Javier Sanz-Valero[I,II]

Marcos Santos Ferreira[III]

Luis David Castiel[IV]

Carmina Wanden-Berghe[V]

Maria Cristina Rodrigues Guilam[VI]

Brazilian academic search filter: application to the scientific literature on physical activity

ABSTRACT

OBJECTIVE: To develop a search filter in order to retrieve scientific publications on physical activity from Brazilian academic institutions.

METHODS: The academic search filter consisted of the descriptor "exercise" associated through the term AND, to the names of the respective academic institutions, which were connected by the term OR. The MEDLINE search was performed with PubMed on 11/16/2008. The institutions were selected according to the classification from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for interuniversity agreements.

RESULTS: A total of 407 references were retrieved, corresponding to about 0.9% of all articles about physical activity and 0.5% of the Brazilian academic publications indexed in MEDLINE on the search date. When compared with the manual search undertaken, the search filter (descriptor + institutional filter) showed a sensitivity of 99% and a specificity of 100%.

CONCLUSIONS: The institutional search filter showed high sensitivity and specificity, and is applicable to other areas of knowledge in health sciences. It is desirable that every Brazilian academic institution establish its "standard name/brand" in order to efficiently retrieve their scientific literature.

DESCRIPTORS: Information Storage and Retrieval. Publications for Science Diffusion. Bibliography as Topic. Exercise. Motor Activity. Bibliometrics.

[I] Departamento de Salud Pública, Historia de la Ciencia y Ginecología. Universidad Miguel Hernández. Elche, España

[II] Departamento de Enfermería Comunitaria, Medicina Preventiva y Salud Pública e Historia de la Ciencia. Universidad de Alicante. Alicante, España

[III] Instituto de Educação Física e Desportos, Universidade do Estado do Rio de Janeiro. Rio de Janeiro, RJ, Brasil

[IV] Escola Nacional de Saúde Pública Sérgio Arouca. Fundação Oswaldo Cruz. Rio de Janeiro, RJ, Brasil

[V] Universidad Cardenal Herrera CEU. Elche, España

[VI] Centro de Estudos em Saúde do Trabalhador e Ecologia Humana. Fundação Oswaldo Cruz. Rio de Janeiro, RJ, Brasil

Correspondence:
Javier Sanz Valero
Departamento de Salud Pública
Historia de la Ciencia y Ginecología
Universidad Miguel Hernández
Carretera Nacional, N-332 , s/n
Sant Joan d'Alacant, Alicante, España
E-mail: jsanz@umh.es

INTRODUCTION

The search for solid and relevant scientific literature has become a necessity for any investigator in the scientific sphere. Having knowledge of existing bodies of work and their content is a pre-condition for resolving any informational problems that arise in the course of professional activity. It is necessary, though, to be aware of logical procedures that allow us to satisfactorily obtain these references in order to actually use them.

Quantitative analysis, to discover and evaluate scientific work in a research area, is currently becoming very important. It is part of the social study of science, and one of its main uses is in the area of science policy, providing tools that allow for evaluating the results of investigation. Therefore, considering the impact these measures have upon the allocation of research funds and upon professional promotion and accreditation of investigators, it is necessary to understand the particular details and the limitations that there use implies.

Being familiar with the scientific work of universities and research centers is very important to be able to perform these evaluations. In every case, it is

indispensable that the indicators of scientific production, as well as other indicators for Science and Technology, are reproducible with a set and generally accepted methodology, so that results can be compared and comparable.

Indices, quotients, obsolescence and other data can provide or end the possibility of accessing public or private financial resources, and also generate classifications which are of extreme importance for managers and evaluators of policy for science and technology.[13,3] In this sense, the evaluation and accreditation agencies, which judge the merit of students and investigators, value the importance of publication according to the prestige of the journal where it was published,[22] which is generally measured through bibliometric indicators.

