

Sale été pour le facteur d'impact ?

Analyse I/IST-n°22 - Octobre 2016



En bref :

Publié comme chaque année à la mi-juin pour plus de 11 000 revues, le *Journal Impact Factor* (JIF) reste au cœur des argumentaires marketing des éditeurs scientifiques. On assiste pourtant cette année à un tir de barrage particulièrement nourri contre le JIF et ses usages dévoyés. La prestigieuse revue *Nature* a fait paraître fin juillet un éditorial au vitriol (« *Time to remodel the Journal Impact Factor* »¹) contre le JIF. Publié en août² sur l'archive BioRxiv un préprint³ (« *A simple proposal for the publication of journal citation distributions* ») démontre que si l'on exclut les 15 à 20% d'articles qui génèrent les plus grands nombres de citations, des revues à facteur d'impact très différent mais appartenant à un même champ disciplinaire ont en fait des retombées similaires en matière de citations. Enfin, paru dans *PLoS Biology* le 6 septembre dernier un article (« *Relative Citation Ratio : A New Metric That Uses Citation Rates to Measure Influence at the Article Level* »⁴) émanant d'un groupe de travail piloté par les NIH⁵ propose un nouvel indicateur bibliométrique « ouvert », le RCR, qui prétend apporter une réponse aux insuffisances de la plupart des indicateurs bibliométriques connus. Pourtant le JIF semble avoir encore de beaux jours devant lui : l'adoption large d'une nouvelle métrique est un processus très long et suppose qu'elle soit labélisée par un « tiers de confiance » à forte notoriété. Surtout, le rôle dominant de l'Impact Factor s'explique parce qu'il sert des intérêts consolidés, en particulier ceux des grands éditeurs. Mais le nœud du problème réside dans l'articulation entre processus d'évaluation de la recherche et usage des indicateurs bibliométriques.

L'analyse d'Intelligence IST :

« *Bad summer for the journal impact factor* », ce titre que nous reprenons ici d'un billet⁶ posté le 6 août sur le blog *Physics Today* de l'*American Institute of Physics*, résume bien le foisonnement de critiques qu'a suscité le JIF dans les deux derniers mois. Il est indispensable d'analyser les contenus mais aussi la configuration d'un tel débat (quels intérêts servent quels arguments?).

En raison du prestige et de l'audience de la revue *Nature*, le scud le plus ronflant lancé récemment contre le JIF est sans doute l'éditorial « *Time to remodel the Journal Impact Factor* » paru dans cette revue. Il n'est pas sûr cependant que ce soit celui à plus longue portée. *Nature* ne mâche pas ses mots :

« *Les métriques sont intrinsèquement réductives, et en tant que telles, peuvent être dangereuses. S'appuyer sur elle pour mesurer les performances de la recherche, et non comme un simple indicateur mettant en évidence des résultats ou des défis de la science, ne peut que conduire à des comportements pathologiques. Le Journal Impact factor est l'une de ces métriques dévoyées.*

« *Les éditeurs ont de longue date célébré les facteurs d'impact élevés qui leurs sont attribués. Pourquoi pas ? C'est bien, faute de mieux, l'une des mesures de la pertinence des contenus qu'ils sélectionnent et publient.*

« *Mais le facteur d'impact est aussi une mesure grossière et trompeuse. Il sous-évalue les articles issus de domaines de la recherche à évolution lente où qui structurellement affichent des taux de citation moins élevés. Étant une moyenne arithmétique, il donne une importance disproportionnée à un nombre restreint de publications générant de très nombreuses citations, et il conforte implicitement l'idée que des recherches suscitant un nombre limité de citations sont relativement peu importantes.*

¹<http://www.nature.com/news/time-to-remodel-the-journal-impact-factor-1.20332>

²La dernière version révisée date du 11 septembre.

³<http://biorxiv.org/content/early/2016/09/11/062109>

⁴<http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002541>. Cet article était disponible depuis octobre 2015 sur l'archive ouverte de pré-print BioRxiv (<http://biorxiv.org/content/early/2015/10/22/029629>)

⁵National Institutes of Health, agence fédérale américaine de financement de la recherche biomédicale, qui a par ailleurs un rôle majeur dans la fourniture d'infrastructures IST (Pubmed, etc.) pour les disciplines des sciences de la vie

⁶<http://scitation.aip.org/content/aip/magazine/physicstoday/news/10.1063/PT.5.8183>

Ces insuffisances sont bien connues, mais elles n'ont pas empêché les scientifiques, les agences de financement ou les universités de trop s'appuyer sur les facteurs d'impact, quand les éditeurs de leur côté (Nature comme les autres dans le passé) ont donné une emphase excessive à leur promotion. En conséquence, les chercheurs utilisent le facteur d'impact pour orienter le choix des revues dans lesquelles ils souhaitent publier – à un degré tel que cela sape les fondements d'une science saine.

