EVALUATING RESEARCH USING IMPACT AND HIRSCH FACTORS

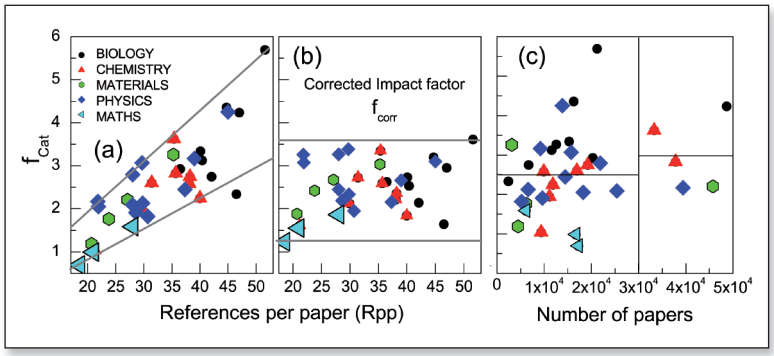
WHY SUCH LARGE DIFFERENCES FOR DIFFERENT FIELDS?

■ **Alicia de Andrés** - Instituto de Ciencia de Materiales de Madrid, Consejo Superior de Investigaciones Científicas,Cantoblanco, 28049 Madrid, Spain - DOI: 10.1051/epn/2011205

Impact factors of the journals and Hirsch index are commonly used to evaluate researchers from close but different areas who are compelled to compete for the same funding and positions. But theses factors depend strongly on the area without a clear foundation. A straightforward normalization would allow comparing different researchers and areas more fairly.

It is widely admitted, from empirical evidence, that the Hirsch indexes [1], *h*, (both, mean values and top values) depend strongly on the area and that only *h* of individuals from the same area should be compared. For example, *h* indexes in biochemistry are, in general, much higher than in mathematics. But the origin of these variations from one area to the other is unclear. **Hirsch index has been criticised** and more or less complicated alternative indexes have been proposed [2,3]. But, in fact, the ***h* index is used everyday and everywhere because it is simple and easily obtained from databases.**

Two issues have been raised to explain the disparity of *h* indexes between different areas: the size of the community (the larger it is, the higher *h*) and the impact factors of the journals of the area. Actually, what the ►

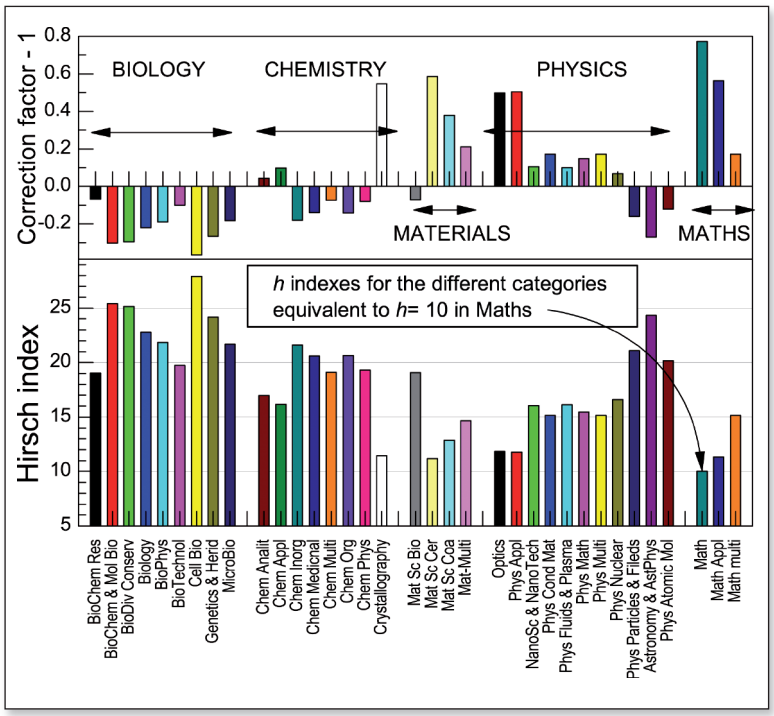


▲ **FIG. 1:** a) Aggregate impact factors (f_{Cat}) vs. average number of references per publication (*Rpp*) for 35 categories from JCR. The different colours correspond to categories grouped into areas (3 in Maths, 11 in Physics, 4 in Materials Sciences, 8 in Chemistry and 9 in Biology). b) Corrected impact factors as explained in the text. c) Aggregate impact factors vs. the number of publications in the category.

