

Revista de Saúde Pública: forty years of Brazilian scientific production

ABSTRACT

OBJECTIVE: To recognize the characteristics and path taken by the through analysis of the scientific production it has published over the period from 1967 to 2005.

METHODS: Scientometric methods were used to analyze reference data on the articles published in the Revista, retrieved from the databases ISI/Thomson Scientific (Web of Science), National Library of Medicine (PubMed) and Scientific Electronic Library Online (SciELO).

RESULTS: The Revista is the only Brazilian publication in the field of public health that is indexed by ISI/Thomson Scientific. It is prominent as a medium for publishing Brazilian scientific production in public health and is displaying a geometric increase in publication and citation, with annual rates of 4.4% and 12.7%, respectively. The mean number of authors per paper has risen from 2 to 3.5 over recent years. Although original research articles predominate, the numbers of reviews, multicenter studies, clinical trials and validation studies have been increasing. The number of articles published in foreign languages has also increased, accounting for 13% of the total, and the leading countries originating these are the UK, USA, Argentina and Mexico. The number and diversity of journals citing the Journal has been increasing, many of which are non-Brazilian. Authorship per author shows good fit to Lotka's Law, but the parameters suggest greater concentration and less dispersion than would be expected. Among the fields of interest of published papers, the following topics account for more than 50% of the total volume: infectious-parasitic diseases and vectors; health promotion, policies and administration; and epidemiology, surveillance and disease control.

CONCLUSIONS: The Revista shows great dynamism, without signs of abating or reaching a plateau any time soon. There are signs of progressively increasing complexity in the studies published, and more multidisciplinary work. The Revista seems to be widening its outreach and recognition, while remaining faithful to the field of public health in Brazil.

KEYWORDS: Scientific publications. Periodicals, history. Periodicals, trends. Public health.

Departamento de Epidemiologia.
Faculdade de Saúde Pública. Universidade
de São Paulo. São Paulo, SP, Brasil

Correspondence:

Júlio Cesar Rodrigues Pereira
Departamento de Epidemiologia - FSP/USP
Av. Dr. Arnaldo, 715
01246-904 São Paulo, SP, Brasil
E-mail: julioicrp@usp.br

Received: 6/7/2006

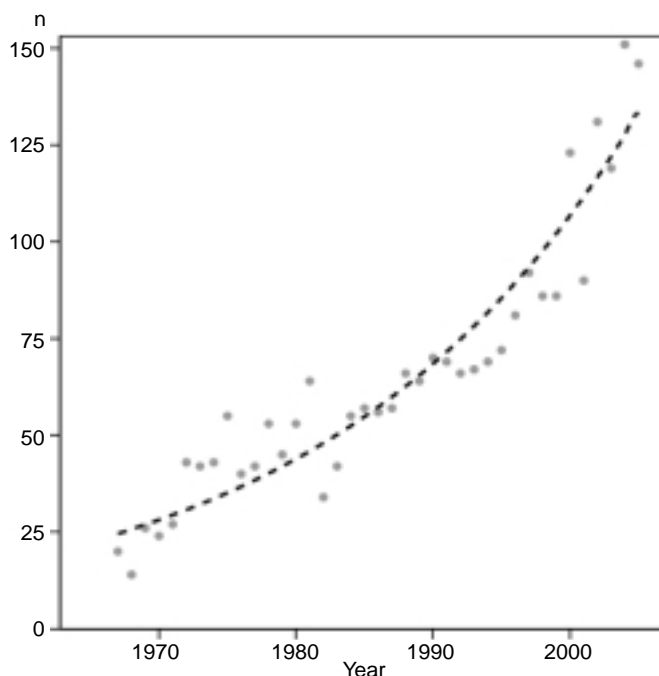
INTRODUCTION

The *Revista de Saúde Pública* is reaching its fortieth anniversary distinguished as one of the leading journals in the field of public health in Brazil. In fact, preliminary analysis of the data from the project “Characterization of Postgraduate Programs in Public Health in Brazil”,* in which all the scientific production from postgraduate lecturers within this field that is registered in the Thomson Scientific database (formerly ISI, the Institute for Scientific Information) up to 2005 is being studied, places the *Revista* as the first choice for the publication of their work. In this analysis, it has been seen that, with regard to indexed journals in that database, these authors concentrate the publication of their work in seven journals, which alone account for one-third of the production from this group of authors:

- 1° *Revista de Saúde Pública*, with 939 articles;
- 2° *Memórias do Instituto Oswaldo Cruz*, with 163 articles;
- 3° *Brazilian Journal of Medical and Biological Research*, with 145 articles;
- 4° *American Journal of Epidemiology*, with 128 articles;
- 5° *Journal of Dental Research*, with 127 articles;
- 6° *Transactions of the Royal Society of Tropical Medicine and Hygiene*, with 57 articles;
- 7° *Circulation*, with 56 articles.

