

The Evaluation of Scientific Production: Towards a Neutral Impact Factor

José María Gómez-Sancho, María Jesús Mancebón-Torrubia

Departamento de Estructura e Historia Económica y Economía Pública, University of Zaragoza (Spain)

Address for correspondence:

José María Gómez Sancho

Departamento de Estructura e Historia Económica y Economía Pública

Facultad de Ciencias Económicas y Empresariales

Universidad de Zaragoza.

Gran Vía, 2

50005 Zaragoza (Spain)

E-mail: jmgomez@unizar.es; mjmance@unizar.es

ABSTRACT

Measurement of research activity still remains a controversial question. The use of the impact factor from the Institute for Scientific Information (ISI) is quite widespread nowadays to carry out evaluations of all kinds; however, the calculation formula employed by ISI in order to construct its impact factors biases the results in favour of knowledge fields which are better represented in the sample, cite more in average and whose citations are concentrated in the early years of the articles.

In the present work, we put forward a theoretical proposal regarding how aggregated normalization should be carried out with these biases, which allows comparing scientific production between fields, institutions and/or authors in a neutral manner. The technical complexity of such work, together with data limitations, lead us to propose some adjustments on the impact factor proposed by ISI which —although they do not completely solve the problem— reduce it and allow glimpsing the path towards more neutral evaluations. The proposal is empirically applied to three analysis levels: single journals, knowledge fields and the set of journals from the Journal Citation Report.

1. INTRODUCCIÓN

Studies related to the measurement of research activity are a constant factor in practically all knowledge fields, whether to elaborate classifications of academic journals, institutions or authors, or to employ it as an explanatory variable in various types of applied studies, such as those aimed at measuring university productive efficiency, national economic growth, technological policy, industry location, behaviour of R&D-intensive industries, etc. (BAR-ILAN, 2008; WILSON, 1999).

Focusing on the field of research evaluation, the development of bibliometric analyses of scientific production constitutes the specific objective of those studies which attempt to establish rankings with regard to authors and varied higher education institutions.

When carrying out these evaluations, the impact factor from ISI (hereafter IFISI) has traditionally been employed as starting point. However, since several authors stand out, their values are biased because they are directly proportional to the number of references per article (KOSTOFF, 2002; SEGLEN, 1997a; WALLIN 2005; ZITT & SMALL, 2008), the time period in which impact is materialized (ROUSSEAU, 2005; SEGLEN, 1997a; SOMBATSOMPOP *et al.*, 2004; WALLIN 2005; ZITT & SMALL; 2008) and, finally, the representation of the field of study within the sample of journals made

by ISI (GARFIELD, 1996; SEGLEN, 1997a; WALLIN 2005). All this means that comparisons among journals, fields or authors carried out from IFISI will be biased, unless a normalization or standardization process is undertaken in order to avoid such biases.

The traditional approach to manage this problem has consisted on undertaking different ex-post-facto normalizations (ZITT & SMALL, 2008). All of them are noted for choosing IFISI as starting point —with no adjustments— and for limiting comparisons to an only field or group of journals (BORDONS & BARRIGON, 1992; MARSHAKOVA-SHAIKEVITCH, 1996; PUDOVKIN & GARFIELD, 2004; SCHUBERT & BRAUN, 1993 AND 1996; SEN 1992; VINKLER, 2002). This approach turns out to be very questionable. Firstly, since it implies considering biased ISI results as valid; and secondly, since the field or journal classifications established by ISI turn out to be quite debatable, given the multidisciplinary nature of its database.

These limitations explain the recent emergence of works which propose that bias correction should be carried out during the impact factor's construction stage (ex-ante normalization). That is the way supported by authors such as SEGLEN (1997a) or WALLIN (2005). The proposal is about constructing an ex-novo impact factor from citations and articles, and comparing the classifications obtained with those provided by IFISI. Thus, SOMBATSOMPOP *et al.* (2004) and ROUSSEAU (2005) create an indicator which attempts to normalize differences in the time period of impact materialization and ZITT & SMALL (2008) propose a methodology which normalizes differences in the propensity to cite.

Although this new ex-ante normalization approach undoubtedly involves a remarkable advance in relation to traditional ex-post normalizations, the proposals developed up to date constitute partial approximations to the achievement of a neutral impact factor between fields and journals, since they focus their attention on the correction of an only bias of those biases involved in the most widely used impact factor: IFISI.

Starting from the last premise, the present article first undertakes a theoretical proposal of how to tackle the construction of an impact factor which jointly avoids biases associated to different citation practices and different field representativity within the sample. The limitations faced by evaluators to obtain the necessary information to undertake it correctly, as well as the technical complexity involved in the empirical application of the proposal, lead us to discard the construction of an impact factor completely neutral in the three biases. Before this situation, the present article proposes another approach, which is necessarily more imperfect and partial but feasible to be undertaken with the data provided by the Institute for Scientific Information: the construction of a new indicator through the application of certain weights on IFISI.

The results obtained indicate that the corrections proposed may drastically change the impacts assigned to each journal by ISI and the differences in impact assigned by ISI for each knowledge field. Thus, one ascertains that the proposed indicator counterbalances the effects that IFISI produces in favour of fields with greater propensity to cite and lower period to materialize impact. It is also ascertained that the journal ranking, which is the result of the application of the impact factor proposed

here, widely differs from that obtained with IFISI, so that the first places are occupied by multidisciplinary or generalist journals. Finally, we also ascertain that our ranking is very similar to those obtained by employing other methodologies (influence weight), which provides the results with greater reliability.

The study is structured in the following way: section 2 puts forward problems associated to impact factors and traditional normalizations, while section 3 presents the methodology proposed to correct the three main biases associated with IFISI, section 4 describes the results obtained by the calculation of a more neutral indicator among those fields provided by IFISI. The study terminates with the traditional conclusions section.

2. PROBLEMS RELATED TO THE USE OF IFISI AND ITS NORMALIZATION

Impact factors are based on the assumption that journals most frequently cited—and which, consequently, make larger contributions to the generation of scientific knowledge— will have higher quality. Due to this conception, its use is gradually increasing as a measurement for the configuration of evaluations of research activity in institutions, authors or journals. At the same time, numerous authors have expressed the limitations raised by its use for this evaluation purpose.

The present epigraph is devoted, firstly, to put forward the problems related to IFISI and, secondly, the problems related to the normalization processes more generally applied to overcome them.

2.1. Criticism to methodology and use of impact factors: The case of IFISI

Despite the fact that the use of impact indices is usually justified on the basis of the relationship established between them and research quality, certain factors—which question whether greater impact is necessarily synonymous with greater quality— must also be taken into account. Some of these objections are generic and refer to the theoretical concept of impact as a measurement of quality (SEGLEN, 1997b). Other doubts concern the specific procedure employed to calculate such indices.

With regard to generic criticism, the following must be underlined: language bias, with an overwhelming predominance of journals in English (MUELLER *et al.*, 2006), seminal articles and review articles receiving excessive citations (AKSNES, 2006; GARFIELD, 1996), the very process of gathering and processing so much information may cause errors (BUCHANAN, 2006; GARFIELD, 1996; JACSÓ, 2006; MOED & VRIENS, 1989) and the inappropriate definition of citable documents (GARFIELD, 1998; GLÄNZEL & MOED, 2002; MOED & VAN LEEUWEN, 1995; MOED *et al.*, 1999). Other criticisms are specifically aimed at the calculation method employed by ISI for impact-factor construction (ADAM, 2002; GLÄNZEL & MOED, 2002; MOED, 2002; SEGLEN, 1997a/b; WHITEHOUSE, 2002).

A further addition to this problem is one of fundamental importance when the measurement of scientific quality is tackled within the specific context of institutional evaluation. Here, we refer to the

biases associated to the impact factor which ISI presents when establishing comparisons whether between institutions, knowledge fields or authors. The importance of this question for the aim of this article requires a more detailed description of this situation.

The biases associated with the impact factor calculated by ISI in evaluation processes stem from three factors: a) the relative importance of the knowledge field in the sample of journals selected by ISI; b) the non-consideration of differences in propensity to cite and c) the non-consideration of differences in the time period in which the impact takes effect (represented by the half-life indicator).