L'effet le plus pernicieux de cette culture, comme Nature l'a déjà souligné dans le passé, est la pratique qui consiste à utiliser les facteurs d'impact comme base de l'évaluation des résultats individuels des chercheurs ».

Et si le facteur d'impact n'était nullement une mesure de l'impact d'une revue...

Ce réquisitoire sévère ne débouche cependant pas sur l'exigence d'une peine très lourde pour les éditeurs : *Nature* s'engage à ce que ses revues publient désormais dans leur propre descriptif, à côté du facteur d'impact, un éventail de 5 autres indicateurs bibliométriques (pour une liste précise voir⁷), et suggère que les autres éditeurs en fassent autant. Ce remède n'est nullement à la hauteur du mal : multiplier dans les descriptifs de revues les indicateurs bibliométriques, que souvent les usagers de la publication scientifique comprennent mal, c'est introduire un degré de complexité qui n'éclaire en rien la qualité d'une revue. C'est un problème bien connu dans la grande distribution : plus l'étiquetage d'un produit est complexe, moins il sera lu et compris ; avec, selon toute probabilité, un faible impact sur une meilleure prise en compte des imperfections du JIF, qui restera l'indicateur de référence pour caractériser l'importance d'une revue. Il faut bien comprendre que l'une des raisons du succès du facteur d'impact est la grossièreté même de cet indicateur dont la nature et le calcul sont facile à comprendre sans que ses insuffisances soient immédiatement manifestes.

Outre l'éditorial de *Nature*, le document *A simple proposal for the publication of journal citations distributions* » paru en juillet dans un première version sur l'archive ouverte de préprint BioRxiv, a suscité un débat assez large (et est en particulier très critiqué par un post du blog collectif⁸, Scholarly Kitchen). Cet article, dans son introduction, livre un réquisitoire contre le JIF très complémentaire de celui publié dans *Nature*, que nous transcrivons en note⁹.

Comme l'indique son titre, cet article livre une « méthodologie simple » (on ne se prononcera pas ici sur que l'on entend ici par « simplicité ») pour construire, pour une revue donnée, un histogramme de la distribution des citations générées par les articles qu'elle a publiés. En clair, un diagramme présente en ordonnées les nombres de citations obtenues et en abscisse le nombre d'articles ayant obtenu ce nombre de citations (cf. figures ci-dessous). Cette approche, conduite sur une dizaine de revues pour la plupart très connues¹⁰, permet de mettre en évidence deux faits majeurs. D'une part, selon la revue considérée, entre 65 et 75% des articles parus dans ces revues à fort facteur d'impact affichent des nombre de citation inférieurs, voire très inférieurs, au facteur d'impact de la revue. Le JIF étant une moyenne arithmétique, qu'il y ait des occurrences en dessous de cette moyenne n'est pas étonnant. Ce qui est étonnant, en revanche, c'est la mise en évidence, par ces diagrammes, que dans tous les cas seul un très petit nombre d'articles génère la très grande majorité des citations obtenues par une revue. Projeter sur l'ensemble de la revue des performances (en nombre de citations obtenues) qui ne reflètent que celles d'un petit nombre d'articles est problématique. Mais le corollaire de ce constat est encore plus dévastateur : si l'on exclut des diagrammes de distribution des citations les 10 ou 15% d'articles qui génèrent 80% des citations comptabilisées pour une revue, on s'aperçoit que pour cette « longue traîne » d'articles générant moins de citations, les revues ont des performances comparables, même quand elles affichent des facteurs d'impacts très différents. ***En d'autres termes l'Impact factor mesure la capacité de chaque revue à attirer (à susciter?) un nombre très réduit d'articles à fort potentiel de citation et nullement l'« impact » réel d'une revue***, fût-il mesuré par des taux de citation. Ce qui bien sûr est très gênant pour un « facteur d'impact » universellement admis comme tel.