Nonetheless, when evaluating the scientific literature, one should remember there are cases that can not be resolved with the use of one or multiple Medical Subject Headings (MeSH), which allow for the retrieval of the existing scientific work on a given subject, produced by specific institutions and countries.[12] This demonstrates the need to create geographical or academic search filters that ensure efficient access to this scientific literature.[15,19]

Therefore, the objective of this study was to develop a search equation to retrieve the academic scientific work pertaining to the theme of physical activity.

METHODS

The filter was created according to the methodology developed and tested by Valderas et al[19] for the creation of geographical search filters, incorporating corrections previously proposed by Sanz-Valero et al.[15] The distinct academic institutions and possible acronyms were associated with the "OR" connector, and different languages were employed for their identification in a pilot search. The academic/institutional search filter was corrected by incorporating the occurrences observed in this study.

- The final search filter consists of the boolean association between four equations:

- Equation 1: name of institution in Portuguese and in the main languages used in MEDLINE.

- Equation 2: official acronyms of Brazilian universities.

- Equation 3: name of the universities and corresponding name of the Brazilian cities, excluding those that could create confusion with cities in other countries.

- Equation 4: names of institutions that could not be included in the previous equations.

The final equation was structured in the following manner:

Exercise[Mesh] AND (((equation 1 OR equation 2) AND ("Brasil"[ad] OR "Brazil"[ad])) OR (equation 1 AND equation 3) OR equation 4)

The searches were performed from the first day available until November 16, 2008 (the last day that the obtained references were verified).

The project was done on MEDLINE which allows for searching mediated by Tags, which are classifiers of a bibliographic database, identified according to a label of two or more letters that can be added to each term with brackets. The Tag [ad] (address) was used to retrieve the institutional affiliation of at least the primary author of the article (e.g.: "Universidade Estadual"[ad] AND "Rio de Janeiro"[ad], is equivalent to having Universidades Estadual and Rio de Janeiro in the field for institutional address). The MEDLINE search engine, through PubMed, was used due to its free and permanent access, even though it is the most consulted biomedical database.[18]

The final search filter can be directly used by copying and pasting it into the PubMed search windows or by performing separate searches and creating the equation using the portal history. Any part of the final equation can be updated by adding new descriptors, changing a part or eliminating an undesired segment.

The filter was evaluated by manually reviewing the references obtained through the proposed search filter, while considering and scoring the previous studies. The institutional affiliation of the article (based on the first author) should belong to Brazilian academic centers or institutions. If articles met these requirements, they were categorized as "without incidence" or as "with incidence". It was documented for later review and correction, if the articles were retrieved from the search equation. The congruence of the retrieved articles to the area studied (exercise = physical activity) allowed for the classification of articles as "pertinent" or "not pertinent". Since a gold standard does not exist, the pertinence of the results was evaluated by comparing the manual review and the subsequent calculation of sensibility and specificity when utilizing the search equation.[15,19]

For the bibliometric evaluation the following variables were considered: document number and type, publication in electronic format (epub), number of authors, institutional affiliation, presence of institutional name in affiliation, language in which the country name appears, language the article was written, journal, year of publication, presence of a link on PubMed to the full text (visibility), open access availability of the article and access to the text on Scientific Electronic Library Online (SciELO).

The bibliometric indicators utilized were:

- Lotka's productivity index – based on the citations of scientific Publications, it allows for classification of the authors in three levels according to their productivity: large producers, with more than ten published works; medium producers that have published between two and nine works; and small producers that have published one work.

- Transience Index – frequency and percentage of authors or institutions that have published only one work about the subject.

- Burton and Kebler half-life – refers to the obsolescence of the works studied and is measured by the median age.

- Price Index – percentage of references equal to or less than five years.

- Bradford Law – indicator of the dispersion of scientific information, which holds that if the journals in a thematic area are divided into groups, then the number of journals in each group would be proportional to 1:n:n², where the main nucleus represents the network of journals of most pertinence to an area of knowledge.

RESULTS

The use of the proposed Boolean search equation retrieved 407 references, corresponding to 0.9% of the references on exercise and 0.5% of the Brazilian academic scientific production indexed in MEDLINE at search date.