With regard to the relative importance of the knowledge field in the sample of journals, we face one of the greatest deficiencies, as ISI itself admits (GARFIELD, 1996; SEGLEN, 1997b), whether by fields (BUTLER AND VISSER, 2006) or languages (ARCHAMBAULT *et al.*, 2006; VAN LEEUWEN *et al.*, 2001). This is one of the most serious problems, since the indicator is multidisciplinary and the fact that certain fields or journals may not be included or underrepresented will necessarily affect the entire system elaborated by ISI. The results are biased in favour of fields (journals) with greater representation or, in the case of sub-fields, those more closely related to those fields most strongly represented. It is important to underline the difficulty of applying this methodology to Humanities and certain fields of Social Sciences in which academic journals are not the main method for knowledge diffusion (MOED *et al.*, 1998).

Propensity to cite is understood to be the number of citations included in each article and its effects upon the value of the impact factor take place from the moment in which two knowledge fields, journals or authors display, for whatever motive, differentiated behaviour regarding this variable. Indeed, *ceteris paribus*, the fact that a field cites double than others will automatically give it twice the impact (ZITT & SMALL, 2008).

With regard to the biases associated with the time period in which impact is materialized (citations' time distribution), these occur because the impact factor calculated by ISI for each year only takes into account the data available for the two previous years, which works to the detriment of research which, by its very nature, requires longer periods of assimilation (SOMBATSOMPOP *et al.*, 2004).

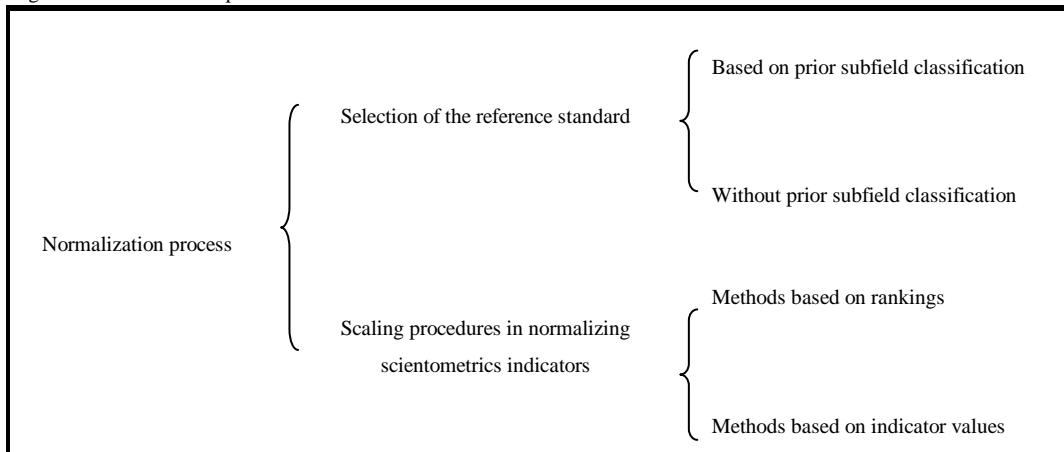
In short, the use of IFISI in comparisons between institutions, fields and/or authors provides biased results in favour of journals belonging to fields better represented in the sample, with greater propensity to cite and a shorter time period for impact materialization. This should not be interpreted as an argument to evade the use of impact factors in quality evaluation. However, it must be remembered that it is essential to refine some of its characteristics in order for its use in making comparisons to be more accurate.

2.2. Criticism to traditional normalization methodologies: Reference standard and scaling procedures

Biases associated to IFISI evaluation processes have lead numerous authors to raise the need to correct it with the aim of making comparisons between articles, journals or institutions more realistic.

With that purpose, they propose to carry out several normalizations to be applied on fields, journals and/or articles. The range of possibilities offered by the articles published up to date is very wide. Traditional normalization approaches are described in SCHUBERT & BRAUN (1996). According to these authors, cross-field normalization of scientometrics indicators demands to face two problems: selection of the reference standard and construction of a proper scaling procedure (see Figure 1).

Figure 1: Normalization process



Source: Elaborated by Gomez-Sancho & Mancebon-Torrubia from Schubert and Braun (1996)

Regarding the selection of the reference standard (selection of the subsample object of comparison), the first option is to resort to traditional classifications such as those provided by ISI itself (standards with prior subfield classification). This approach has been widely criticized by RINIA *et al.* (2001), SCHUBERT & BRAUN (1996) and ZITT *et al.* (2005), who find arbitrariness in the categorization of journals into subfields and especially in its application to multidisciplinary journals, where very differentiated behaviours are observed among both authors and journals, especially in the widest fields (LEYDESDORFF, 2008). The second option for reference standard selection is working without prior subfield classification. In this case, different alternatives are possible. One of these is the use of journals as reference standards (comparison of articles published in the same journal), although articles published in broader scope or multidisciplinary journals also present problems. Another option is the use of related records as reference standards (comparison of articles with the articles cited in them), conceived for individual articles. Finally, another proposal is the use of the set of cited journals as reference standard. This last possibility establishes relations between the journal and other journals which it provides or from which it receives citations. In our opinion, this is the most correct option when selecting an appropriate frame for comparison, since —as SCHUBERT & BRAUN (1996) point out: “These journals were selected by the most authentic persons, the authors themselves, as references; therefore, they can justly be regarded also as reference standards”. The problem is that putting this option into practice is very costly. Besides, it does not allow appropriate treatment of all information contained in the Cited Journal Data Table (both unknown journals included within the field “all others” and many of the relations established with journals without IF). Finally, it should also

be taken into account that even reaching minimum consensus about the most correct frame for comparison, the procedures to normalize the results obtained through IFISI still remain problem.

This last issue is usually tackled through scaling procedures. There are two possible methods: methods based on ranking (the ranks themselves or quantities derived from them serve as a basis for comparison) and methods based on indicator value (applicable if one wants to distinguish the magnitudes of differences between journals). The main disadvantage of all of them is that the starting point for comparisons is original IFISI with no bias correction.

This important limitation has given rise in the last few years to the emergence of certain works proposing the *ex-ante* normalization of some of the previously-mentioned biases. It is about constructing *ex-novo* indicators from original data of citations and articles with varied proceedings which avoid the biases currently affecting IFISI. Thus, SOMBATSOMPOP *et al.* (2004) and ROUSSEAU (2005) present a proposal aimed at normalizing citation age (CHALIF), while ZITT & SMALL (2008) focus their attention on the normalization of the propensity to cite (Audience Factor). This working line is that which, in our opinion, presents greater potentiality when designing an impact factor which is expected to be neutral between fields and/or journals and, therefore, appropriate when undertaking institution, journal and author evaluations. Research presented here can be placed within this new working trend.

Before concluding this section, Table 1 (four first columns) summarizes the different above-detailed normalization proposals.

Table 1: Methods of field normalization of impact.

	Impact factor	Ex-post-facto field-normalized IF	Audience factor	CHALIF	Neutral teórico	IFPCMP
Propensity to cite	not normalized	jointly normalized (a)	normalized	not normalized	normalized	normalized
Aging	not normalized		not normalized	normalized	normalized	normalized
Across-fields citations exchanges	not normalized		not normalized	not normalized	normalized	not normalized

(a) In typical proposals.

Source: Elaborated by Gomez-Sancho & Mancebon-Torrubia from Zitt and Samall (2008)

3.- TOWARDS A MORE NEUTRAL IMPACT FACTOR

In this section, on the basis of a numerical example, we put forward the biases involved by IFISI and propose several alternatives for correction.

3.1. *The problem*

Let us consider the relative situation of two journals: A and B (one per field and, thus, journal and field are equivalent), which will be object of an evaluation in a specific year (t). We assume that both journals have the same objective quality and publish 10 articles per year. Journal A awards 20 citations per article or, in other words, 200 citations per year. For purposes of simplification, we assume that this journal (field) also receives those 200 citations in turn. With regard to citation distribution, we assume that journal A distributes them uniformly over 10 years (20 per year, including the year under evaluation). Journal B is prone to double the number of citations of journal A; that is to say, it awards 400 citations per year (40 per article). By contrast, journal B concentrates its 400

citations received in 5 years (80 citations per year). The first journal may, therefore, belong to a basic science field and the second to a cutting-edge research field. Finally, let us consider that journal A awards 50 per cent of its citations to journals contained in the sample from which the impact factor is calculated (the remainder citations may end up in books, other journals not included in the sample, etc.) This means that only 10 citations per year will be taken into account by IFISI calculation. However, journal B awards 100 per cent of its citations to journals in the sample. In this case, journal A would be identifiable to a journal on Social Sciences or Humanities, while journal B would correspond to one on Sciences.

