

# Robustesse du score SIGAPS, critère bibliométrique pour valoriser les publications des établissements de santé

Stefan J. Darmoni<sup>1</sup>, Joël Ladner<sup>2</sup>, Patrick Devos<sup>3</sup>, Jean-François Gehanno<sup>1</sup>

1. CISMef, CHU de Rouen & GCSIS, TIBS, LITIS EA 4108, Université de Rouen, France
2. Département de Santé Publique, CHU de Rouen
3. Délégation à la Recherche Clinique, CHRU de Lille & CERIM, Université de Lille II

Reçu le 11 février 2009  
Accepté le 27 mars 2009

Disponible sur internet le :  
6 mai 2009

## Correspondance :

Jean-François Gehanno, Université de Rouen, 1 rue de Germont F-76031 Rouen, France.  
jean-francois.gehanno@univ-rouen.fr

## ■ Summary

### Reliability of a bibliometric tool used in France for hospital founding

*SIGAPS is a bibliometric score that aims at making an inventory, evaluating and promoting scientific publications of hospitals that perform research. It has become a major stake in France since it is one of the most important components of the MERRI (Mission Training, Research, Reference and Innovation) founding of hospitals. This score is based on the points attributed to the authors of articles published in journals indexed in Medline, according to the rank of the authors and the Impact Factor of the journal.*

**Objectives** > to compare the reliability of the score when applying different way of computing it, and different weights for the rank or the Impact Factor.

**Material and methods** > we computed the scores of all the physicians of a University Hospital, using the rules that are actually applied at the national level. We then used 4 different scenarios, with different weight given to the rank of authors or the Impact Factor. We compared the scores obtained by each author according to the different scenarios with the Spearman's rank and Pearson's correlation coefficients.

**Results** > The score is not significantly affected when no points are given to the fourth authors and above, when the last author get

## ■ Résumé

**Contexte** > Le projet SIGAPS (Système d'interrogation, de gestion, et d'analyse des publications scientifiques) est un projet dédié au recensement, à l'évaluation et la valorisation des publications scientifiques d'un établissement ayant des activités de recherche médicale. C'est un enjeu pour ces établissements car il constitue un des indicateurs essentiels intervenant dans la part modulable de l'enveloppe MERRI (Mission Enseignement, Recherche, Référence et Innovation). Le score SIGAPS est fondé sur un calcul bibliométrique, accordant des points aux auteurs d'articles dans des journaux indexés dans Medline en fonction du facteur d'impact de la revue et du rang des auteurs.

**Objectifs** > Comparer la robustesse du score SIGAPS en fonction de différents scénarios d'attribution des points selon le rang des auteurs ou le type de revue.

**Méthodes** > Nous avons calculé les scores individuels des médecins d'un CHU, en utilisant les principes appliqués au niveau national dans le cadre de ce projet. Nous avons ensuite recalculé ces scores en appliquant 4 différents modèles d'attribution de points en fonction du rang des auteurs ou du facteur d'impact. Nous avons comparé ces différents résultats en utilisant les tests de corrélation des rangs de Spearman et de régression de Pearson.

**Résultats** > Le fait de ne pas attribuer de points au delà du 3<sup>e</sup> auteur, d'augmenter les points du dernier auteur ou de diminuer le poids de

more points or to change the points according to the Impact Factor of the journal.

**Conclusion** > The different scenarios do not lead to significant changes for the physicians' scores, and therefore for the cumulated score of the hospital. Despite the well known limits of bibliometric indicators, the SIGAPS score appears reliable to compare the hospitals for founding decisions.

