistics about the number of articles and publication shares by authorship type (Table 1), the classification of articles by scientific field and geographical area (Table 2), the absolute number of citations at the critical citation line in every field, the multiple of the mean that this number represents, and the percentage of the total number of citations received by the high-impact articles in each case (Table 3).
8. Recall from Section 3.2 that a detailed justification of all statements about ordinal comparisons made in this section is available in Albarrán *et al* (2010b).

References

Aksnes, D and G Sivertsen 2004. The effect of highly cited papers on national citation indicators. Scientometrics, 59, 213–224.

Albarrán, P and J Ruiz-Castillo 2011. References made and citations received by scientific articles. *Journal of the American Society for Information Science and Technology*, **62**, 40–49.

Albarrán, P, J Crespo, I Ortuño and J Ruiz-Castillo 2010a. A comparison of the scientific performance of the U.S. and the European Union at the turn of the 21st century. *Scientometrics*, **85**, 329–344.

Albarrán, P, I Ortuño and J Ruiz-Castillo 2010b. Average-based versus High- and Low-impact Indicators for the Evaluation of Scientific Distributions. Working Paper 10-40, Universidad Carlos III.

Albarrán, P, I Ortuño and J Ruiz-Castillo 2011a. The measurement of low- and high-impact in citation distributions: technical results. *Journal of Informetrics*, **5**, 48–63.

Albarrán, P, I Ortuño and J Ruiz-Castillo 2011b. High- and low-impact citation measures: empirical applications. *Journal of Informetrics*, **5**, 122–145.

Albarrán, P, J Crespo, I Ortuño and J Ruiz-Castillo 2011c. The skewness of science in 219 sub-fields and a number of aggregates. *Scientometrics*, **88**, 385–397.

Bornmann, L, R Mutz, C Neuhaus and H-D Daniel 2008. Citation counts for research evaluation: standards of good practice for analyzing bibliometric data and presenting and interpreting results. *Ethics in Science and Environmental Politics*, **18**, 93–102.

Foster, J E, J Greer and E Thorbecke 1984. A class of decomposable poverty measures. *Econometrica*, **52**, 761–766.

Glänzel, W 2002. Coauthorship patterns and trends in the sciences (1980–1998): a bibliometric study with implications for database indexing and search strategies. *Library Trends*, **50**, 461–473.

Glänzel, W 2007. Characteristic scores and scales: a bibliometric analysis of subject characteristics based on long-term citation observation. *Journal of Informetrics*, **1**, 92–102.

Glänzel, W, A Schubert and T Braun 2002. A relational charting approach to the world of basic research in twelve science fields at the end of the second millennium. *Scientometrics*, **55**, 335–348.

Moed, H F and A F J van Raan 1988. Indicators of research performance. In *Handbook of Quantitative Studies of Science and Technology*, ed. A F J van Raan, pp. 177–192. Amsterdam: North–Holland.

Moed, H F, W J Burger, J G Frankfort and A F J van Raan 1985. The use of bibliometric data for the measurement of university research performance. *Research Policy*, **14**, 131–149.

Moed, H F, R E de Bruin and T van Leeuwen 1995. New bibliometrics tools for the assessment of national research performance: database description, overview of indicators, and first applications. *Scientometrics*, **133**, 381–422.

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

Schubert, A, W Glänzel and T Braun 1987. A new methodology for ranking scientific institutions. *Scientometrics*, **12**, 267–292.

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

Tijssen, R, M Visser and T van Leeuwen 2002. Benchmarking international scientific excellence: are highly cited research papers an appropriate frame of reference. *Scientometrics*, **54**, 381–397.

Van Leeuwen, T, M Visser, H Moed, T Nederhof and A van Raan 2003. The holy grail of science policy: exploring and combining bibliometric tools in search of scientific excellence. *Scientometrics*, **57**, 257–280.

Van Raan, A F J 2004. Measuring science. In *Handbook of Quantitative Science and Technology Research*, eds H F Moed, W Glänzel and U Schmoch, pp. 19–50. Dordrecht: Kluwer Academic Publishers.