This group of journals has been identified as the hard core of publications from postgraduate lecturers in public health, by applying Bradford's Law.¹ This predicts that, among the set of publications within a scientific discipline, the numbers of journals in the first, second and third terciles of production obey an order of 1:n:n². The records in the aforementioned project include 4,842 publications and the Bradford sets have been identified in a ratio that is close to the theoretical prediction: 1:n:n^{1.8}.

Bradford's theory was among the first contributions towards originating scientometry, a scientific discipline that is well established today. From the middle of last century onwards, it became greatly boosted through procedures for bibliometric analysis, thanks to the consolidation of the records relating to scientific production, on a regular basis. In the present article, bibliometric records are analyzed with a view to offering a portrait of the scientometric base for these forty years of the *Revista de Saúde Pública*.



Source: PubMed, May 2006

Figure 1 - Scatter plot for the numbers of articles by the *Revista de Saúde Pública*, by year of publication for the period from 1967 to 2005, and curve fitted to the data.

METHODS

The *Revista* has had its publications registered in the MEDLINE database of the National Library of Medicine, in the United States, since 1967; in the Social Science Citation Index and the Web of Science, since 1982; and in SciELO (Scientific Electronic Library Online), an initiative coordinated by BIREME-PAHO/WHO (Latin American and Caribbean Center on Health Sciences Information) and by the *Fundação de Amparo à Pesquisa do Estado de São Paulo* (Fapesp - State of São Paulo Research Foundation), since 1977. On the basis of the records in the Web of Science, MEDLINE/PubMed and SciELO, it was sought to characterize the *Revista* over the period from 1967 to 2005, according to the patterns of:

- 1) production, through studying the publication dynamics and characteristics of the articles over this period;
- 2) reach and repercussion, through studying their citations;
- 3) authorship, through fitting this to scientometric laws and by institutional authorship; and
- 4) topics of interest, through analyzing keywords.

Details of the methodological procedures for the analysis will be presented together with the results from each of these databases. The methodological description will be restricted to what is necessary for

*Interim data from a project still in progress.

correct interpretation of the results and as required for clarifying their presentation.

RESULTS

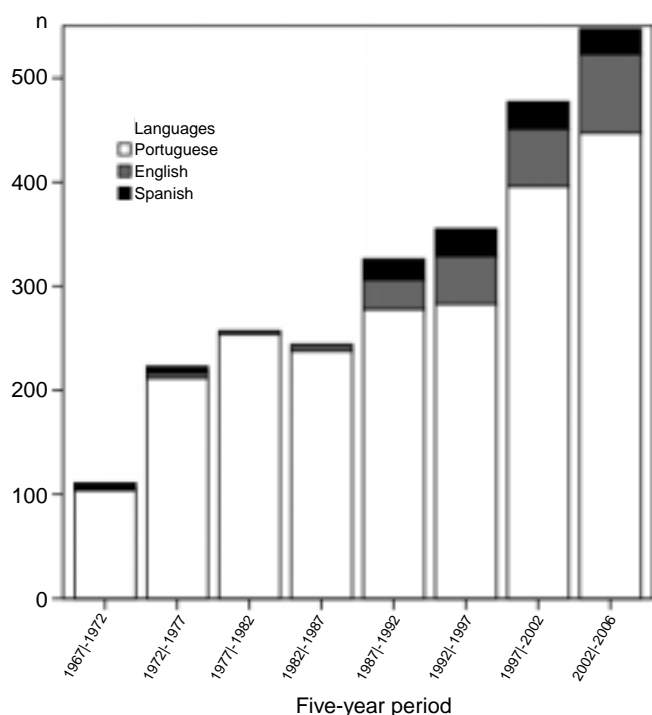
Production

It is the MEDLINE/PubMed databases that gather together the greatest quantity of information on the articles published by the *Revista*: there are records of 2,540 documents of different types, of which 20 were in 1967 and 146 were in 2005. Figure 1 shows the scatter of the annual numbers of items published over this period, and also a curve fitted using the least-squares method as a suggestion for the inferred behavioral pattern.