**L**e projet SIGAPS (Système d'interrogation, de gestion, et d'analyse des publications scientifiques), initié au CHRU de Lille en 2002, est dédié au recensement, à l'évaluation et la valorisation des publications scientifiques d'un établissement ayant des activités de recherche médicale [1]. En France, la recherche s'effectue essentiellement dans les Centre hospitalo-universitaires (CHU), les Centres de lutte contre le cancer (CLCC) et quelques Centres hospitaliers généraux (CHG). Initié dans le cadre de la mise en place des réformes de financement des hôpitaux, le projet SIGAPS s'articule aujourd'hui autour d'un logiciel de bibliométrie qui automatise le processus de recensement, de chargement et d'analyse des références issues exclusivement de la base de données *Medline* créée par la *National Library of Medicine*, aux États-Unis. En 2007, SIGAPS est un enjeu pour les établissements ayant des activités de recherche, car il constitue un des indicateurs essentiels intervenant dans la part modulable de l'enveloppe MERRI

certaines journaux, de façon isolée (scénario 1 à 3) ou cumulée (scénario 4) n'a eu aucun effet significatif sur le score SIGAPS d'un établissement.

**Conclusion** > Le changement du mode d'attribution des points en fonction du rang des auteurs ne modifie pas significativement le score cumulé des auteurs et donc d'un établissement. En dépit des limites des approches bibliométriques, le score SIGAPS apparaît robuste et utilisable pour la comparaison des établissements entre eux.

(Mission enseignement, recherche, référence et innovation), qui est une incitation financière au sein des missions d'intérêt général [2].

Les principes généraux de SIGAPS sont :

- l'automatisation d'un processus de recensement et d'analyse des publications (fait manuellement auparavant) ;
- une transparence complète des informations relatives aux publications du CHU contenues dans la base de données *Medline* ;
- un outil utile à beaucoup d'acteurs des hôpitaux (responsables, médecins, chercheurs, ...) ;
- un outil de valorisation de la recherche du site CHU (à travers l'enveloppe MERRI).

Dans le domaine de la recherche, le critère bibliométrique le plus reconnu est l'influence d'un article sur les autres articles, et qui mesure la notoriété des journaux [3], traduit par le facteur d'impact (voir encadré 1).

La principale innovation de SIGAPS, est de transformer le facteur d'impact de l'*Institute for Scientific Information* (ISI) qui est un chiffre absolu, en un critère bibliométrique relatif pour tenter d'égaliser toutes les disciplines biomédicales.

### Ce qui était connu

- Le score SIGAPS est un enjeu majeur de valorisation des publications, pour le financement des CHU. Il est utilisé pour comparer des CHU entre eux
- Il ne se fonde que sur les publications indexées dans *Medline*, et sur le facteur d'impact relatif des journaux par discipline
- Le mode de calcul fait intervenir une pondération en fonction du rang des auteurs

### Ce qu'apporte l'article

- La modification des points attribués aux auteurs ou en fonction du facteur d'impact de la revue change les scores SIGAPS des publiants.
- Cette modification ne change pas significativement leur classement, et donc la comparaison entre eux ou, globalement, entre CHU
- D'autres pistes de réflexion devraient être examinées, telles que la pondération en fonction du type de publication (article original, éditorial, cas clinique, revue de la littérature).

#### ENCADRÉ 1

##### Le facteur d'impact

Le facteur d'impact (*impact factor* en anglais) est publié dans *Science Citation Index (SCI) Journal Citation Reports (JCR)* ; il est fondé sur l'analyse bibliométrique des journaux scientifiques indexés dans la base de données de l'*Institute for Scientific Information* (ISI) (devenu *Thomson Scientific*) [4]. Cette base de données est distincte de *Medline*, et bien qu'elles soient assez superposables en termes de journaux indexés, un certain nombre de journaux ne sont indexés que dans une de ces deux bases. Le facteur d'impact d'une revue est le rapport à un temps  $t$  du nombre de citations de cette revue au cours des années  $a-1$  et  $a-2$  par rapport au nombre d'articles publiés durant ces mêmes années dans la revue. Par exemple pour calculer le facteur d'impact 2007 (IF 2007) dans une revue, on applique la formule suivante :  

$$IF\ 2007 = \frac{\text{Nombre de citations en 2007 des articles parus en 2005 et 2006}}{\text{Nombre d'articles parus en 2005 et 2006}}$$

Certaines disciplines de recherche fondamentale (comme la biologie cellulaire, l'immunologie, ou la génétique) ont des facteurs d'impact moyen (ISI) par discipline beaucoup plus élevés que la plupart des disciplines cliniques. Cette différence compliquait la promotion des hospitalo-universitaires en n'obtenant pas une comparaison aisée entre deux dossiers (notamment un de recherche fondamentale et un de recherche clinique). Il est à noter que le facteur d'impact a été utilisé pour la promotion des hommes alors qu'il a été conçu pour mesurer la notoriété des journaux [3].