⁷http://www.nature.com/npg/_company_info/journal_metrics.html

⁸« How to manipulate a citation histogram », <https://scholarlykitchen.sspnet.org/2016/08/08/how-to-manipulate-a-citation-histogram/>

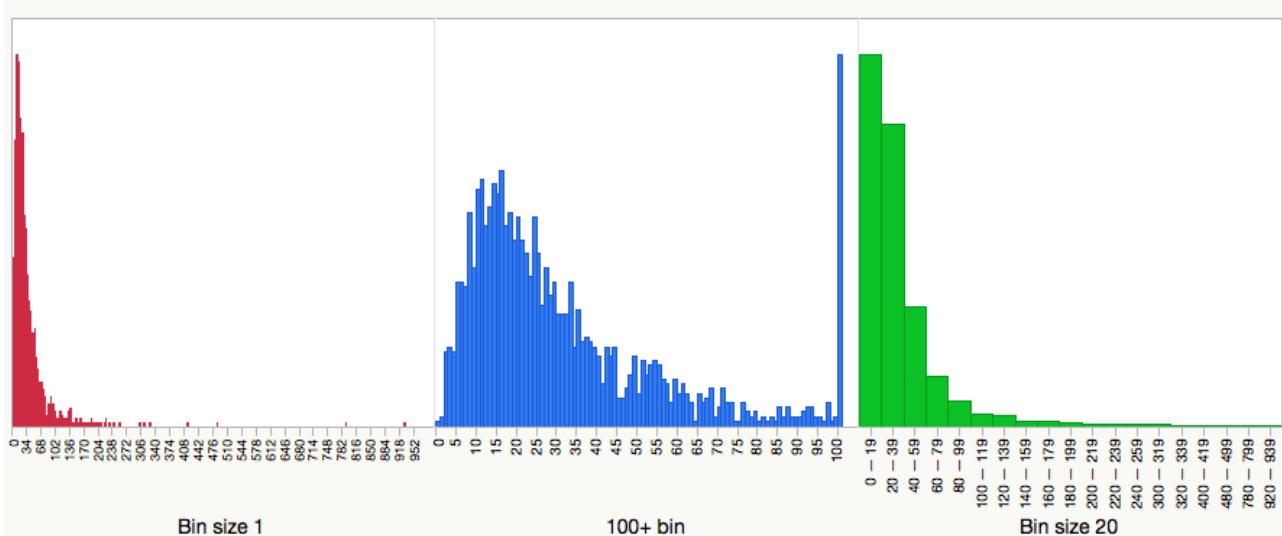
⁹ First of all, the journal IF is a journal-level metric, not an article-level metric, and its use to determine the impact of a single article is statistically flawed since citation distribution is skewed for all journals, with a very small number of articles driving the vast majority of citations. Furthermore, impact does not equal importance or advancement to the field, and the pursuit of a high IF, whether at the article or journal level, may misdirect research efforts away from more important priorities... High-IF journals limit the number of their publications to create an artificial scarcity and generate the perception that exclusivity is a marker of quality. The relentless pursuit of high-IF publications has been detrimental for science... Individual scientists receive disproportionate rewards for articles in high-IF journals, but science as a whole suffers from a distorted value system, delayed communication of results as authors shop for the journal with the highest IF that will publish their work, and perverse incentives for sloppy or dishonest work. [Emphasis added.]

¹⁰Nature, Nature Communications, PLoS One, PLoS Biology, PLoS Genetics, eLife, EMBO Journal, Journal of Infometrics, Proceedings of the Royal Society B

Histogrammes de distribution des citations d'une revue : un outil à faible potentiel