Using the data above, we shall attempt to isolate the biases causing these differences. Firstly, we calculate the impact factor by using the methodology applied in IFISI:

$$IFISI'_{Journal A} = \frac{10+10}{20} = 1 \quad IFISI'_{Journal B} = \frac{80+80}{20} = 8$$

According to these parameters, it can be concluded that the quality of journal B is eight times greater than that of journal A. Below, we demonstrate that part of the differences attributed by IFISI to both journals is partly due to divergences in the propensity to cite, partly due to the different period for impact materialization and partly due to the different representativity of both fields involved in the sample of journals.

To this end, firstly, the differences in propensity to cite of both journals are corrected. We therefore assume that journal B receives (awards) 20 rather than 40 citations per article (maintaining the same citation period and the representativity of the field in the sample). In this case, this journal reaches 200 citations to be distributed among the first five years (40 citations per year). The impact factor of each journal becomes:

$$IFISI'_{Journal A} = \frac{10+10}{20} = 1 \quad IFISI'_{Journal B} = \frac{40+40}{20} = 4$$

The impact assigned to journal B, which cites twice the number, now becomes four times of that assigned to journal A. This is due to the fact that, prior to correction, receiving cites from journal B, on average, doubles the value of receiving them from A. However, as it can be seen, this is not due to the fact that journal B is of greater quality, but instead solely to a greater number of citations.

Secondly, we raise in an isolated manner the correction of the effect produced by different field representativity in the sample. While all citations in journal B are processed in the calculation of its impact factor, only half of them are processed in the case of journal A. Correction in this case should double the value of the citations which journal A awards to journals in the sample. Thus, we achieve to make equal the value awarded by all journals through their citations to the remainder articles of the journals in the sample.

$$IFISI'_{Journal A} = \frac{20+20}{20} = 2 \quad IFISI'_{Journal B} = \frac{80+80}{20} = 8$$

Like in the previous case, the difference becomes half of that before bias correction.

Thirdly, we undertake isolated normalization of the period of impact materialization. To this end, we assume that journal B distributes the citations it receives in the same period as journal A (10 years), keeping the tendency for initial citation (20 citations per article for journal A and 40 for journal B) and the same ratio of citations to journals included in the sample (50 and 100 per cent for journals A and B, respectively). In this case, journal B will receive its 400 citations uniformly over 10 years (40 citations per year). The impact factor assigned to each journal by ISI is the following:

$$IFISI_{Journal A}^t = \frac{10+10}{20} = 1 \quad IFISI_{Journal B}^t = \frac{40+40}{20} = 4$$

The differences between the quality assigned to journals is once again reduced by half with regard to their initial situation, which demonstrates that part of the impact assigned by ISI is due to differences in the citation period in favour of journals (fields) concentrating it during the first two years.

In short, the facts put forward in this section prove that IFISI favours journals with greater propensity to cite and greater speed in obtaining citations, as well as fields with better sample representation, thus generating distortions in comparisons between knowledge fields following very different citation patterns in these aspects. These three biases should therefore be corrected jointly if one wants to make neutral and impartial comparisons between articles, journals, knowledge fields or institutions.

3.1. Proposal of theoretical ex-ante normalization

In this section we develop a theoretical proposal to undertake individual and joint ex-ante normalization of each of the three previously-analysed biases. While normalization of biases due to propensity to cite and field representativity in the sample requires working on the data contained in each article, normalization of the time period for impact materialization should be carried out on the journals' aggregated data, whether normalized or not. For that reason, we firstly present how, in our opinion, a complete normalization from the articles should be carried out and, subsequently, it is about showing how the journals' aggregated data can be approached. The underlying ideas throughout this process are the following two: that quality values awarded by all articles should be identical and that the indicator should reflect, in a homogeneous manner between fields/journals, the influence of past researches on the works of the year evaluated.

A possible normalization process to make equal the quality value awarded by articles would involve weighting the citations awarded (received) by the value 1/total citations per article. This way, biases caused by differences in the propensity to cite are eliminated.

The normalization required to homogenize field representativity in the sample would be similar to the previous: weighting awarded (received) citations but only including the journals in the sample, so that all of them award the same quality values. This can be carried out by weighting citations with the following factor: 1/total citations of articles published in journals with impact factor.

Finally, the normalization of the impact materialization period demands the indicator to include, in a homogeneous manner, the influence of past researches on the works of the year evaluated. Once the previous two effects have been normalized, a possible normalization process would involve making equal the percentage of citations and articles considered in the impact factor of all journals, instead of homogenizing according to years, as it is carried out by ISI.

In short, joint normalization of the three biases demands the construction of a new *ex-novo* indicator from original data. Firstly, citations per article should be normalized and, once these data have been normalized, new Cited and Citing Journal Data Tables would be generated and applied the normalization of the impact materialization period.

In our opinion, this joint ex-ante normalization of the three biases would avoid problems associated to traditional standardizations. Therefore, it is what we may call a first optimal solution. Nevertheless, the huge technical complexity involved in correcting the millions of citations handled every year prevents its put into practice in evaluation processes.

A more imperfect possibility of carrying out this normalization is undertaking it on the journals' aggregated data.

Regarding differences in propensity to cite, in this case it would be about applying weights to the citations awarded (received) by each journal; that is, on the Cited and Citing Journal Data Tables. The work by ZITT & SMALL (2008) opts for this solution by acting on citations awarded in the first five years and from the citing side. The availability of a rich data set, provided by Thomson Reuters, allows them carrying out this detailed analysis which, nevertheless, turns out to be unfeasible when one does not dispose of such information. In any case, problems derived from differences in the impact materialization period and the fields' different sample representativity still remain in their results.

Regarding normalization of field representativity in the sample, we underline that nowadays we do not know the existence of an indicator which allows carrying it out successfully. For that purpose it would be necessary to dispose of information on the percentage of citations to journals included in the sample. Besides, we should be aware that, in IFISI calculations, citations are obtained from journals and other publications included in the Master Journal List and that within this sample, JCR journals approximately amount a half. Therefore, there are journals which, for IFISI calculation, award citations but whose articles are not taken into account in calculations. We thus face a case similar to the inappropriate definition of citable documents. All this makes that this problem will be ignored throughout the rest of the present article.

Finally, normalization of differences in the impact materialization period may be carried out on normalized or non-normalized data. In our opinion, the process is much more complicated than the previous one, since it makes equal the ex-ante percentage of citations and articles considered (numerator and denominator) in the impact factor of all journals, instead of homogenizing by years, as ISI currently does (the two previous years, both for citations received and articles published). With regard to citations (impact factor numerator), comparison between journals may be performed by

using citations themselves or the time necessary to obtain them (GÓMEZ SANCHO, 2005; ROUSSEAU, 2005; SOMBATSOMPOP *et al.*, 2004). With regard to computed articles (impact factor denominator), its normalization among the different journals is performed, as mentioned above, by equalizing the percentage of journals taken into consideration.

The main restriction when it comes to put this proposal into practice is, once more, the limitations of the existing data and specifically those related to the number of articles published by the different journals throughout their history and, similarly, the changes which have occurred since their appearance (name changes, journal mergers or divisions, etc.)

As it has been pointed out previously, recent works by SOMBATSOMPOP *et al.* (2004) and ROUSSEAU (2005) have attempted to normalize this bias ex-ante. These authors propose a new method for calculating the impact factor by using *Half-Life* (HL). Specifically, their proposal suggests the calculation of an impact factor for each journal which includes in the numerator and denominator the citations received and the articles published in the time period shown by half-life. Taking as reference our numerical example, where HL of journals A and B were of 5 and 2.5 years, respectively, the application of this correction provides the following impact-factor values:

$$CHALIF_{Journal A}^t = \frac{50}{50} = 1 \quad CHALIF_{Journal B}^t = \frac{200}{25} = 8$$

These values coincide with those obtained with the direct application of IFISI, which demonstrates that the proposed correction does not succeed in bias elimination. In our judgement, this is due to the fact that this proposal only homogenizes the numerator (50 per cent of received citations) but not the denominator (50 per cent of articles for journal A and 25 per cent of articles for journal B). If this factor is not taken into account, journal B will continue to be favoured in impact factor calculation, due to the mere fact of concentrating its citations within a shorter time period and, consequently, will ‘need’ fewer articles in spite of the fact that its citation base is identical to that of journal A.