The manual review of the retrieved references allowed for the consideration of the 407 (100%) articles pertinent to the theme of "exercise" (physical activity).

In regards to the institutional affiliation, 377 (92.6%) were in Portuguese or Spanish and 30 (7.4%) were not. The retrieval errors that occurred were as follows: on 19 (4.7%) of occasions the Southern Cross University of Australia, due to the translation of the Universidade Cruzeiro do Sul; three (0.7%) times the Santa Cruz de Tenerife (Spain) and one time the University of California at Santa Cruz (USA) occurred, due to the use of the term "Santa Cruz" for the Universities of Santa Cruz do Sul and Estadual de Santa Cruz; and in three other occasions the Catholic University of Sacred Heart (Italy) was retrieved when translating Universidade do Sagrado Coração. The other four (1.0%) incidences were due to confusion in abbreviations and word fragments that are difficult to predict and correct.

After corrections considering the observed incidences, the equation retrieved 381 references, of which 377 had a Brazilian affiliation. Therefore, comparing the

manual revision to verify the pertinence of the articles retrieved by the search equation (descriptor + institutional filter) showed a sensibility of 99.0% and a specificity of 100.0%.

The institutional affiliation was present in 194 (51.5%) of Portuguese publication, in 182 (48.3%) of English publications and in one (0.3%) of Spanish publications. In 12 (2.8%) occasions there was difficulty in recognizing the institution; seven (1.9%) of these were identified by referring to dependent institutions, and six (1.6%) were identified by the acronym.

In regard to the nomenclature of the country, it was written in 265 (18.3%) occasions in English (Brazil), in 69 (70.3%) occasions in Portuguese (Brasil), and in 43 cases (11.4%), the country did not appear.

The regards to the affiliation of the scientific work there were 54 academic institutions identified, of which three were in the first tertile of productivity: Universidade de São Paulo (USP), Universidade Federal de São Paulo (UNIFESP) and Universidade Estadual de Campinas (UNICAMP) (Table 1). The classification of productivity by institution according to Lotka's Index resulted in three levels of performance: 26 low production center, with only one work (48.2%); 18 medium production centers (between two and nine works) (33.3%); and ten high productions centers (ten or more works) (18.5%).

Of the 377 articles studied, 44 (11.7%) were reviews with an mean of 5.05 (SD = 0.13) authors per publication (CI 95%: 4.80; 5.30), with a minimum of one and a

Table 1. Universities with more than ten publications on physical activity, indexed in MEDLINE when using the Brazilian academic search filter.

University	f_0	%
Universidade de São Paulo (USP)	105	27.9
Universidade Federal de São Paulo (UNIFESP)	52	13.8
Universidade Estadual de Campinas (UNICAMP)	28	7.4
Universidade Federal do Rio Grande do Sul (UFRGS)	23	6.1
Universidade Federal de Pelotas (UFPEL)	17	4.5
Universidade Federal de Minas Gerais (UFMG)	16	4.2
Universidade Federal de São Carlos (UFSCAR)	15	4.0
Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP)	13	3.4
Universidade Federal do Rio de Janeiro (UFRJ)	11	2.9
Universidade do Estado do Rio de Janeiro (UERJ)	11	2.9

maximum of 17 (median = 5 and mode = 3). The main language of publication was English in 280 (74.3%) of articles, followed by Portuguese in 77 (20.4%) of articles; 17 (4.5%) of articles were found in English/ Portuguese, and three (0.8%) articles were in Spanish. The mean age of the retrieved articles was 4.25 years (SD = 0.22) (CI 95%: 3.81; 4.69), and their obsolescence (median of the Burton and Kebler index) was equal to three years. The percentage of documents with ≤5 years (Price index) was 74.3%.

It is notable that electronic versions (epub) exist for 105 (27.9%) articles, with a significant difference between publications before and after the year 2000 (p<0.001). The presence of a link from PubMed to the full text article (visibility) occurred in 274 (72.7%) occasions – of which 148 (39.2%) are available for free, 113 (30.0%) through the SciELO network and 126 (33.4%) by paying access fees – with similar significant differences between publications before and after the year 2000 (p<0.001).