Si ces constats sont frappants et éclairent un aspect des faiblesses du facteur d'impact jusqu'ici peu pris en compte, on ne peut s'empêcher d'être sceptique, malgré la qualité des éditeurs qui ont appuyé ces travaux (PloS, EMBO, Royal Society, Nature...) sur le fait que l'approche par les diagrammes de distribution de citations soit une alternative ou même un complément utile du JIF. D'une part, comme le montrent les figures suivantes (issus du billet critique de Scholarly Kitchen), en faisant conventionnellement varier les échelles sur la base desquelles sont construits ces diagrammes, on modifie sensiblement le profil de ces figures, rendant impossible leur interprétation immédiate et non équivoque. Les auteurs de l'article suggèrent que les éditeurs s'approprient cet outil et publient en complément du JIF ces histogrammes de distribution de citations pour chacune de leurs revues. Quand bien même la publication de ces histogrammes se généraliserait, il est peu probable que cela fasse avancer l'appréciation qualitative des revues au sein d'un même système « universel » d'évaluation. Certes, leurs codes couleurs agrémenteraient les descriptions marketing des revues publiées par les éditeurs, mais leur interprétation est tout sauf intuitive et chaque histogramme nécessite une analyse spécifique, incompatible avec l'immédiateté de lecture qui doit être celle d'un indicateur. De plus, faute d'une méthodologie standardisée, les histogrammes publiés par différents éditeurs seraient difficilement comparables. Ce qui amène Phil Davis, auteur du billet critique paru dans Scholarly Kitchen, à suggérer que si ces histogrammes devaient être généralisés ce soit Thomson Reuters et/ou Scopus d'Elsevier qui servent de « tiers de confiance » (lire ci-dessous) pour l'élaboration centralisée de ces diagrammes... On serait là à l'opposé d'une approche de « science ouverte » puisque les données et les algorithmes que manipulent Thomson Reuters ou Elsevier pour la production de leurs indicateurs bibliométriques sont verrouillés dans des boîtes noires.

Figure 1 : Variation du profil d'un histogramme de distribution des citations de la revue Nature en fonction des échelles adoptées



Légende : Histogramme de distribution des citations des articles parus dans la revue *Nature* en 2013 et 2014, mesurées en 2015 ; source : Thomson Reuters pour les données, Scholarly Kitchen pour les diagrammes

La Relative Citation Ratio : une meilleure prise en compte de l'impact au niveau de l'article

Paru le 6 septembre dernier dans la revue *PLoS Biology* l'article « *Relative Citation Ratio : A New Metric That Uses Citation Rates to Measure Influence at the Article Level* »¹¹) synthétise les travaux de scientifiques réunis dans un groupe de travail par les NIH¹². Cet article propose un nouvel indicateur bibliométrique, le RCR, mesurant l'impact d'un article scientifique donné dans un champ disciplinaire donné. On se détache ici de l'appréciation de l'impact d'une revue dans son ensemble. Ce travail décoche pourtant une autre flèche de taille en direction du JIF : en comparant sur 10 ans (2002-2012) et pour 85 000 articles scientifiques les résultats fournis par cet indicateur et par les indications qualitatives d'un panel d'expert sur les articles les plus influents à long terme dans le domaine des biosciences, les auteurs de l'article

¹¹ <http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002541>. Cet article était disponible depuis octobre 2015 sur l'archive ouverte de pré-print BioRxiv (<http://biomedrxiv.org/content/early/2015/10/22/029629>)

¹²National Institutes of Health, agence fédérale américaine de financement de la recherche biomédicale, qui a par ailleurs un rôle majeur dans la fourniture d'infrastructures IST (Pubmed, etc.) pour les disciplines des sciences de la vie

montrent que seuls 11% des « papers » les plus influents (et qui affichent un RCR supérieur ou égal à 3) sont apparus dans des revues à facteur d'impact élevé (JIF supérieur ou égal à 28) alors que 89% des articles majeurs ont été publiés dans des revues à facteur d'impact moyen ou faible.

Sans entrer dans une explication méthodologique détaillée, le RCR est comme son nom l'indique une mesure relative, un ratio qui, pour un article donné, inscrit au numérateur le nombre exact de citations générées par cet article et au dénominateur la « probabilité » calculée *a priori* qu'a cet article de générer des citations. L'originalité de la méthode est de calculer cette probabilité à partir de la modélisation des réseaux de citations et de co-citations dans un corpus de références bibliographiques (corpus extrait de Pubmed) pertinentes, corpus propre à chaque article. Si un article affiche un RCR de 1, cela signifie qu'il n'a pas généré plus de citations que ce que la modélisation du corpus de référence laissait prévoir : c'est un article à « impact neutre ». Si l'article affiche un RCR de 3 c'est qu'il a généré trois fois plus de citations que ce à quoi on aurait pu s'attendre (c'est donc un article d'un intérêt certain). Selon les auteurs, la validité de cette méthode est démontrée par le fait que les valeurs des RCR calculés automatiquement sont corroborées par les avis d'experts qui, sur la base d'une analyse qualitative, procèdent à une notation de ces mêmes articles. Toujours selon ses auteurs, l'avantage de cette méthode est qu'en modifiant le périmètre, l'extraction et l'analyse statistique du corpus de référence propre à chaque article, (tout cela est réalisé par un algorithme ouvert) le RCR efface l'influence possible des «biais de comportement» spécifiques au pratiques de citation dans un champ disciplinaire donné.