In short, correct comparison of journal quality or knowledge fields through impact factors demands indicators whose construction requires disposing of certain amount of information of which we lack nowadays. However, the need to dispose of indicators to evaluate scientific production is undisputable, especially due to the generalization of the evaluation of the teaching-staff activity in universities, a process which is nowadays taking place in many countries (Spain among them). This need points out that further deepening into the design of procedures to obtain indicators is necessary in order to reduce the above-mentioned problems. With this aim, we put forward the following two sections.

3.2. Towards the correction of differences in the propensity to cite

We must remember that the fact that a field cites double than another will automatically give it twice the impact. Citing to a larger or lesser extent depends on multiple factors —authors, fields, kind of publication, etc. — an aspect whose analysis remains beyond the aim of the present article.

The solution proposed in the previous section to correct this bias consists of equalizing the value awarded to each article via citations, independently of the number of citations associated to it.

In our judgement, this aim may be achieved by penalizing those journals citing above sample average and favouring those awarding a lower-than-average value. The technical impossibility to carry out this procedure directly on articles or journals, lead us to design an indirect correction. The adjustment proposed to correct differences in the propensity to cite is as it follows:

$$\text{IFCPC} = \text{IFISI} * (\text{average of citations of sample}/\text{average of the journal's citations per article}).$$

Where 'IFCPC' stands for the impact factor corrected by the propensity to cite. This proposal is essentially similar to that by ZITT & SMALL (2008), although the term on which weight is applied is different: IFISI in our case, citations in the case of Zitt & Small.

The practical instrumentation of this proposal requires calculating the number of citations per article in the sample of analyzed journals: $30 [(200 + 400)/20]$ in our reference numerical example. The application of these corrections for journals A and B, supposing that field representativity in the sample has been corrected, provides us with the following results:

$$\text{IFCPC (A)} = 2 * (30/20) = 3 \qquad \text{IFCPC (B)} = 8 * (30/40) = 6.$$

As a result of the correction applied, the impact factor of journal B changes from four times that of A (IFISI) to another impact factor (IFCPC) which is only twice as great, thereby eliminating the influence of the propensity to cite, which in field B was twice that of A. Thus, we achieve that citations awarded in each field per article have the same value, eliminating the effect produced if one of them cites to a greater extent than another one.

In any case, it should be taken into account that the lack of citations may be related to the fact that the article is not purely a research article (today's or opinion articles, editorials), which would question the suitability of these 'rising' corrections (ZITT & SMALL, 2008).

3.3. Towards correction in the impact materialization period

The second bias affecting comparability between fields and/or journals, as explained earlier, refers to the existing differences in the time required for impact materialization (citation distribution). This is due to the fact that in articles published "in highly dynamic research fields a large proportion of citations are captured within a short-term; however, fields with a more durable literature have a smaller fraction of short-term citation and hence lower impact factors" (SEGLEN, 1997a).

As it has already been mentioned, the normalization process which may solve this limitation requires the homogenisation of the percentage of citations and articles considered in the impact factor of all journals. An approximation to that would be carrying out some adjustments to IFISI which counteract the bias it involves. Concretely, we propose IFISI weighting with *Half-Life*, which allows favouring journals with longer maturity periods (i.e., greater HL) and works to the detriment of those with more immediate impact (shorter HL), thereby counteracting the effect of the opposite sign produced by the procedure for IFISI calculation. This proposal is similar to those by SOMBATSOMPOP *et al.* (2004) and ROUSSEAU (2005), since HL data are employed to normalize in both cases.

Consequently, the adjustment proposed to correct the citations' time differences is the following:

$$\text{IFCMP} = \text{IFISI} * \text{Half-Life}$$

Where 'IFCMP' denotes the impact factor, once the differences in the materialization period has been corrected. Its application to the referential example —once again supposing that field representativity in the sample has been corrected— provides us with the following results:

$$\text{IFCMP (A)} = 2 * 5 = 10 \quad \text{IFCMP (B)} = 8 * 2.5 = 20.$$

The result of the example is conclusive, since the impact-factor differential between both journals evaluated becomes half of the previous.

Finally, the simultaneous correction of both biases provides an indicator which we term impact factor corrected by the propensity to cite and corrected by the impact materialization period (IFPCMP) and takes the following form:

$$\text{IFPCMP} = \text{IFISI} * (\text{average of citations of the sample/average of the citations per article of the journal}) * \text{Half-Life}$$

As final result of the example, the simultaneous elimination of both biases leads to identical impact factors:

$$\text{IFPCMP (A)} = 2 * (30/20) * 5 = 15 \quad \text{IFPCMP (B)} = 8 * (30/40) * 2.5 = 15$$

4. RESULTS

This section presents the results obtained by applying the methodology recently described in those journals included in ISI's JCR for the year 2000. Firstly, an explanation is given regarding the methodology followed. Empirical analysis is carried out at three different levels: journal (section 4.1), knowledge field (s. 4.2) and set of JCR journals (s. 4.3).

Construction of the indicators proposed in this study requires gathering the information necessary for calculations: filtering the sample of journals, number of articles, citations per article in each journal, ISI impact factors and HL.

Deuration labours (elimination of journals with no impact in the year 2000, duplicated journals, name changes, merged journals, etc.) shaped the final sample of selected JCR journals on which our study is applied into 6855 units.

4.1. Results for a specific journal

In order to explain how the new indicator was constructed, we shall take as an example the *American Economic Review* (hereafter *AER*) in the year 2000. The JCR journal database counts on the following information regarding *AER*: ISSN, total number of citations received, impact factor (IFISI), immediacy factor (IIF), number of articles published, cited HL and citing HL. The remaining information necessary to undertake the adjustments proposed must be processed individually, datum per datum, by the researcher.

Firstly, in order to correct the effect of the propensity to cite, it is necessary —as seen earlier— to calculate both the journal's and the sample's propensity to cite. This information may be calculated on the basis of the data provided by the *Journal Source Data*.

In the case of *AER*, 2214 references were awarded (citable items) in the year 2000 and 169 was the number of articles published in that year, so that its propensity to cite is 19.6 citations per article. With regard to the sample of JCR journals, these values were 21342847 and 739763, respectively, which gives rise to an average propensity to cite of 28.85 citations per article.

The previous data allow calculating the impact factor corrected the propensity to cite (IFCPC):

$$\text{IFCPC (AER)} = \text{IFISI} * (\text{Average citations of the sample} / \text{average citations of the review}) = 1.795 * (28.9 / 19.6) = 2.646.$$

As it can be seen, *AER* benefits from the proposed correction, since it cites below average. IFISI is 1.795, while the adjustment produces an impact of 2.646.

Concerning the correction of the effect produced by differences in the period for impact materialization among the different journals, the need for information increases considerably.

Firstly, it should be taken into account in practice that the HL indicator provided by ISI includes the year evaluated, while IFISI is referred to the two years previous to such period. This time gap advises against carrying out HL weighting directly on IFISI in the calculation of IFCMP. The appropriate weight should be applied on an impact factor which includes both the year evaluated and the two previous ones (ROUSSEAU, 2005). We call IF(3) to the impact factor including the last three years. Its calculation is identical to that of IFISI, —but taking the last three years into account— and may be carried out in the following way.

The JCR database shows that the articles published in 1998 and 1999 by *AER* amounted to 161 and 142, respectively; while 344 and 200 citations were received in these years. The Immediacy Index (IIF), on the other hand, provides information for the year 2000: 169 articles published and 54 citations received. On the basis of the previous data, IF(3) is calculated as it follows:

$$\text{IF(3)} = [(344+200+54) / (161 +142 +169)] = 1.267.$$

We take this factor as a basis for IFCMP and subsequent IFCPCMP construction. In order to calculate both, information on the HL indicator also becomes necessary. Given that the HL of the journal under analysis exceeds 10 years, this indicator is not supplied by ISI and, thus, must be calculated manually. Once again, all the basic information is to be found in the Source Data. Particularly, the *Journal Cited Half-Life* includes the time distribution of received citations. This clearly shows that in the 10 years prior to 2000, the *AER* received 41.29 % of its citations. If we assume, for the sake of simplicity, that citation time distribution is proportional to that achieved throughout the first ten years, we can proxy the time necessary to receive 50 % of its citations is 12.1 years. Another possibility to calculate the HL indicator in case it is not provided by ISI is that suggested by BROOKES (1970) and EGGHE & ROUSSEAU (1990).