The retrieved references were from 156 journals. The dispersion study of the retrieved scientific literature found a concentration of 127 (33.7%) articles in seven (4.5%) journals (Table 2); these documents consist of the main Bradford nucleus (Figure), which together with the other tertiles constitutes the dispersion of the publications.

DISCUSSION

The institutional search filter allowed for the retrieval of Brazilian academic publications in the MEDLINE database. The evaluation demonstrated a very good sensibility (capability of retrieving the desired publications) and an adequate percentage of pertinent articles, after the correction for the observed incidents. The retrieval of very pertinent publications may cause a large decrease in specificity, making the search less exhaustive. This fact is explained by the correction of the search filter and also the use of MeSH descriptors, but in any case, this situation favors the pertinence of retrieved publications.

Wild cards were not utilized in the search equations (for example: univers*) if they were not recognized by Tags, since the results would have been affected in a way contrary to the goal for formulating the equation.

A gold standard does not exist for comparison, but an evaluation can be done with already utilized techniques.[15,19] In comparison to previous publications, the performance obtained in this study shows a similar or superior sensibility as other recent studies[6,9,15,17,19,23,24] about search methodology. This same result also occurs for specificity.[6,9,11,23,24]

The progressive appearance of journals with electronic publication (epub), especially since the year 2000 in the case of Brazil, coincides with the progressive establishment of the SciELO.[2,10,21] The contribution of SciELO to the visibility of Brazilian and Latin American scientific literature is also notable,[14] since it provides links to full text articles found through MEDLINE searches, greatly increasing the visibility of these documents.[1,5,7]

The proposed filter can be improved by utilizing it and identifying new incidences not covered in the corrected version proposed here. A similar situation has successfully happened in the case of a Spanish geographic filter,[20,16] which is of a modular structure and can be easily modified through the addition or subtraction of any of its parts.

In many occasions the scientific documents can be retrieved for a given county without having such elaborate search strategies, but this depends on the characteristics of the study approach and the amount of error that can be accepted.

Table 2. Journals belonging to the main Bradford nucleus, where articles about exercise (physical activity) were published and obtained by utilizing the Brazilian academic search filter.

Abbreviation of the journal name	References		IF
	n	(%)	
Braz J Med Biol Res	30	8.0	1.150
Arq Bras Cardiol	28	7.4	-
Cad Saude Publica	22	5.8	-
Electromyogr Clin Neurophysiol	14	3.7	-
Rev Saude Publica	13	3.5	0.557
Eur J Appl Physiol	11	2.9	1.752
Arq Bras Endocrinol Metabol	9	2.4	-

* FIF = Impact factor, data obtained from 2007 JCR Science Edition Database of the ISI Web of Knowledge, Thomson Reuters ©
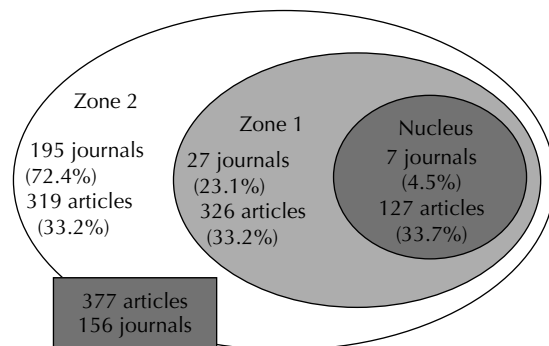

Figure. Bradford nucleus of dispersion of journals and articles on physical activity published by Brazilian institutions and indexed in MEDLINE.