L'objectif de ce travail a été de mesurer la robustesse du score SIGAPS, score qui tient compte de la revue et de la position de l'auteur, en testant la méthode nationale retenue pour le calculer, et en la comparant à 4 modifications de son calcul. Il s'agit d'un enjeu important puisque le budget de la recherche dans les CHU et dans les CLCC (par l'enveloppe MERRI) dépend grandement de ce score depuis 2007.

## Méthodes

SIGAPS repose sur la relativisation par discipline biomédicale du facteur d'impact. Ainsi, SIGAPS utilise la classification des journaux d'une discipline par valeur ordonnée des facteurs d'impact de chaque journal. Dans chaque discipline reconnue par ISI, les facteurs d'impact des journaux sont classés par ordre décroissant. Les journaux sont dits de type A si le facteur d'impact est dans les 10 % meilleurs, B entre 10 et 25 %, C entre 25 et 50 %, D entre 50 et 75 %, puis E pour 75-100 %, et enfin NC pour les journaux qui n'ont pas de facteur d'impact. À titre d'exemple, la discipline gastro-entérologie comporte 48 journaux. Le journal *Gastroenterology* est classé premier et lui est donc affecté un rang A alors que la revue *Liver Transplant*, classée huitième pour cette même discipline est de niveau B. Les journaux indexés dans *Medline* mais non recensés par *Thomson Scientific* (certains actes de congrès internationaux, par exemple le congrès mondial d'informatique médicale MEDINFO) sont affectés à la catégorie NC. Cette catégorisation permet d'obtenir un facteur d'impact relatif par discipline. Cette catégorisation est revue chaque année en fonction du facteur d'impact : ainsi une revue X pourrait être classée A en année Y, et moins bien classée l'année suivante, ou inversement.

Le score SIGAPS est obtenu par l'addition de 2 facteurs indépendants : la catégorisation du facteur d'impact et le rang de publications des auteurs. Pour la catégorisation du facteur d'impact, le score SIGAPS fournit 4 points pour la catégorie A, 2 points pour la catégorie B et 1 pour les catégories C, D, E et NC. Pour le rang de publications des auteurs, le score SIGAPS fournit 4 points pour le premier auteur, 3 points pour le second, 2 points pour le troisième et le dernier auteur, et un point pour les autres. Ce calcul a été choisi sur le plan national : nous l'appellerons le scénario 0.

Nous avons imaginé de tester la robustesse de ce score en le comparant à 4 autres scénarios (Sc).

- Scénario 1 : diminuer le score SIGAPS des auteurs, qui ne sont ni dans les 3 premiers auteurs, ni le dernier auteur. Au lieu d'accorder 1 point, ce scénario accorde 0 point (mais le score SIGAPS tient toujours compte des points issus du facteur d'impact relatif de la revue dont est issu l'article).
- Scénario 2 : diminuer le score SIGAPS pour les journaux des catégories D, E et NC qui sont toutes à 1 point dans le scénario actuel (ou scénario 0). Dans le scénario 2, les journaux classés D vaudraient 0,5 point et les journaux classés E ou NC 0,25 point.
- Scénario 3 : donner plus de poids au dernier auteur (le chef de service ou l'auteur correspondant de l'article), c'est-à-dire attribuer 4 points pour le dernier auteur, comme pour le premier auteur.
- Scénario 4 : ce scénario cumule les modifications apportées aux 3 scénarios 1, 2, et 3.