IFCMP can be calculated from all the previous information:

$$\text{IFCMP (AER)} = \text{IF(3)} * \text{Half-Life} = 1.267 * 12.1 = 15.33.$$

Given that *AER* needs much more time than average to obtain its citations (see Table 4, section 4.3), the adjustment proposed benefits *AER*.

Finally, the integration of both adjustments carried out on *AER* allows obtaining a new impact factor once the propensity to cite and HL have been corrected, which is called as IFCPCMP and whose value is obtained as follows:

$$\text{IFCPCMP} = \text{IF}(3) * (\text{Average citations of the sample}/\text{Average citations of the journal}) * \text{Half-Life} = \text{IFCMP} * (\text{Average citations of the sample}/\text{Average citations of the journal}) = 1.267 * (28.9/19.6) * (12.1) = 15.33 * (28.9/19.6) = 22.6.$$

Thus, it is demonstrated that the non-consideration of biases caused by citation patterns (propensity to cite and period for impact materialization) may have wide-ranging effects upon the impact-factor values provided by ISI to JCR journals and, therefore, on the evaluation of the research activity of authors and/or institutions. In particular, *AER* is tarnished by ISI calculations, since it has lower propensity to cite than most journals contained in the sample and requires longer time in order to gather 50 % of citations.

4.2. Results among fields

Secondly, in order to show the effects of the new impact factor on the comparison between knowledge fields, we select two fields with very distinct citation patterns regarding propensity to cite and HL. One of these, *Mathematics*, which may be included within basic research, includes 155 journals. The other, *Biophysics*, associated with cutting-edge research, includes 61 journals. The hypothesis is that IFISI benefits cutting-edge rather than basic research, while the impact factor we have created (IFCPCMP) diminishes the existing differences between both fields.

Tables 2 and 3 show 30 journals of both fields; the first column displays the number of articles published in the year 2000 and the second, third and fourth columns show total citations received by the journal (CITED), HL indicator and citations per article (CPA), respectively. The last five columns show the different impact factors: IFISI, IF(3), IFCPC, IFCMP and IFCPCMP, respectively. In this case, in order to improve comparisons, IFCPC is calculated by weighting on IF(3) instead of weighting on IFISI. The last row in Tables 2 and 3 shows the values corresponding to a ‘meta-journal’ of the field constructed by aggregating the data of each of the journals by using the data available. Comparing the values of both tables, one observes that if no correction were undertaken on IFISI, the ‘meta impact’ of the field of *Biophysics* would be almost seven times higher than that of *Mathematics* (2.951 against 0.442); an almost identical difference is also obtained if their IF(3) are compared (2.124 against 0.319). However, when using the impact factor which corrects HL and propensity to cite (IFCPCMP), this difference does not reach 50 % (12.224 against 9.702). This reduction is explained—almost equally— by the greater propensity to cite (somewhat more than double) and the lower period necessary for impact materialization (somewhat less than half) which *Biophysics* journals have against *Mathematics* ones. If only one of the biases is corrected, the differences in impact factor would clearly favour *Biophysics* journals, since their values would triplicate those of *Mathematics* journals

(2.452 against 0.795 in IFCPC and 17.114 against 5.396 in IFCMP), since both biases favour the field of cutting-edge research against the field of basic research and, thus, their impact factors, as calculated by ISI, are far greater.

The impact factor proposed in this work therefore shows the need to correct these biases with the aim of establishing more neutral comparisons between knowledge fields.

Table 2: Differences between Impact Factors in the Biophysics area (Year 2000)

ABR NAME	ART 2000	CITED	HLIFE	CPA	IFISI	IF(3)	IFCPC	IFCMP	IFCPCMP
ARCH BIOCHEM BIOPHYS	566	21117	8.5	38.9	2.576	1.802	1.912	21.896	16.252
BBA-BIOENERGETICS	161	3439	2.8	55.6	6.347	4.132	3.301	17.771	9.242
BBA-BIOMEMBRANES	272	3521	3.9	46.3	2.313	1.657	1.444	9.022	5.631
BIOCHEM BIOPH RES CO	2078	60358	6.3	29.7	3.055	2.095	2.969	19.246	18.705
BIOFIZIKA+	151	1033	11.4	21.4	0.303	0.221	0.410	3.456	4.677
BIOPHYS J	586	22779	5	43.4	4.462	3.394	2.972	22.308	14.859
BIOPOLYMERS	163	6807	10.9	42.6	2.405	1.656	1.630	26.219	17.771
BIORHEOLOGY	37	707	13.1	33.1	0.788	0.472	0.689	10.329	9.023
CHEM PHYS LIPIDS	76	2702	8	45.7	2.328	1.754	1.473	18.625	11.782
FEBS LETT	1254	53658	5.5	29.9	3.440	2.693	3.321	18.922	18.268
INT J BIOMETEOROL	36	314	8.7	26.3	0.652	0.448	0.716	5.674	6.227
J BIOMECH	215	5901	9.9	24.7	1.474	0.963	1.721	14.589	17.040
PHOTOCHEM PHOTOBIOLOG	221	8261	8.1	36.1	2.278	1.624	1.822	18.451	14.762
Q REV BIOPHYS	3	1534	10.4	115.3	6.625	5.632	1.660	68.900	17.265
RADIAT RES	188	6388	7.5	35.4	2.752	2.009	2.248	20.640	16.863
PROG BIOPHYS MOL BIO	18	1539	8.9	118.9	4.931	3.855	1.199	43.886	10.668
J BIOL PHYS	16	109	6.1	21.3	0.512	0.404	0.695	3.124	4.237
J BIOENERG BIOMEMBR	45	2292	5.5	54.5	3.355	2.587	1.778	18.450	9.778
J BIOMECH ENG-T ASME	95	1733	8.5	28.9	0.990	0.685	0.990	8.415	8.415
MET IONS BIOL SYST	20	946	5.1	117.4	4.385	3.288	1.080	22.362	5.507
J BIOCHEM BIOPH METH	84	785	10.9	25.2	0.926	0.527	1.064	10.093	11.595
BBA-GENE STRUCT EXPR	284	2988	3.4	30.9	2.243	1.547	2.095	7.626	7.124
BBA-PROTEIN STRUCT M	277	2568	3.7	48.2	1.688	1.259	1.011	6.245	3.742
BBA-MOL CELL RES	137	2764	3.7	46.4	3.172	2.407	1.975	11.735	7.306
SPATIAL VISION	31	419	7.2	23.7	1.324	0.831	1.613	9.529	11.616
EUR BIOPHYS J BIOPHY	46	1092	5.1	34.8	2.188	1.739	1.818	11.161	9.273
BIOELECTROMAGNETICS	67	1340	5.8	29.7	1.947	1.362	1.896	11.292	10.999
GEN PHYSIOL BIOPHYS	15	281	7.6	55.1	0.417	0.366	0.218	3.167	1.660
INDIAN J BIOCHEM BIO	51	403	7.9	27.8	0.256	0.194	0.266	2.021	2.102
BIOPHYS CHEM	120	2453	6.8	36.1	1.578	1.136	1.262	10.730	8.582
BIOPHYSICS (61)	10231	272808	5.8	34.8	2.951	2.124	2.452	17.114	14.224
ART 2000: articles published in the year 2000									
CITED: total citations received by the journal in articles published in the year 2000									
HLIFE: time period required to receive 50% of the citations									
CPA: cites per article									
IF ISI: impact factor supplied by ISI									
IF(3): impact factor 3 years									
IFCPC: impact factor corrected propensity to cite									
IFCMP: impact factor corrected materialization period									
IFCPCMP: impact factor corrected by the propensity to cite and materialization period									

Table 3: Differences between Impact Factors in the Mathematics area (Year 2000)