Zone 2
195 journals (72.4%)
319 articles (33.2%)

Zone 1
27 journals (23.1%)
326 articles (33.2%)

Nucleus
7 journals (4.5%)
127 articles (33.7%)

377 articles
156 journals

It should be forewarned that the use of the filter currently generates the message "Quoted phrase not found" which is not an error message and does not interfere at all in the search process. The message is due to the non-recognition of some terms, such as "Municipal de Sao Caetano do Sul" because there is no reference that includes this in the Address field. Nonetheless, it was decided to include these locations to improve the topicality of the filter and in case of new publications by investigators from the affected institutions.

The percentage of retrieved documents will depend on the thematic area studied, independent of utilizing the proposed filter. Also, the institutions included in the Brazilian academic filter will have varying importance, depending on the document content considered.

The search equation included the academic and research institutions that are part of the classification in the Coordination for Training of Graduate Education (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, CAPES) for interuniversity agreements.

The results of the bibliometric analysis of scientific publications on exercise showed similar data to those presented by previous studies about health sciences,[4,8,22]

except in the case of obsolescence. This is due to the current situation of the theme studied and from the dispersion, since a pair of key journals, where the majority of authors want to publish, does not exist.

It can be concluded that the institutional filter is suitable, with high sensibility and specificity, for retrieving Brazilian academic scientific publications in MEDLINE, and the filter can be applied to any thematic area related to health sciences.

When using the entire institutional name together with the corresponding acronym, a great ease of identification was observed, especially in cases where the institutional name had been translated to a language other than Portuguese. Therefore, it is recommended that each Brazilian academic institution establishes a normalized name in order to facilitate the efficient retrieval of its scientific work in the different bibliographic databases. This recommendation is valid for any academic institution, independent of country.

In conclusion, this study offers an institutional filter to efficiently and easily retrieve the scientific work of Brazilian academic institutions and is applicable to political science studies.

REFERENCES

1. Antelman K. Do open access articles have a greater research impact? *Coll Res Libr.* 2004;65(5):372-82.

2. Barradas Barata R. SciELO Saúde Pública: o desempenho dos Cadernos de Saúde Pública e da Revista de Saúde Pública. *Cad Saude Publica.* 2007;23(12):3031-40. DOI: 10.1590/S0102-311X2007001200025

3. Camí J. Evaluación de la investigación biomédica. *Med Clin (Barc)*. 2001;117(13):510-3.

4. Casterá VT, Sanz-Valero J, Juan-Quilis V, Wanden-Berghe C, Culebras JM, García de Lorenzo y Mateos A. Estudio bibliométrico de la revista Nutrición Hospitalaria en el periodo 2001 a 2005: parte II, análisis de consumo; las referencias bibliográficas. *Nutr Hosp.* 2008;23(6):541-6.

5. Harnard S, Brody T. Comparing the Impact of Open Access (OA) vs. Non-OA articles in the same Journals. *D-Lib Magazine.* 2004[cited 2008 Dec 31];10(6). Available from: http://www.dlib.org/dlib/june04/harnad/06harnad.html

6. Jenuwine ES, Floyd JA. Comparison of Medical Subject Headings and text-word searches in MEDLINE to retrieve studies on sleep in healthy individuals. *J Med Libr Assoc.* 2004;92(3):349-53.

7. Lawrence S. Free online availability substantially increases a paper's impact. *Nature.* 2001 [cited 2008 Dec 31];411(6837):521. Available from: http://www.nature.com/nature/journal/v411/n6837/pdf/411521a0.pdf

8. Miralles J, Ramos JM, Ballester R, Belinchon I, Sevila A, Marangon M. Estudio bibliométrico de

la revista Actas Dermo-Sifiliográficas (1984-2003) II. Análisis de las referencias bibliográficas. *Actas Dermosifiliogr.* 2005;96(9):563-71. DOI:10.1016/S0001-7310(05)73137-9

9. Montori VM, Wilczynski NL, Morgan D, Haynes RB, Hedges Team. Optimal search strategies for retrieving systematic reviews from Medline: analytical survey. *BMJ.* 2005 [cited 31 Dec 2008]8;330(7482):68. Available from: http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=543864&blobtype=pdf

10. Rodrigues Pereira JC. Revista de Saúde Pública: forty years of Brazilian scientific production. *Rev Saude Publica.* 2006;40(N Esp):148-59. DOI:10.1590/S0034-89102006000400020.