La comparaison entre les scénarios s'est effectuée dans un premier temps 2 à 2. Les scores des scénarios ont été comparés par le test de corrélation des rangs de Spearman ( $r_s$ ) et par le test de la pente de la droite de régression de Pearson ( $R^2$ ). Les différences ont été considérées comme significatives au seuil de 5 %.

Pour tester ces 4 scénarios, nous avons utilisé le score SIGAPS du CHU de Rouen pendant la période 2003-2006, représentant un plan quadriennal. Le score de tous les médecins publiants du CHU de Rouen ont été inclus dans l'étude, soit 290 personnes.

## Résultats

Le calcul actuel du score SIGAPS a été comparé aux 4 scénarios décrits dans la section « Méthodes ». Dans les 4 cas, la corrélation entre le scénario actuel et les 4 autres scénarios était très forte et très significativement positive, que ce soit par la méthode des rangs de Spearman, ou par la méthode de la droite de régression.

La comparaison du score SIGAPS actuel par rapport aux 4 autres scénarios de calcul est dans le [tableau I](#) et la [figure 1](#) [test de la pente de la droite de régression de Pearson ( $R^2$ )] et dans le [tableau II](#) [test de corrélation des rangs de Spearman ( $r_s$ )].

TABLEAU I

Comparaison du score SIGAPS actuel (Sc.0) par rapport à 4 autres scénarios de calcul (test de la pente de la droite de régression de Pearson ( $R^2$ ))

	Sc 0	Sc 1	Sc 2	Sc 3	Sc 4
Sc 0	1	0.997*	0.998*	0.991*	0.983*
Sc 1	0.997*	1	0.994*	0.987*	0.985*
Sc 2	0.998*	0.994*	1	0.985*	0.981*
Sc 3	0.991*	0.987*	0.985*	1	0.994*
Sc 4	0.983*	0.985*	0.981*	0.994*	1

\* La corrélation est significative avec un  $p < 0.001$  (bilatéral).