► Journal Citation Reports (JCR) from the ISI Web of Knowledge (WoK) calls "Aggregate impact factor" (f_{Cat}) of a category (which is the average impact factor of the journals of this category weighted by the number of articles in each journal) largely varies for different categories [4].

It is clear that the size of a community favours large *h* indexes for individuals since there are more candidates to cite one paper. But, when looking at mean values, this cannot explain the large differences in the factors f_{Cat} of the various categories, not even for the impact factors of individual journals with thousands of articles per year: as the community increases, the number of published articles also increases, so the ratio between the citations to theses papers and the papers themselves cannot be solely determined by the community size (or by the number of papers).

▼ **FIG. 2:** Upper panel: (correction factor-1) for the 35 categories. Lower panel: Hirsch index, when normalized by the correction factors, equivalent to $h = 10$ for the category "Mathematics".



What influences the impact factors of journals and categories?

Here we show that what matters are the different habits of the different communities when writing papers: there is a **strong correlation between the average number of references included in each publication (*Rpp*) of a given category and its aggregate impact factor, f_{Cat}.** Intuitively, it is clear that the more references per article, the more overall citations and therefore the higher impact factor.

Fig. 1 (a) shows the aggregate impact factors, f_{Cat}, for 35 categories with more than 3000 articles each, published in 2008. The values of f_{Cat} vary from 0.695 for the category "Mathematics" to 5.696 for "Cell Biology", while the values of *Rpp* range from 18.5 to 51.6 references per article in these two categories. The f_{Cat} values of the different areas are plotted in different colours and all lie between two straight lines, evidencing a correlation with *Rpp*.

Many of these categories are not independent in the sense that i) the researchers usually publish in several categories although, in fact, mainly do it at most in two or three categories that share some journals [5]; ii) researchers from different sets of these categories are compared since they are compelled to compete for the same funding and positions. (For example, researchers that mainly publish in one or two of the following categories: "Applied Physics", "Materials Science Multidisciplinary", "Chemistry Multidisciplinary", "Physics Condensed Matter", and, recently "Chemistry Organic" are frequently compared.)

Normalization of impact factors

Here, a simple normalization is proposed to eliminate, or reduce, the biasing effect of *Rpp*. Form JCR data, the average number of references per article for all these categories can be obtained: $R_{mean} = 32.8$ (always weighted by the number of published items in each category). The proposed normalization is simple: for a category *i*, the corrected impact factor is defined as: $f_{Corr}(i) = f_{Cat}(i) R_{mean} / Rpp(i)$. The resulting f_{Corr} values are plotted in Fig. 1 (b). The effect is that large values of f_{Cat} related to large *Rpp* are reduced, while low values of f_{Cat} increase if they are associated to low *Rpp*. The large variation range of the impact factors is now reduced by a factor of two: from 5 (from 0.695 to 5.696) to 2.4 (from 1.2 to 3.6).

To check whether there is a correlation between f_{Cat} and the size of the community, we have plotted f_{Cat} in Fig. 1 c) vs. the number of published papers in the category (the colours correspond to the different

areas). There is a small difference between the mean impact factor of the categories with less than 30000 publications per year (2.45) than for those (five) with more than 30000 publications per year (3.0). However, for these highly productive categories, the dispersion is too large, the number of categories too low and the difference in impact factors small to be really significant.