11. Royle P, Waugh N. A simplified search strategy for identifying randomised controlled trials for systematic reviews of health care interventions: a comparison with more exhaustive strategies. *BMC Med Res Methodol.* 2005[cited 2008 Dec 31];5:23. Available from: http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=1183214&blobtype=pdf

12. Sampson M, Zhang L, Morrison A, Barrowman NJ, Clifford TJ, Platt RW, et al. An alternative to the hand searching gold standard: validating methodological search filters using relative recall. *BMC Med Res Methodol.* 2006[cited 2008 Aug 30];6:33. Available from: http://www.biomedcentral.com/content/pdf/1471-2288-6-33.pdf

13. Sancho R. Evaluación comparativa (Benchmarking) de las políticas nacionales de ciencia y tecnología en Europa. *Rev Esp Doc Cient.* 2001;24(3);315-8.

14. Sanz-Valero J. Visibility, the current challenge for publishing policy! *Cad Saude Publica.* 2007;23(12):2804-5. DOI:10.1590/S0102-311X2007001200001

15. Sanz-Valero J, Veiga de Cabo J, Rojo-Alonso C, D'Agostino MJ, Wanden-Berghe C, Espulgues Pellicer JX, et al. Los filtros metodológicos: aplicación a la búsqueda bibliográfica en la medicina del trabajo española. *Med Segur Trab.* 2008;54(211):75-83. DOI:10.4321/S0465-546X2008000200009

16. Sanz-Valero J, Wanden-Berghe C. El filtro geográfico español: Spain NOT Trinidad & Tobago. *Gac Sanit.* 2009;23(3):253. DOI:10.1016/j.gaceta.2008.09.002

17. Sladek RM, Tieman J, Currow DC. Improving search filter development: a study of palliative care literature. *BMC Med Inform Decis Mak.* 2007 [cited 2008 Dec 31];7:18. Available from: http://www.biomedcentral.com/content/pdf/1472-6947-7-18.pdf

18. Steinbrook R. Searching for the Right Search – Reaching the Medical Literature. *N Engl J Med.* 2006;354(1):4-7. DOI:10.1056/NEJMp058128

19. Valderas JM, Mendivil J, Parada A, Losada-Yáñez M, Alonso J. Construcción de un filtro geográfico para la identificación en PubMed de estudios realizados en España. *Rev Esp Cardiol.* 2006;59(12):1244-51. DOI:10.1016/S1885-5857(07)60080-2

20. Valderas Martínez JM. Usos de las restricciones geográficas en PubMed. *Rev Esp Salud Publica.* 2007;81(4):421-2. DOI: 10.1590/S1135-57272007000400009

21. Veiga de Cabo J. Visibilidad de revistas científicas e iniciativas para incrementar la difusión de las publicaciones españolas. *Nutr Hosp.* 2003;18(4):177-80.

22. Villar Álvarez F, Estrada Lorenzo JM, Pérez Andrés C, Rebollo Rodríguez MJ. Estudio bibliométrico de los artículos originales de la Revista Española de Salud Pública (1991-2000). Parte tercera: análisis de las referencias bibliográficas. *Rev Esp Salud Publica.* 2007;81(3):247-59. DOI:10.1590/S1135-57272007000300002

23. Wilczynski NL, Haynes RB, Hedges Team. Developing optimal search strategies for detecting clinically sound prognostic studies in MEDLINE: an analytic survey. *BMC Med.* 2004[cited 2008 Dec 31];2:23. Available from: http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=441418&blobtype=pdf

24. Wilczynski NL, Haynes RB, Hedges Team. Optimal search strategies for identifying mental health content in MEDLINE: an analytic: survey. *Ann Gen Psychiatry.* 2006 [cited 2008 Dec 31];5:4. Available from: http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=1440861&blobtype=pdf