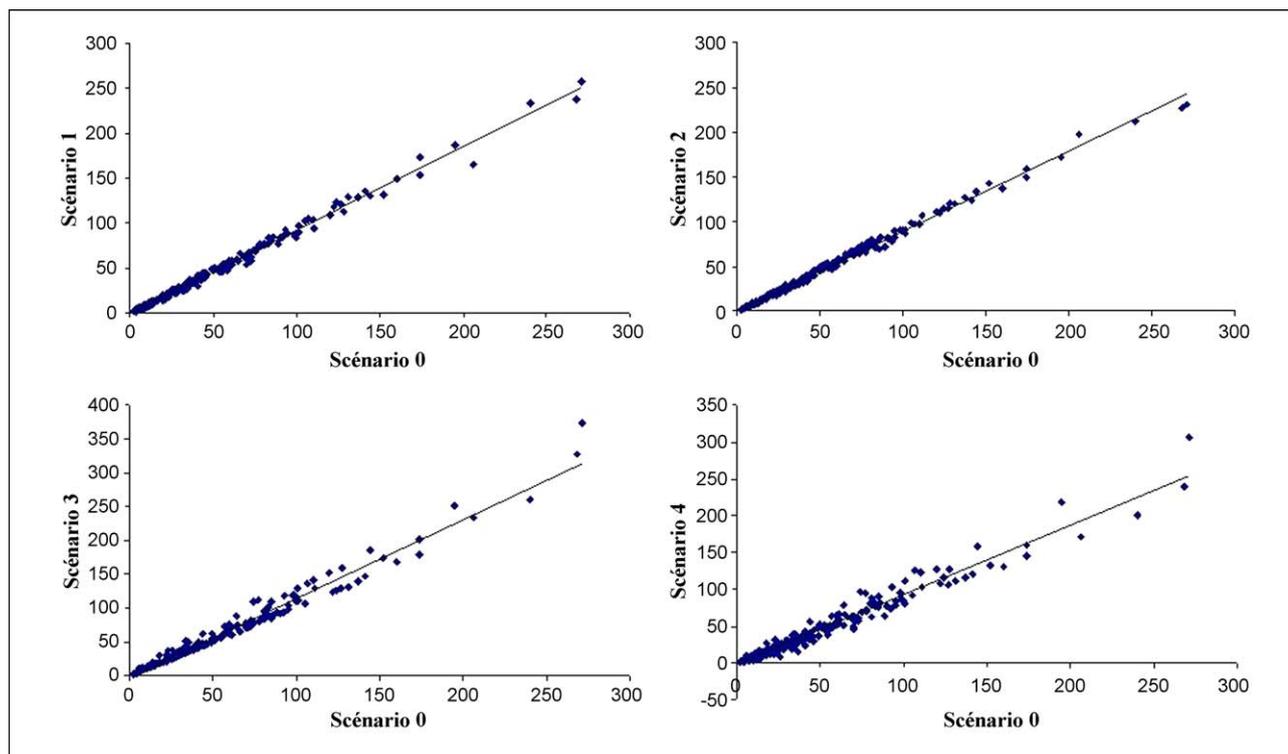


FIGURE 1  
 Comparaison du score SIGAPS actuel (Sc.0) par rapport à 4 autres scénarios de calcul (test de Pearson)

TABLEAU II  
 Comparaison du score SIGAPS actuel (Sc.0) par rapport à 4 autres scénarios de calcul (test de corrélation des rangs de Spearman ( $r_s$ ))

	Sc 0	Sc 1	Sc 2	Sc 3	Sc 4
Sc 0	1	0.996*	0.998*	0.996*	0.984*
Sc 1	0.996*	1	0.995*	0.993*	0.990*
Sc 2	0.998*	0.995*	1	0.992*	0.985*
Sc 3	0.996*	0.993*	0.992*	1	0.988*
Sc 4	0.984*	0.990*	0.985*	0.988*	1

\* La corrélation est significative avec un  $p < 0.001$  (bilatéral).

## Discussion

Ce travail sur la production d'un CHU français a montré la très grande robustesse du score SIGAPS. Il nous semble que cette étude renforce, s'il en était besoin, le choix stratégique effectué par la Direction des hôpitaux et de l'organisation des soins (DHOS) de l'utiliser pour calculer en partie la part modulable de l'enveloppe MERRI. Ce score SIGAPS devient également un

critère bibliométrique « global » de plus en plus utilisé dans les CHU (et bien entendu d'abord à Lille [5]) pour promouvoir les hospitalo-universitaires. Il a également permis de réaffirmer la place importante que les hospitalo-universitaires occupaient au sein de la recherche biomédicale française : une étude d'agrégation nationale des SIGAPS des différents établissements, (citée par son auteur dans [6]) a identifié 80 000 publications réalisées sur 5 ans, avec un facteur d'impact moyen de 2,87.

Le choix stratégique de restreindre SIGAPS aux journaux indexés dans *Medline* nous semble également licite en terme de faisabilité ; en effet, outre le fait essentiel que *Medline* est de très loin la base de données internationale la plus utilisée dans le monde, les bases de données françaises, comme *Francis* ou *Pascal*, ne sont toujours pas d'accès libre et gratuit pour les enseignants-chercheurs de notre pays, alors que *Medline* l'est pour tout Internaute depuis plus de 10 ans ! Néanmoins, les publications indexées dans *Medline* ne couvrent globalement qu'une partie de la littérature (francophone ou non), variable selon les disciplines et les sujets. Ainsi, ce pourcentage est de 32 % en médecine générale [7] et de 48 % en toxicologie environnementale [8].

