

Epidemiologists (of All People) Should Question Journal Impact Factors

Miguel A. Hernán

Each year Thomson Scientific, a private company, computes the bibliographic impact factor (BIF) for many journals, including general epidemiology journals. The 2006 BIF was 5.2 for the *American Journal of Epidemiology*, 4.5 for the *International Journal of Epidemiology*, and 4.3 for *EPIDEMIOLOGY*.

The literature on the shortcomings of the BIF as a criterion for ranking journals is extensive. The main criticisms of the BIF as a measure of research quality, as well as the vulnerability of the BIF to editorial manipulation and the distortions encouraged by the use of the BIF, have been recently discussed by several authors,^{1–4} including the creator of the BIF.⁵ This commentary does not reiterate those criticisms. Rather, I would like to highlight some flaws of the BIF that epidemiologists are especially well trained to detect. To do so, let me tell you the apocryphal story of a paper that I recently handled as an Editor of *EPIDEMIOLOGY*.

Thomson et al submitted a paper whose implicit goal was to compare the quality of medical care for epileptic patients among the neurology clinics of 3 hospitals located in Baltimore, Maryland, Durham, North Carolina, and Bristol, England. To accomplish this goal, the authors identified all new diagnoses of epilepsy in each hospital during the years 2004 and 2005. They then conducted an exhaustive search to count the number of seizures experienced by each of the hospital's patients during the year 2006. The authors were able to find all occurrences of seizures in these patients no matter where in the world they were in 2006. Pretty impressive, I thought. For each clinic, Thomson et al computed the ratio of number of seizures among its patients divided by the number of patients with epilepsy. They referred to this ratio as the brain irritability factor (BIF). Thomson et al cautioned against the misuse of the BIF, but nonetheless announced their intention to compute the BIF for comparisons among all major hospitals in the world. The 2006 BIF was 5.2 for Baltimore, 4.5 for Bristol, and 4.3 for Durham. I sent the

paper for review to 3 fellow epidemiologists. They raised the following criticisms:

1. **Bad Choice of Denominator:** The authors included information from more subjects in the numerator than in the denominator of the BIF. Specifically, for each hospital, the denominator was the number of patients admitted with a diagnosis of epilepsy in 2004–2005, whereas the numerator was the total number of seizures experienced in 2006 by *all* patients admitted to the clinic in 2004–2005 (regardless of their diagnosis). The reviewers asked to see a corrected BIF that includes all admitted patients (regardless of their diagnosis) in the denominator. Otherwise, the BIF could not be interpreted as “a measure of the frequency of seizures of the ‘average patient’ in a clinic during a particular period,” as proposed by Thomson et al in their article. Had the authors responded to this criticism, they would have reported that the corrected 2006 BIF was approximately 4.1 for Baltimore, 2.8 for Durham, and 2.1 for Bristol.
2. **Need for Adjustment:** The proportion of patients with a diagnosis of epilepsy varied greatly among the 3 hospitals: approximately 86% for Baltimore, 72% for Durham, and 59% for Bristol. Because the number of seizures is expected to be greater among epileptic patients, a crude comparison of the average number of seizures across hospitals would be misleading. Figure 1 represents this problem. Thus the reviewers requested that either the BIF be standardized to some common distribution of epilepsy frequency, or the numerator of the BIF be modified to include only seizures in patients with epilepsy. Had the authors responded, they would have reported that the 2006 BIF restricted to patients with epilepsy was approximately 5.1 for Baltimore, 3.8 for Durham, and 3.1 for Bristol. Some characteristics of the patients with epilepsy (eg, comorbidities) may also be differentially distributed by hospital. If these characteristics are strongly associated with the number of seizures, then even the restricted BIF may be misleading.
3. **Questionable Summary Measure:** Because of the highly skewed distribution of the number of seizures, the mean may not be the most informative summary. Other measures such as the median number of seizures (4 in Baltimore, 2 in Durham and Bristol) or the proportion of

Submitted 7 December 2007; accepted 25 January 2008.

From the Harvard School of Public Health, Boston, MA.

Editors' note: Related articles appear on pages 369, 370, 372, and 373.

Correspondence: Miguel A. Hernán, Harvard School of Public Health, 677 Huntington Ave, Boston, MA 02115. E-mail: miguel.hernan@post.harvard.edu.

Copyright © 2008 by Lippincott Williams & Wilkins

ISSN: 1044-3983/08/1903-0366

DOI: 10.1097/EDE.0b013e31816a9e28

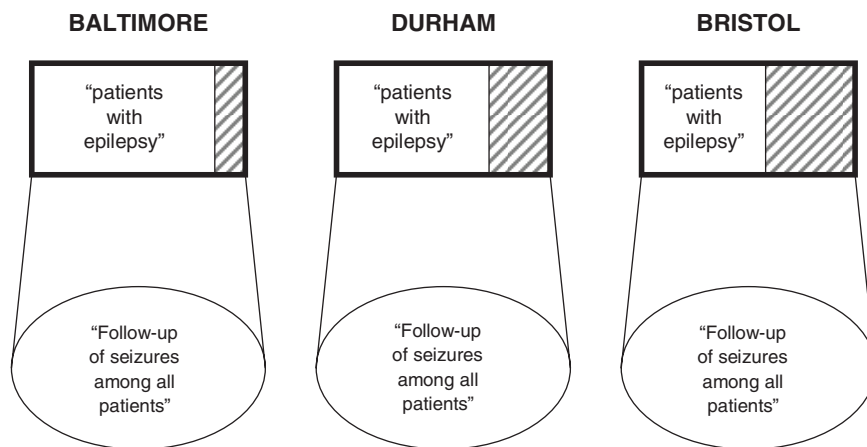


FIGURE 1. The different proportion of patients with epilepsy makes the BIF hard to interpret.

epilepsy patients with no seizures (approximately 4.7% in Baltimore, 11.7% in Durham, 11.3% in Bristol) or the proportion above a certain number of seizures may also provide important information.