ABR NAME	ART 2000	CITED	HLIFE	CPA	IFISI	IF(3)	IFCPC	IFCMP	IFPCMP
MATH NOTES+	211	897	12	10.4	0.166	0.124	0.463	1.996	5.552
ACTA MATH-DJURSHOLM	11	1637	21.6	35.0	1.941	1.489	1.603	41.929	34.622
ADV MATH	83	1803	12.7	22.3	1.118	0.780	1.447	14.199	18.381
ALGEBRA UNIV	39	244	12.1	14.9	0.068	0.044	0.131	0.818	1.581
AM J MATH	49	2230	21.6	21.2	0.823	0.566	1.121	17.775	24.203
AM MATH MON	89	944	20.8	10.0	0.238	0.175	0.689	4.957	14.326
P AM MATH SOC	602	4188	11.6	10.8	0.394	0.262	1.051	4.569	12.191
T AM MATH SOC	309	5981	18	21.1	0.651	0.432	0.894	11.725	16.086
ANN MATH	46	4949	21.6	29.6	1.542	1.131	1.507	33.308	32.559
ARCH MATH	133	926	12.5	10.0	0.305	0.209	0.881	3.810	11.012
ARK MAT	26	357	16.4	28.9	0.511	0.352	0.511	8.382	8.382
B AUST MATH SOC	112	402	8.7	10.0	0.257	0.189	0.742	2.234	6.456
CAN J MATH	54	1280	20.2	20.7	0.476	0.354	0.663	9.611	13.392
CAN MATH BULL	57	375	13	12.0	0.230	0.162	0.554	2.984	7.207
COMMENT MATH HELV	32	857	18.4	20.4	0.646	0.443	0.917	11.889	16.864
COMMUN PUR APPL MATH	46	3462	16.8	22.6	1.674	1.225	2.142	28.122	35.982
COMPOS MATH	68	1000	11.3	17.9	0.600	0.418	0.969	6.780	10.948
CZECH MATH J	73	303	15.8	11.8	0.103	0.100	0.251	1.626	3.967
DIFF EQUAT+	184	1441	11.7	10.0	0.265	0.209	0.765	3.096	8.948
DISCRETE MATH	363	1815	9.1	13.8	0.294	0.218	0.617	2.671	5.613
DUKE MATH J	104	2385	11.1	22.9	0.944	0.662	1.189	10.474	13.199
ANN SCI ECOLE NORM S	31	994	18.8	27.1	0.623	0.464	0.663	11.706	12.470
P EDINBURGH MATH SOC	45	286	8.6	15.9	0.430	0.290	0.780	3.700	6.711
FIBONACCI QUART	58	269	11.2	10.0	0.183	0.129	0.530	2.053	5.934
FUNCT ANAL APPL+	56	1286	21.6	10.0	0.238	0.174	0.688	5.143	14.863
FUND MATH	37	701	21.6	15.7	0.297	0.246	0.548	6.417	11.831
GLASGOW MATH J	48	202	7.6	11.1	0.539	0.358	1.398	4.099	10.628
ILLINOIS J MATH	42	612	21.6	14.5	0.301	0.230	0.600	6.503	12.962
INDAGAT MATH NEW SER	39	119	5.2	12.2	0.200	0.144	0.474	1.040	2.463
INDIAN J PURE AP MAT	134	197	6.9	11.5	0.062	0.043	0.156	0.428	1.074
MATHEMATICS (155)	11891	127544	12.2	16.1	0.442	0.319	0.795	5.396	9.702

ART 2000: articles published in the year 2000

CITED: total citations received by the journal in articles published in the year 2000

HLIFE: time period required to receive 50% of the citations

CPA: cites per article

IF ISI: impact factor supplied by ISI

IF(3): impact factor 3 years

IFCPC: impact factor corrected propensity to cite

IFCMP: impact factor corrected materialization period

IFPCMP: impact factor corrected by the propensity to cite and materialization period

4.3. Results of the sample as a whole

Lastly, we present the results obtained by applying the new indicator to the entire sample of journals. Table 4 shows data which describe the variables for the set of journals: citations received (CITED), articles published in the year 2000 (ART-2000), immediacy impact factor (IIF), impact factor (IFISI), impact factor for three years (IF(3)), *Half-Life* (HLIFE), citations per article (CPA) and impact factor corrected by propensity to cite and impact materialization period (IFPCMP).

Table 4: Descriptive statistics of completed and corrected variables

	N	Minimum	Maximum	Mean	Standard dev.
CITED	6855	1	344256	2495	10720
ART 2000	6855	1	5549	108	201
IIF	6855	0	25.000	0.221	0.540
IFISI	6855	0	50.340	1.275	2.162
IF(3)	6855	0.001	33.548	0.920	1.530
HLIFE	6855	0.5	21.6	7.1	3.4
CPA	6855	10	1334.6	34.2	31.9
IFCPCMP	6855	0	137.679	5.144	6.041

Source: ISI and authors' compilation

Data show great variability both in the journals' total citations received and the number of articles published in the year 2000. The sample's average HL is somewhat over 7 years and citations per article range between 1334 and 10. With regard to impact, IIF is on average almost six times lower than IFISI. Its maximum value (50) doubles that of IIF (25) and in both cases the minimum value is zero. IF(3) is on average somewhat lower (0.917) than IFISI (1.271) and clearly exceeds IIF (0.220) with values ranging between 0.001 and 33.548. Lastly, IFCPCMP displays on average values ranging between 0.003 and 137.679, while standard deviation is close to 6.

Table 5 classifies the top-50 journals on the basis of IFISI, IF(3) and IFCPCMP. In these calculations, we have corrected those journals citing below 10, an idea which was also put into practice by ZITT & SMALL (2008).

In order to ascertain if there are important variations either when employing one or the other indicator or in the classifications drawn, we have calculated the Pearson's correlation coefficients between the IFISI and IFCPCMP quantitative variables, thus obtaining a value of 0.636, while the Spearman's rho and Kendall's tau-b applied to rankings are 0.739 and 0.551, respectively. There are remarkable variations when employing one or the other indicator, in spite of not including bias correction corresponding to differences in the relative importance of the knowledge field in the sample of journals, which undoubtedly affects Social Sciences journals to a greater extent. The results prove that corrections are not disproportionate since no journal ever goes from the first places to the last ones or vice versa.

Individually —and focusing our attention on the classification top-50— the first outstanding fact is the existing similarity between the rankings obtained with IFISI and IF(3), since only two journals featuring among the top-50 list according to IFISI do not appear in the classification drawn with IF(3) (in the classification according to IFISI, these are *Annual Review of Genetics and Clinical Microbiology Reviews*, replaced by *Archives of General Psychiatry and Biochimica et Biophysica Acta-Reviews on Biomembranes* in the classification according to IF(3)). However, variations in ranking positions become noticeable from the fourth position onwards. By contrast, following the bias corrections displayed by IFCPCMP, journal classification is drastically modified: Only 23 journals of those in the previous ranking remain, while 27 new journals enter. Qualitative changes in positions are also drastic. By way of example, *Annals* are significantly penalized by our corrections and attain

much lower positions in the classification (in some of them, it even disappears from the top-50 positions).