Comme le facteur d'impact, dont il est en partie déduit, le score SIGAPS peut néanmoins être critiqué. Il a notamment été

démontré la non validité de l'extrapolation du facteur d'impact d'une revue aux articles qu'elle contient, pris individuellement [9]. Ce point indique que le facteur d'impact d'une revue est composé de facteur d'impact d'articles individuels, avec une très forte variance entre ceux-ci. Une autre critique peut être adressée spécifiquement au score SIGAPS, liée à la classification des spécialités médicales de *Thomson Scientific* sur laquelle il repose. Ainsi, si la plupart des disciplines médicales ou chirurgicales (comme la dermatologie, la pédiatrie, l'orthopédie ou la gynécologie-obstétrique) constituent chacune des catégories autonomes et à part entière, certaines disciplines plus transversales sont incluses dans une catégorie plus générale. C'est par exemple le cas de la neurochirurgie incluse dans la discipline plus générale Neurosciences ou les disciplines "hygiène hospitalière" et "médecine du travail" incluses dans la "santé publique". Alors que les meilleurs journaux de ces 2 disciplines "hygiène hospitalière" et "médecine du travail" ont un facteur d'impact respectivement de 2,989 (*Infection Control and Hospital Epidemiology*) et de 2,817 (*Occupational and Environmental Medicine*), la meilleure revue (en terme de facteur d'impact) de cette rubrique "santé publique" (*Annual Review of Public Health*) a un facteur d'impact de 8,978, ce qui sous-estime dans SIGAPS le poids relatif des publications dans ces 2 autres spécialités, puisque leur meilleure revue n'est classée qu'en catégorie B.

Au niveau national, il a été choisi d'utiliser l'addition entre les 2 critères choisis, à savoir la catégorisation du facteur d'impact et le rang de publications des auteurs. Le choix de la multiplication aurait été beaucoup plus discriminant, mais en retrouvant sans doute les biais décrits précédemment sur le facteur d'impact. Actuellement, les différences maximales pour un article varient de 8 à 2 : 8 points pour un premier auteur dans une revue de rang A (comme le *New England Journal of Medicine*), 2 points pour un auteur dont le rang que certains instituts de recherche définissent comme « non significants » (à savoir ni dans les 3 premières places ni en dernier auteur) et dans une revue de rang C, D, E ou NC (c'est-à-dire dans la deuxième moitié des journaux d'une discipline en terme de facteur d'impact). Si le score SIGAPS s'appuyait sur la multiplication, les différences maximales passeraient de 16 à 1 (pour les mêmes cas de figure que précédemment). Il serait intéressant de tester à l'avenir la robustesse du score SIGAPS avec la multiplication en la comparant avec l'addition. Avec le score SIGAPS actuel, il (re) devient efficace de publier en français, sous réserve que la revue soit indexée dans *Medline*, pour l'enveloppe MERRI de votre établissement, au détriment (peut-être) d'une carrière hospitalo-universitaire.

Les trois scénarios contre lesquels nous avons mesuré le score SIGAPS actuel sont des pistes pour amender celui-ci pour les années à venir :

- comme suggéré dans le scénario 1, diminuer le score SIGAPS des auteurs en position « non significative » ; mais dans ce cas, il

faut conserver l'addition, car la multiplication rend nul ce score pour tous les auteurs dans ces positions. Or, celles-ci sont souvent attribuées aux auteurs les plus transversaux dans une communauté hospitalo-universitaire, par exemple les biostatisticiens ou les anatomopathologistes ;

- comme proposé dans le scénario 2, diminuer le score SIGAPS pour les journaux D à 0,5 point (au lieu de 1) et E et NC à 0,25 (au lieu de 1) ;
- enfin, comme dans le scénario 3, donner plus de poids au dernier auteur, c'est-à-dire attribuer 4 points à celui-ci, comme pour le premier auteur. Le dernier auteur correspond le plus souvent à la personne ayant encadré le travail, et non plus au chef de service par défaut, qui n'encadre que l'équipe. Certains instituts de recherche appliquent déjà ce mode d'égalité dans leur démarche d'évaluation. On peut noter d'ailleurs que dans SIGAPS, à compter de 2009, le poids du dernier auteur sera renforcé, passant de 2 à 3 points. Cette tendance à valoriser spécifiquement le dernier auteur, en posant comme principe qu'il a encadré l'étude publiée, ne correspond pas à la proposition du Comité international des rédacteurs des journaux médicaux. En effet, l'*International Committee of Medical Journals Editors* ([www.icmje.org](http://www.icmje.org)) n'indique aucune règle pour décider de l'ordre des auteurs, y compris pour le dernier auteur. Il ne précise que le fait que les auteurs doivent avoir participé significativement à chacune des 3 étapes suivantes :
  - conception de l'étude, acquisition des données ou analyse et interprétation des résultats ;
  - participation à l'écriture de l'article ;
  - accord explicite de la version finale du manuscrit.