I asked the authors to address these standard epidemiologic criticisms. I specifically directed the authors' attention to the fact that restriction to patients with epilepsy changed the value of the BIF differentially among hospitals (a change of about 2% for Baltimore, 10% for Durham, and 30% for Bristol), and wondered whether the sensitivity of the estimates could be explained by a combination of the differential proportions of seizures detected in the same hospital where patients were admitted (about 5.2% in Baltimore, 4.7% in Durham, and 12.0% in Bristol) and of patients with epilepsy.

I also asked the authors for the rationale underlying the use of seizures that occurred only in 2006, and requested a better description of their methods. Specifically, it was un-

clear what procedure was used to diagnose epilepsy, and thus the number of patients that should contribute to the denominator of the BIF. This vagueness ensures that the authors' BIF cannot be replicated by other investigators who may wish to assess its accuracy. In fact, I could not exactly reproduce the unadjusted BIFs reported by the authors, even when provided the raw data consisting of each patient's number of seizures and medical records (as a result, the restricted BIFs for Durham and Bristol reported above are probably 0.1–0.2 lower than they should be).

The authors rejected these criticisms. Paraphrasing Hoeffel,⁶ they responded that the BIF "is not a perfect tool to measure the quality of clinics but there is nothing better and it has the advantage of already being in existence and is, therefore, a good technique for scientific evaluation." They also responded that the diagnosis of epilepsy was "based on human judgment" and the diagnostic criteria were not meant

TABLE 1. Parallels Between Thomson et al's Brain Irritability Factor and Thomson Scientific Journal Impact Factor

	Thomson et al BIF—Medical Care	Thomson Scientific BIF—Impact
Subjects	Patients	Published items
Source population	Hospitals	Journals
Eligibility criterion	Admission in 2004–2005	Publication in 2004–2005
Outcome	No. seizures occurring anywhere in the world in 2006 among those admitted patients	No. citations in any journal in 2006 to any of the published items
Home outcomes	Seizures detected in same hospital in which patient was admitted	Citations in journal of items published in same journal
Subjects contributing to the BIF denominator	Patients with epilepsy	Original research reports, reviews, and other substantive papers
Subjects not contributing to the BIF denominator	Patients without epilepsy	Commentaries, letters, etc. unless they are considered by Thomson Scientific to be substantive papers
Criteria to determine which subjects contribute to the denominator	Unspecified (no explicit definition of diagnosis of epilepsy)	Unspecified (no explicit definition of substantive paper)
Other possible adjustment factors	Comorbidities (cardiovascular, neurologic, HIV . . .)	Research area (cardiovascular, neurology, HIV, methods . . .)

to be publicly available. These responses left me with no choice but to reject the paper. I later learned that other editors in similar situations had actually received similar responses from Thomson, or none at all.⁷

The parallels between this hypothetical BIF and the journal BIF are summarized in Table 1. Many epidemiologists use the Thomson Scientific impact factor to rank journals. Some even decide where to submit their own papers based on the journals' BIF—which confers the BIF rankings with the power of a self-fulfilling prophecy, as journals with higher BIFs (1) get the right of first refusal of many papers, including a disproportionate number of the best ones, and (2) are read more and thus tend to have more citations. Interestingly, some epidemiologists who would be quite critical of Thomson et al's brain irritability factor seem to put their critical faculties on hold when considering the Thomson Scientific bibliographic impact factor, even though both BIFs go against the fundamental epidemiologic principles that epidemiologists abide by in their teaching and research.

Developing a good impact factor is a nontrivial methodologic undertaking that depends on the intended goal of the rankings. Hence, a scientific discussion about any impact

factor requires that its goal is made explicit and its methodology is described in enough detail to make the calculations reproducible. Paradoxically, the methodology of the impact factor that is used to evaluate peer-review journals cannot be fully evaluated in a peer-reviewed journal. As illustrated above, a manuscript describing the Thomson Scientific impact factor would be a hard sell for most journals, and hardly acceptable for the *American Journal of Epidemiology*, the *International Journal of Epidemiology*, or *EPIDEMIOLOGY*.

REFERENCES

1. Seglen PO. Why the impact factor of journals should not be used for evaluating research. *BMJ*. 1997;314:498–502.
2. Porta M, Fernández E, Bolívar F. Commentary: the 'bibliographic impact factor' and the still uncharted sociology of epidemiology. *Int J Epidemiol*. 2006;35:1130–1135.
3. Adler R. The impact of impact factors. *IMS Bull*. 2007;36:4.
4. Chew M, Villanueva EV, Van der Weyden MB. Life and times of the impact factor: retrospective analysis of trends for seven medical journals (1994–2005) and their Editors' views. *J R Soc Med*. 2007;100:142–150.
5. Garfield E. The history and meaning of the journal impact factor. *JAMA*. 2006;295:90–93.
6. Hoeffel C. Journal impact factors [letter]. *Allergy*. 1998;53:1225.
7. PLoS Medicine Editors. The impact factor game [editorial]. *PLoS Med*. 2006;3:e291, 1–2.