Normalization of Hirsch index:

For an easy visualization of the data, we have represented in the upper panel of Fig. 2 the correction factor (R_{mean} / $Rpp(i)$) – 1 for each category. Consequently, positive and negative values appear in the plot illustrating the effects of our correction on f_{Cat}: positive, the normalized factor increases, and if negative, it decreases. This correction factor may be a reasonable approach to the normalization of different areas so it can be used to compare average *h* indexes of researchers from the different categories. Using the correction factors of the upper panel we obtain the "equivalent *h* indexes" for all categories, see Fig. 2 lower panel. **It is found that *h* = 10 for the category "Mathematics", is equivalent to *h* = 28 in "Cell Biology" or that *h* = 24 for "Astronomy & Astrophysics" is equivalent to *h* = 12 for "Applied Physics".**

Temporal evolution

Also interesting is the evolution with time of the citation habits of scientists. Focussing on the category "Physics, Condensed Matter", Fig. 3 shows that in only seven years (2003 to 2009) its mean impact factor f_{CM} has increased from 1.56 to 2.34, corresponding to the increase of the mean number references per paper (*Rpp*) from 20.8 to 28. This also has implications for the comparison between the *h* indexes of researchers from different times.

The scientific community size is not relevant for mean factors

We showed in Fig.1c that the number of publications per category is not correlated with the mean impact factors (f_{Cat}). Moreover, note that the number of publications in 2008 in the categories grouped in areas of Physics, Biology and Chemistry is similar (Physics: 179558, Biology: 155076, Chemistry: 149546) while the mean impact factors of the whole areas are different (2.51, 3.93 and 2.73, respectively). So, the reason at the root of the observed differences is not in the number of papers (or the community size) but rather lies in the average number of references in each paper. **So the size does matter, but it does not correspond to the size of the community but rather to the habit of some communities to be more generous in citations than others.**

Impact factors and Hirsch index should be used carefully, at least normalized

The final remark would be that **impact and *h* factors have to be used with substantial care since this simple approach evidences that an important part of the large differences between areas are related to publication habits and do not directly reflect different scientific quality. Therefore, these factors, which are usually compared from one area to another, should not be used without some normalization**. The normalization presented here is straightforward and, even if it does not overcome other disadvantages of h (like the number and order of authors in the paper or self-citations, for example) it allows comparing different researchers and areas more fairly. ■

Acknowledgements

I thank G. Benedek for the interesting discussions maintained on this and closely related subjects.

About the Author

Alicia de Andrés studied Physics at the "Universidad Autónoma de Madrid" where she received her PhD in 1987. Engaged at CSIC since 1988 and research professor since 2008, her current research is focused on oxides and hybrid organic/oxide materials for spintronics and optoelectronics.

References

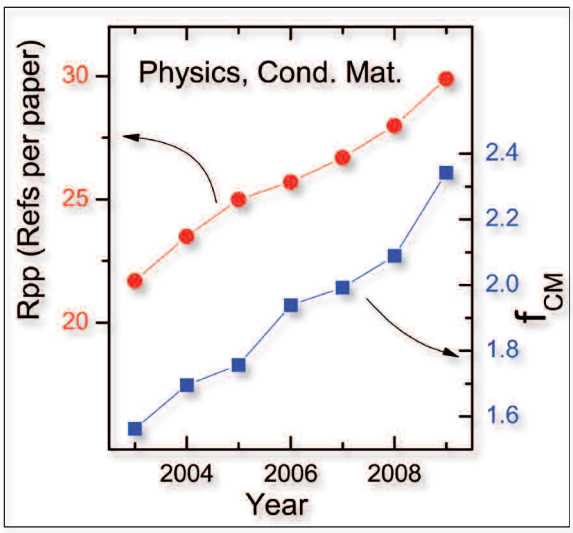
[1] A scientist has Hirsch index *h* if *h* of his/her papers have been cited at least *h* times; see J.E. Hirsch, *PNAS* **102**, 16569 (2005).

[2] M. Wendl, *Nature* **449**, 403 (2007).

[3] A. Sidiropoulos, D. Katsaros and Y. Manolopoulos, *Scientometrics* **72**, 253 (2007).

[4] Data for 2008 for categories from the Journal Citation Reports (JCR), Thomson Reuters-ISI web of knowledge.

[5] The lists of journals associated to each category can be found in JCR .



◀ **FIG. 3:** Evolution with time of the mean impact factor (f_{CM}) and of *Rpp* of the category "Physics, Condensed Matter"