Automated grading of research performance clearly fails to measure up.

Nicole Haeffner-Cavaillon, Claude Graillot-Gak, Christian Bréchot

Nature **438**, 559 (1 December 2005) | doi:10.1038/438559a

Automated grading of research performance clearly fails to measure up

N. Haeffner-Cavaillon[1], C. Graillot-Gak[1] and C. Bréchot[1]

1. Cellule de Bibliométrie, Département de l'Évaluation Scientifique, INSERM, 101 rue de Tolbiac, 75654 Paris Cedex 13, France

More contentious than national rankings of research quality, as shown, for example, by David A. King (Nature 430; 311–316; 2004), is the application of such measures to research institutions. Several leading organizations — such as Thomson Scientific (formerly Thomson ISI), the centres for science and technology studies in Leiden (CWTS) and Bern (CEST) and European bibliometric analysts (see A. F. J. van Raan *Scientometrics* **62,** 133–143; 2005) — have emphasized the risk of reaching erroneous conclusions through using inappropriate data.

We have compared an automated and a manual analysis of the performance, between 1994 and 2003 of two European national organizations: the UK Medical Research Council (MRC) and the French biomedical research agency (Inserm) in France. The two agencies are both devoted to biomedical research and are of comparable size.

We first used Thomson Scientific's Web of Science, which correctly identified all 17,829 publications from the MRC and all 46,978 from Inserm. We then compared the Essential Science Indicators (ESI) from the Thomson ranking with a manually extracted list of the 'top 1%' of publications affiliated to France and Britain.

The results turn out to be very different. The manual analysis took affiliations into account carefully, whereas the automated index missed many Inserm-affiliated papers. ESI rankings show 253 'top 1%' publications for the MRC and 117 for Inserm, whereas the manual count has the two organizations on a more equal footing, with 513 top 1% publications for the MRC and 535 for Inserm. As many as 50% of the MRC's and 80% of Inserm's highly cited publications are not identified by the automatic extraction.

Hal-Inserm Author manuscript

Given the use to which these figures are put by funding agencies and governments, these discrepancies, and discrepancies in other types of citation studies, emphasize the problems that can arise from the use of bibliometric analyses.

It is important to ensure that affiliations are captured correctly before performing an analysis, and to use the appropriate citation measure.

For both the MRC and Inserm, only about 20% of papers published in high-impact journals are in the 'highly cited' category, demonstrating that the two indicators should not be confounded.

The research organization of France is extremely complex, which renders assessment difficult. But we believe that France and other countries must collaborate and reach agreement on benchmarks for assessment of research performance, including a simplified, generally accepted affiliation nomenclature.

Hal-Inserm Author manuscript