Un outil sponsorisé par les NIH

Outre sa nouveauté et son apparence robustesse, le RCR retient l'attention en raison de la crédibilité de leur sponsor : les NIH ont toujours été un foyer d'innovation important en IST. Un autre aspect du RCR est qu'il semble inaugurer l'ère des « métriques ouvertes » (qui seront sans doute une composante de l'Open Science) : les algorithmes sophistiqués qui permettent à quiconque – individu ou organisation - de calculer le RCR d'un article ou d'un portefeuille d'articles sont en libre accès¹³. Or, il faut souligner que les métriques bibliographiques les plus courantes, que ce soit celles de Thomson Reuteurs (qui ne se limitent pas au JIF), d'Elsevier (avec Scopus) ou même celles de Google Scholar, ne reposent nullement sur des logiques de « science ouverte » : les algorithmes et les données qu'elles mobilisent sont un « secret de fabrication » jalousement gardé par ces opérateurs.

Le RCR parviendra-t-il à s'installer durablement comme indicateur bibliométrique bien adapté pour apprécier l'ampleur des retombées d'un article donné ? Rien n'est moins sûr. *Il serait en effet naïf qu'un indicateur bibliométrique puisse s'imposer dans les usages au point de devenir un « référentiel universel » (ce qu'est à sa façon le JIF) par ses seules qualités intrinsèques.*

Plusieurs facteurs qui n'ont rien à voir avec la pertinence scientifique doivent être réunis pour qu'un indicateur bibliométrique s'impose comme une mesure standard.

De forts intérêts économiques appuient la prédominance du Journal Impact Factor

Le premier de ces facteurs est le temps : publié pour la première fois en 1972 le tableau des facteurs d'impact des revues a longtemps vivoté, utilisé essentiellement par les bibliothécaires pour orienter leurs politiques d'abonnement aux revues. Ce n'est qu'au début des années 2000 (avec la mention des facteurs d'impact dans les bibliographies personnelles publiées par les chercheurs et en parallèle la prise en compte des facteurs d'impact dans l'évaluation des chercheurs) que s'est amorcée la « spirale infernale » du recours dévoyé au facteur d'impact.

Le second facteur décisif est la force des intérêts (et en particulier des intérêts économiques) qui œuvre à promouvoir une métrique plutôt qu'une autre. Si le JIF a acquis une telle proéminence c'est que les éditeurs - et tout particulièrement les plus grands - y ont un intérêt vital : les quatre premiers éditeurs, s'ils éditent au plan mondial 24,8% des titres, publient 50,1% des revues à plus fort facteur d'impact¹⁴. Le facteur d'impact est un paramètre important de la valeur patrimoniale d'une revue. Ne serait-ce que parce que le niveau des tarifs d'abonnement et l'ampleur de l'audience – et par voie de conséquence la rentabilité – d'une revue sont largement indexés sur son facteur d'impact. La valeur d'actif d'un portefeuille de revues est largement corrélée à sa « qualité » mesurée indument par le facteur d'impact. De plus, le facteur d'impact est un devenu un paramètre essentiel de l'attractivité d'une revue pour les chercheurs publant. Or, la concurrence pour satelliser les meilleurs auteurs ou les auteurs les plus prometteurs est la seule concurrence qui aujourd'hui oppose réellement les éditeurs. Le facteur d'impact est donc un levier stratégique pour les éditeurs auquel

¹³<https://icite.od.nih.gov/>

¹⁴ Ce dénombrement est celui de Sally Morris, "Mapping the Journal publishing landscape", Learned Publishing, vol. 20, N° 4, octobre 2007. Cette étude de 2007 n'a jamais été actualisée. Mais le phénomène de concentration des revues à facteur d'impact aux mains des grands éditeurs s'est probablement accru.

ils ne sont pas prêts de renoncer¹⁵. La « course au facteur d'impact » ne s'arrêtera que quand les méthodes d'évaluation tourneront le dos au JIF. Mais malgré quelques signaux encourageants on est loin d'une telle perspective.