Table 5: Ranking of Journals according to IF, IF(3) and IFPCMP

Abbreviated Name	IF	Abbreviated Name	IF(3)	Abbreviated Name	IFPCMP
ANNU REV IMMUNOL	50.340	ANNU REV IMMUNOL	33.548	NATURE	137.679
ANNU REV BIOCHEM	43.429	ANNU REV BIOCHEM	28.414	NEW ENGL J MED	128.274
CELL	32.440	CELL	23.927	SCIENCE	103.929
NAT GENET	30.910	CA-CANCER J CLIN	23.436	LANCET	98.754
NEW ENGL J MED	29.512	NEW ENGL J MED	21.697	NAT GENET	89.925
NAT MED	27.905	NAT GENET	21.080	CELL	88.481
PHYSIOL REV	27.677	NAT MED	20.729	ENDOCRINOLOGY	75.056
ANNU REV NEUROSCI	26.676	PHYSIOL REV	19.263	JAMA-J AM MED ASSOC	66.373
ANNU REV CELL DEV BI	26.300	SCIENCE	18.274	HARVARD BUS REV	62.486
NATURE	25.814	ANNU REV NEUROSCI	17.468	NAT MED	59.950
PHARMACOL REV	25.381	PHARMACOL REV	17.391	BRIT MED J	54.359
CA-CANCER J CLIN	24.674	NATURE	17.305	ADV NUCL PHYS	50.400
SCIENCE	23.872	ANNU REV CELL DEV BI	16.889	ANN INTERN MED	50.213
CURR OPIN CELL BIOL	22.754	MICROBIOL MOL BIOL R	16.632	PHYS TODAY	47.103
ADV CANCER RES	21.680	CURR OPIN CELL BIOL	15.931	ARCH GEN PSYCHIAT	46.598
IMMUNITY	21.083	IMMUNITY	15.341	SCI AM	44.249
MICROBIOL MOL BIOL R	20.639	GENE DEV	14.992	IMMUNOL TODAY	42.168
CHEM REV	20.036	ENDOCR REV	14.545	J NATL CANCER I	42.078
GENE DEV	19.676	TRENDS CELL BIOL	14.004	EVOL BIOL	41.885
ENDOCR REV	19.524	ADV CANCER RES	13.977	J CLIN INVEST	41.728
ANNU REV PHARMACOL	19.289	CHEM REV	12.809	ARCH NEUROL-CHICAGO	40.607
ANNU REV PHYSIOL	18.848	ANNU REV PHARMACOL	12.703	J EXP MED	40.305
TRENDS CELL BIOL	18.815	ANNU REV PHYSIOL	12.394	ARCH DERMATOL	39.734
MOL CELL	18.195	ADV IMMUNOL	12.229	ACCOUNTS CHEM RES	39.466
TRENDS NEUROSCI	17.417	MOL CELL	12.100	P NATL ACAD SCI USA	38.913
ANNU REV BIOPH BIOM	16.194	JAMA-J AM MED ASSOC	12.081	TRENDS BIOCHEM SCI	38.793
JAMA-J AM MED ASSOC	15.402	ANNU REV PLANT PHYS	11.549	BRIT J PSYCHIAT	37.204
J EXP MED	15.236	J EXP MED	11.344	TRENDS PHARMACOL SCI	36.995
ANNU REV PLANT PHYS	15.094	TRENDS NEUROSCI	11.265	GASTROENTEROLOGY	36.789
NEURON	15.081	BEHAV BRAIN SCI	11.079	J CELL BIOL	36.304
IMMUNOL TODAY	14.954	NEURON	10.991	CIRCULATION	36.066
SURF SCI REP	14.952	NAT NEUROSCI	10.975	PHYS REV LETT	35.878
BEHAV BRAIN SCI	14.250	REV MOD PHYS	10.796	ANN NEUROL	35.567
J NATL CANCER I	14.159	ANNU REV BIOPH BIOM	10.471	ANNU REV BIOCHEM	35.197
ANNU REV ASTRON ASTR	14.000	SURF SCI REP	10.250	ANN SURG	34.788
EMBO J	13.999	EMBO J	10.105	COMMUN ACM	34.700
J CELL BIOL	13.955	J CELL BIOL	10.074	AM J MED	34.646
CURR OPIN GENET DEV	13.810	J NATL CANCER I	10.019	ANAL BIOCHEM	32.453
ADV IMMUNOL	13.800	CURR OPIN GENET DEV	9.858	CIRC RES	32.147
ADV PHYS	13.611	IMMUNOL TODAY	9.581	GENE DEV	31.498
ANNU REV GENET	13.450	ACCOUNTS CHEM RES	9.536	J PEDIATR	31.377
ACCOUNTS CHEM RES	13.262	J CLIN INVEST	9.492	NEURON	31.222
TRENDS BIOCHEM SCI	13.246	TRENDS GENET	9.418	CURR OPIN CELL BIOL	31.154
TRENDS GENET	12.912	TRENDS BIOCHEM SCI	9.101	J POLIT ECON	31.127
REV MOD PHYS	12.774	GASTROENTEROLOGY	9.085	ANNU REV IMMUNOL	30.493
NAT NEUROSCI	12.636	CURR OPIN IMMUNOL	9.007	IMMUNITY	30.272
CURR OPIN IMMUNOL	12.549	BBA-REV BIOMEMBRANES	8.762	AM SCI	30.056
GASTROENTEROLOGY	12.246	ANNU REV ASTRON ASTR	8.733	RECENT PROG HORM RES	29.935
CLIN MICROBIOL REV	12.141	ARCH GEN PSYCHIAT	8.729	TECHNOMETRICS	29.926
J CLIN INVEST	12.015	ADV PHYS	8.667	TRENDS GENET	29.922

Source: ISI and authors' compilation

Another outstanding fact is observed when valuating the field of the journals thereby affected. In IFISI almost all (45) of the highest-ranking journals belong to Health Sciences and related fields and, more specifically, to some of its sub-fields (*Biochemistry & Molecular Biology, Immunology, Genetics & Heredity*). It gathers 20 of the highest-ranking journals, since they are included in the most highly represented fields in the sample where the ‘super-cited’ articles are included (GARFIELD, 1996 and 1999). In the classification obtained by using IFCPCMP, journals belonging to the field of Health Sciences continue to form majority; these are lower in number (38) and their concentration in specific fields (now the generic field of *General & Internal Medicine* accumulates most journals: namely 6). Another remarkable fact is that by using IFCPCMP, multidisciplinary or generic journals become particularly important in various fields (medicine or physics). This result is reasonable given the longer time necessary to diffuse knowledge among these fields (RINIA *et al.*, 2001), which was already forecasted by SEGLEN (1997). This is the case of journals such as *Nature, Science, Scientific American, Proceedings of the National Academy of Sciences of the United States of America* or *American Scientist*. As it is logical from the foregoing, the inclusion of journals from the field of Social Sciences (Economics and Business Administration, IT, and Statistics and Probability) is noticeable, something unthinkable when using IFISI.

Another aspect to be underlined is that the results obtained with IFCPCMP are similar to those obtained by BOLLEN, RODRIGUEZ & VAN DE SOMPEL (2006) employing a methodology based on PINSKI & NARIN’s ‘influence weight’ (1976). In their work they obtain a journal-status classification when combining popularity (IFISI) and prestige (Weighted PageRank). Their results (applied to the year 2003) provide a top-10 classification which turns out to be almost identical to that obtained by us: 7 out of the first 10 first journals coincide in both rankings. More precisely, the journals occupying the first three positions coincide: *Nature, England Journal of Medicine* and *Science*, the last two with exchanged positions. Another two journals classified into top-10 positions by BOLLEN, RODRIGUEZ & VAN DE SOMPEL appear in positions 12 and 30 in our classification. We believe that this provides the results with greater consistency, since both methods coincide when valuating the ‘best’ journals, which helps answer the question raised by REEDIJK & MOED (2008): “Do we face a genuine alternative to citation or rather a complementary parameter”.

In short, the empirical analysis performed in the three previous sections clearly shows that the direct use of the impact factor compiled by the Institute for Scientific Information, as employed by university administrative managers in general, provides highly questionable results, since it incorporates significant biases which work in favour of certain fields/journals to the detriment of others.

5. CONCLUSIONS

Measurement of research activity continues to be a controversial question. As time goes by, it seems that minimum consensus is being achieved, such as the recommendation of the use of impact factors, although these are not unanimous.

Focusing on impact factors, this study has underlined the general shortcomings produced by both the employment of these indicators and traditional normalizations used to compare different knowledge fields.

In the case of the impact factor designed by ISI, our first conclusion is that its employment will impair or benefit certain knowledge fields or articles. The system benefits those fields better represented in the sample, those with greatest propensity to be cited and those which, in general, accumulate a higher percentage of citations in the two years prior to the year evaluated. Its indiscriminate use may cause, therefore, undesirable effects, meaning that this question should be treated with rigour and caution. The conclusion to be extracted from the foregoing has important implications. The first of these implications is that IFISI should not be directly or indirectly employed to establish comparisons—whether between universities, research institutes, departments or persons—without the pertinent corrections. This is not an obstacle to recognise that the employment of IFISI, instead of other subjective options, may help improve research evaluation; however, it must be taken into account that it is essential to refine some of its characteristics in order to use it to make comparisons be more accurate.

In this article we put forward a theoretical proposal which attempts to correct the three most important biases when undertaking comparisons. Its application would entail the creation of a new impact factor from normalized original data with the aim that all articles award the same quality level and collect homogeneously the influences of past researches on the works from the year evaluated.

Before the impossibility to put the foregoing into practice, we propose several corrections of IFISI which counteract the effect of two of its three main biases (propensity to cite and impact materialization period). Such adjustments have been performed ‘indirectly’ and approximately, due to data limitations (they were performed with regard to the journal and not to the article or citation) and to the practical impossibility of being undertaken by an individual researcher (thus, the best possible approximations had to be employed in some of the questions). All this requires extreme caution when interpreting the empirical results obtained, which should be interpreted as an illustration of the relevance of biases associated to IFISI and not as a definitive result. Besides, the non-correction of differences in the representativity in the field-wise journal sample invalidates its employment to undertake comparisons, unless some kind of normalization or standardization of this aspect is carried out.