D'autres pistes de modification de ce score existent : la première serait de tenir compte des types de publication de la base de données bibliographique *Medline*. Ainsi, le score SIGAPS devrait accorder plus de points aux articles originaux qu'aux journaux de la littérature ou surtout aux cas cliniques et lettres à l'éditeur. Cela paraît techniquement faisable puisqu'il existe dans *Medline* un champ spécifique précisant le type de publication.

Le score SIGAPS pourrait également tenir compte de nouveaux critères bibliométriques, comme le H-index [10], déjà utilisé par des institutions de recherche en France. Ce critère permet d'analyser l'impact réel des articles, et donc des auteurs, en décomptant le nombre réel de citations des auteurs, mais il n'est pas lui-même pas exempt de biais puisqu'il repose, en accès libre, sur *Google Scholar*, dont le mode de sélection n'est ni clair ni stable dans le temps<sup>1</sup>.

**Conflits d'intérêts :** aucun.

**Remerciements :** Les auteurs remercient Mme Jessica Hautin de la Direction de la Recherche et de l'Innovation du CHU de Rouen qui nous a fourni tous les fichiers nécessaires à cette étude.

<sup>1</sup> A noter que le H-index peut aussi être calculé dans le *Web of Knowledge* sur les données de *Thomson Scientific*, dont le mode de sélection est plus transparent.

## Références

- [1] Devos P, Dufresne E, Renard JM, Beuscart R. SIGAPS: a prototype of bibliographic tool for medical research evaluation. *Stud Health Technol Inform* 2003;95:721-6.
- [2] Douste-Blazy P. Discours pendant les IXèmes Assises nationales hospitalo-universitaires. URL : [http://www.sante.gouv.fr/htm/actu/33\\_050317pdb.htm](http://www.sante.gouv.fr/htm/actu/33_050317pdb.htm).
- [3] Ojasoo Tiu, Maisonneuve H, Matillon Y. Le facteur d'impact des revues, un indicateur bibliométrique à manier avec prudence. *La Presse Médicale* 2002;31:775-81.
- [4] Garfield E. Fortnightly review: how can impact factors be improved? *BMJ* 1996;17(313):411-3 (Available from Internet: <http://www.bmj.com/cgi/content/full/313/7054/411>).
- [5] Devos P, Lefranc H, Dufresne E, Beuscart R. From bibliometric analysis to research policy: the use of SIGAPS in Lille University Hospital. *Stud Health Technol Inform* 2006;124:543-8.
- [6] Devos P. De la bibliométrie au financement : le logiciel SIGAPS. *J Neuroradiol* 2008;35: 31-3.
- [7] Wilkins T, Gillies RA, Davies K. EMBASE versus MEDLINE for family medicine searches: can MEDLINE searches find the forest or a tree? *Can Fam Physician* 2005;51: 848-9.
- [8] Gehanno JF, Paris C, Thirion B, Caillard JF. Assessment of bibliographic databases performance in information retrieval for occupational and environmental toxicology. *Occup Environ Med* 1998;55(8): 562-6.
- [9] Seglen PO. Why the impact factor of journals should not be used for evaluating research. *BMJ* 1997;314(7079):498-502.
- [10] Hirsch JE. An index to quantify an individual's scientific research output". *Proc Natl Acad Sci U S A* 2005;102(46):16569-72.