La notion essentielle de « tiers de confiance »

Il y aurait certes au sein de la communauté internationale de la recherche publique un intérêt (y compris économique) que se développe, sur la base de méthodologies admises de tous, des métriques alternatives ouvertes, tant au niveau de la revue qu'au niveau de l'article. Mais cet intérêt diffus n'est relayé par aucun acteur puissant. Si le NIH a par exemple financé les travaux et les développements algorithmiques qui sous-tendent le RCR, il est peu probable que les NIH, qui pourtant disposent d'une légitimité et d'un label fort prennent la responsabilité et le risque de devenir, au travers l'un de ses services, une « agence de notation » de la publication scientifique en biosciences. Or, un RCR ouvert et diffus (ou chaque chercheur ou chaque entité de recherche étonnerait ses publications) n'aurait ni la centralité, ni la visibilité ni l'indépendance requise pour que cet indicateur soit universellement admis.

Ce dernier point amène à identifier un troisième facteur indispensable pour qu'un indicateur bibliométrique s'impose largement dans les usages : il faut qu'il soit proposé par un « tiers de confiance » disposant d'une forte notoriété ayant la force d'un label. Lorsque Phil Davis, dans son billet pour *Scholarly Kitchen* suggère que ce soient Thomson Reuters ou Elsevier qui produisent des histogrammes de distribution des citations d'une revue, il ne dit pas autre chose. Si malgré tous ses défauts criants, le Journal Impact Factor continue année après année de faire l'actualité IST du mois de juin c'est parce que Thomson Reuters joue ce rôle de tiers de confiance, disposant d'une marque forte perçue à tort ou à raison comme un label de qualité, malgré les critiques récurrentes de l'outil. Le JIF a une histoire longue, on peut donc supposer que sa méthodologie est stabilisée et que le Journal Citation Index (à partir duquel est calculé le JIF) offre des relevés sincères et vérifiables.

Une opportunité de nouveaux services bibliométriques pour les éditeurs

Or, là aussi, on ne voit pas quelle(s) entité(s) internationale(s) (l'international est dans le champ de la publication scientifique la seule échelle pertinente) émanant des communautés de recherche elles-mêmes seraient à même de jouer ce rôle d'« agence de notation » produisant des indicateurs bibliométriques de mesure d'impact fiables, uniques, faciles à déchiffrer, et sur la base d'une logique de « science ouverte » donnant accès à toutes les données et tous les algorithmes utilisés pour produire ces indicateurs. Et si l'innovation semble vivace sur le front des indicateurs bibliométriques, ce sont finalement les grands éditeurs qui sont peut-être les mieux à même de se l'approprier. Thomson Reuters (qui dans son service *InCite* a déjà un indicateur qui n'est pas sans rappeler le RCR) ou Elsevier ont les moyens de s'inspirer de ces innovations, de construire des indicateurs bibliométriques à partir de leurs riches bases de citations, et demain de revendre fort cher (c'est la logique à l'œuvre dans *InCite*) ces palettes d'outils bibliométriques à leurs clients.

La DIST/CNRS organise le 17 octobre la Journée « Innométriques »

sur le campus Gérard Mégie du CNRS (3, rue Michel-Ange, 75016 Paris)

Le colloque Innométriques vise à répondre aux attentes globales de l'ESR sur les usages des bases et outils de métrique. Il s'agit en particulier de :

- ◆ Clarifier les attentes et les usages des outils de métriques à l'heure du text and data mining (TDM).
- ◆ Mieux maîtriser les usages des bases et outils de métrique dans le cadre de l'activité de recherche et de la gestion de la recherche.
- ◆ œuvrer à l'amélioration de la visibilité de la production scientifique française.
- ◆ Construire une vision prospective partagée sur les nouveaux usages de la métrique.

<http://www.cnrs.fr/dist/z-outils/documents/Innometriques/pre-programme-journee-innometriques.pdf>

¹⁵Comment interpréter alors la vive prise de position de Nature (fleuron de Springer Nature, N°2 mondial de l'édition scientifique) contre l'impact factor et la participation de Nature aux travaux sur les histogrammes de distribution des citations ? Il faut faire ici la part du positionnement éditorial propre de Nature. Cette revue, bien qu'ayant des pratiques tarifaires et commerciales aussi contestables que celles des autres grands éditeurs scientifique, s'est toujours voulu en phase avec les débats traversant la communauté scientifique. Le site nature.com a ainsi été dès l'origine un important relais du débat sur l'Open Access. Nature, en programmant sa charge contre le JIF est cohérent avec cette politique d'image. Il serait cependant tout à fait illusoire de penser que cette position va changer quoi que ce soit au fait que Springer Nature continuera à promouvoir activement le facteur d'impact.