Whatever the case, and bearing in mind the above-mentioned limitations, we consider that our proposal to adjust the indicators which approximate the value of research manages to overcome part of the problems inherent to the impact factor most commonly used worldwide, that supplied by ISI. The results obtained in the empirical section point out, in any case, the relevance of undertaking

adjustments on all biases jointly, thus diminishing differences between knowledge fields or journals strengthened by the IFISI construction method.

We would not wish to end this article without referring to the basic issue behind its elaboration; in practice, and especially in Spain, in recent years the ISI impact factor has become the cornerstone for the evaluation of university lecturers' research, with regard to both their promotion and the evaluation of projects. The limitations of this indicator, as evidenced in this study, are not irrelevant insofar as they may have extremely important consequences for the adequate working of the system of incentives in higher education institutions and, consequently, upon the motivation of their human capital.

REFERENCES

- ADAM, D. (2002), The counting house. *Nature*, 415 : 725–729.
- AKSNES, D. W. (2006), Citation rates and perceptions of scientific contribution. *Journal of the American Society for Information Science and Technology*, 57 (2) : 169–185.
- ARCHAMBAULT, E., VIGNOLA-GAGNE, E., CÔTE, G., LARIVIÈRE, V. & GINGRAS, Y. (2006), Benchmarking scientific output in the social sciences and humanities—The limits of existing databases. *Scientometrics*, 68 (3) : 329–342.
- BAR-ILAN, J. (2008), Informetrics at the beginning of the 21st century – A review. *Journal of Informetrics*, 2 : 1–52.
- BORDONS, M. & BARRIGÓN, S. (1992), Bibliometric analysis of publications of Spanish pharmacologists in the SCI (1984–89). Part II. Contribution to subfields other than “Pharmacology & Pharmacy (ISI)”. *Scientometrics*, 25 (3) : 425–446.
- BROOKES, B. C. (1970), The growth, utility, and obsolescence of scientific periodical literature. *Journal of Documentation*, 26 : 458–461.
- BUCHANAN, R. A. (2006), Accuracy of cited references—The role of citation databases. *College & Research Libraries*, 67 (4) : 292–303.
- BUTLER, L. & VISSER, M. S. (2006), Extending citation analysis to non-source items. *Scientometrics*, 66 (2) : 327–343.
- EGGHE, L., ROUSSEAU, R. (1990), *Introduction to Infometrics. Quantitative Methods in Library, Documentation and Information Science*. Elsevier, Amsterdam.
- GARFIELD, E. (1996), How can impact factors be improved?. *British Medical Journal*, 313 : 411–413.
- GARFIELD, E. (1998), Long-Term Vs. Short-Term Journal Impact: Part II. Cumulative Impact Factors. *The Scientist*, 12 (14) : 12–13.
- GARFIELD, E. (1999), Journal Impact Factor: a brief review. *Canadian Medical Association* 161 (8) : 979–980.
- GLÄNZEL, W., & MOED, H. F. (2002), Journal impact measures in bibliometric research. *Scientometrics*, 53 (20) : 171–193.
- GÓMEZ SANCHO, J. M. (2005), *La evaluación de la eficiencia productiva de las Universidades Públicas Españolas*, tesis doctoral, Universidad de Zaragoza, Spain.
- JACSÓ, P. (2006), Deflated, inflated and phantom citation counts. *Online Information Review*, 30 (3) : 297–309.
- KOSTOFF, R. N. (2002), Citation analysis of research performer quality. *Scientometrics*, 53 (1) : 49–71.
- LEYDESDORFF, L. (2008), Caveats for the Use of Citation Indicators in Research and Journal Evaluations. *Journal of the American Society for Information Science and Technology*, 59 (2) : 278–287.
- MARSHAKOVA-SHAIKEVICH, I. (1996), The Standard Impact Factor as an Evaluation Tool of Science Fields and Scientific Journals, *Scientometrics*, 35 (2) : 283–290.
- MOED, H. F. (2002), The impact-factors debate—The ISI's uses and limits. *Nature*, 415 : 731–732.
- MOED, H. F. & VAN LEEUWEN, T. N. (1995), Improving the accuracy of institute for scientific information's journal impact factors. *Journal of the American Society for Information Science and Technology*, 46 (6) : 461–467.
- MOED, H. F., VAN LEEUWEN, T. N. & REEDIJK, J. (1998), A new classification system to describe the ageing of Scientific journals and their impact factors. *Journal of Documentation*, 54 (4) : 387–419.
- MOED, H. F., VAN LEEUWEN, T. N. & REEDIJK, J. (1999), Towards appropriate indicators of journal impact. *Scientometrics*, 46 (3) : 575–589.
- MOED, H. F. & VRIENS, M. (1989), Possible inaccuracies occurring in citation analysis. *Journal of Information Science*, 15 : 95–107
- MUELLER, P. S., MURALI, N. S., CHA, S. S., ERWIN, P. J. Y GHOSH, A. K. (2006), The association between impact factors and language of general internal medicine journals. *Swiss Medical Weekly*, 136 (27/28) : 441–443.

- PINSKI, G. & NARIN, F. (1976), Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics, *Information Processing & Management*, 12 (5) : 297-312.
- PUDOVKIN A. I., GARFIELD E. (2004), Rank-normalized impact factor: a way to compare journal performance across subject categories. In: *Proceedings of the 67th Annual Meeting of the American Society for Information Science & Technology*. Vol 41. Silver Spring, Md: American Society for Information Science & Technology: pp. 507-515.
- REEDIJK, J. AND MOED, H. F. (2008), Is the impact of journal impactfactors decreasing?, *Journal of Documentation*, 64 (2) : 183-192.
- RINIA, E. J., VAN LEEUWEN, T. N., BRUINS, E. E. W., VAN VUREN, H. G., & VAN RAAN, A. F. J. (2001), Citation delay in interdisciplinary knowledge exchange. *Scientometrics*, 51 (1) : 293-309.
- ROUSSEAU, R. (2005), Median and percentile impact factors: A set of new indicators. *Scientometrics*, 63 (3) : 431-441.
- SCHUBERT, A. & BRAUN, T. (1993), Reference Standards for Citation Based Assessments. *Scientometrics*, 26 (1) : 21-35.
- SCHUBERT, A. & BRAUN, T. (1996), Cross-field normalization of scientometric indicators. *Scientometrics*, 36 (1) : 311-324.
- SEGLIN, P.O. (1997a), Why the impact factor of journals should not be used for evaluating research. *British Medical Journal*, 314 (7079) : 498-502.
- SEGLIN, P.O. (1997b), Citations and journal impact factors: Questionable indicators of research quality. *Allergy*, 52 (11) : 1050-1056.
- SEN B.K. (1992), Documentation note normalized impact factor. *Journal of Documentation*, 48 (3) : 318-325.
- SOMBATSOMPOP, N., MARKPIN, T., & PREMKAMOLNETR, N. (2004), A modified method for calculating the Impact Factors of journals in ISI Journal Citation Reports—Polymer Science Category in 1997-2001. *Scientometrics*, 60 (2) : 235-271.
- VAN LEEUWEN, T. N., MOED, H. F., TIJSSEN, R. J. W., VISSER, M. S. & VAN RAAN, A. F. J. (2001), Language biases in the coverage of the Science Citation Index and its consequences for international comparisons of national research performance. *Scientometrics*, 51 (1) : 335-346.
- VINKLER, P. (2002), Subfield problems in applying the Garfield (Impact) Factors in practice. *Scientometrics*, 53 (2) : 267-279.
- WALLIN, J. A. (2005), Bibliometric Methods: Pitfalls and Possibilities. *Basic & Clinical Pharmacology & Toxicology*, 97 (5) : 261-275.
- WHITEHOUSE, G. H. (2002), Impact factors: facts and myths. *European Radiology*, 12 (4) : 715-717.
- WILSON, C. S. (1999), Informetrics. In: M. E. WILLIAMS (Eds), *Annual review of information science and technology*. Medford, NJ: Information Today, pp. 107-247.
- ZITT, M.; RAMANA-RAHARY, S. & BASSECOULARD, E. (2005), Relativity of citation performance and excellence measures: From cross-field to cross-scale effects of field-normalisation, *Scientometrics*, 63 (2)
- ZITT, M. & SMALL, H. (2008), Modifying the Journal Impact Factor by Fractional Citation Weighting: The Audience Factor, *Journal of the American Society for Information Science and Technology*, 59 (11) : 1856-1860.