The curve suggested in Figure 1 has a good fit ($R^2=0.89$) and is described by the following function:

$$n = 24,64 \cdot e^{0,044 \cdot (\text{ano} - 1967)} \quad (1)$$

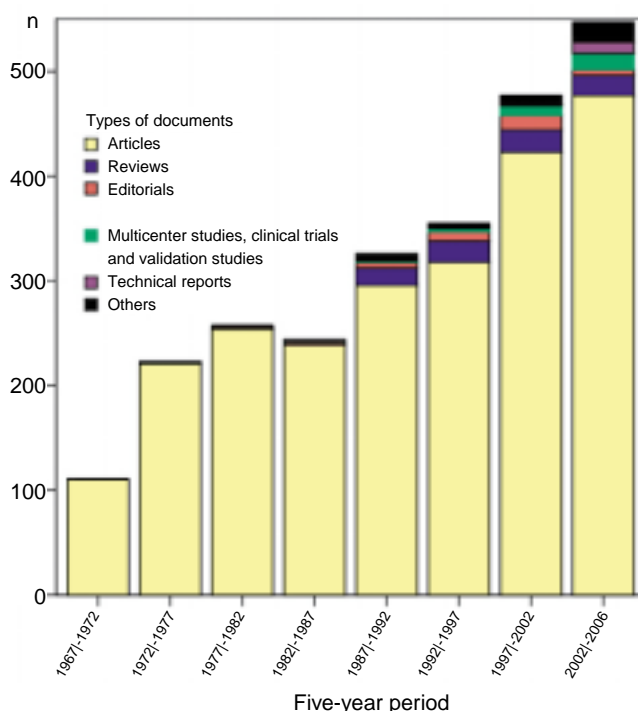
in which the parameters indicate, respectively:



*From 2003 onwards, the Journal also started to be published in English in its electronic version.

**ISI/Thomson Scientific (Web of Science), May/2006. The most recent five-year period is incomplete, only covering four years.

Figure 3 - Languages* of the articles published by the *Revista de Saúde Pública*, by five-year period between 1967 and 2005.**



*ISI/Thomson Scientific (Web of Science), May/2006. The most recent five-year period is incomplete, only covering four years

Figure 2 - Types of documents published in the *Revista de Saúde Pública*, by five-year period between 1967 and 2005.*

- 24.64: The *Revista* started by publishing around 25 articles per year;
- 0.044: It has been growing geometrically at a mean rate of 4.4% per year;
- 1967: The initial year was 1967, i.e. the counting of time started with that year ($t_0=1967$)

According to the classification by types of publication in the PubMed database, scientific articles account for 92% of all types of documents published, but by examining the trends, it can be seen that, over more recent five-year periods, other types have progressively gained space in the *Revista* (Figure 2).

For the same period, the PubMed records also make it apparent that, despite the predominance of the Portuguese language (87% of the total), both Spanish and English have been increasing the participation over recent years, as shown by Figure 3. It should be noted that, in its electronic format, the *Revista* has also been published in the English language since 2003.

In the Web of Science records, the countries of origin of the authors of documents pub-

Table 1 - Country of origin of the publications in the Revista de Saúde Pública, according to the periods indicated, from the ISI records.*

Country of origin (original notation)	Period										Total n %	
	1982 - 1987		1987 - 1992		1992 - 1997		1997 - 2002		2002 - 2006			
	n	%	n	%	n	%	n	%	n	%	n	%
Brazil	343	97.7	419	89.0	306	82.5	419	85.5	457	83.4	1.944	87.1
USA	3	0.9	11	2.3	9	2.4	12	2.4	13	2.4	48	2.2
England	3	0.9	11	2.3	8	2.2	7	1.4	14	2.6	43	1.9
Argentina			5	1.1	17	4.6	7	1.4	13	2.4	42	1.9
México			2	0.4	7	1.9	21	4.3	11	2.0	41	1.8
Chile			7	1.5	5	1.3	7	1.4	2	0.4	21	0.9
Venezuela					5	1.3	1	0.2	5	0.9	11	0.5
Colombia			4	0.8	1	0.3	3	0.6	2	0.4	10	0.4
Spain					2	0.5	4	0.8	4	0.7	10	0.4
Cuba			5	1.1	1	0.3	1	0.2	2	0.4	9	0.4
Germany	2	0.6	3	0.6					3	0.5	8	0.4
Canada			2	0.4	2	0.5	2	0.4	1	0.2	7	0.3
France			2	0.4	1	0.3	2	0.4	1	0.2	6	0.3
Portugal					1	0.3			3	0.5	4	0.2
Croatia									3	0.5	3	0.1
Switzerland					1	0.3			2	0.4	3	0.1
Australia							2	0.4	1	0.2	3	0.1
Japan									2	0.4	2	0.1
Peru					2	0.5					2	0.1
Uruguay					1	0.3			1	0.2	2	0.1
Austria									1	0.2	1	0.0
Denmark									1	0.2	1	0.0
Bolivia							1	0.2			1	0.0
India					1	0.3					1	0.0
Israel									1	0.2	1	0.0
Italy									1	0.2	1	0.0
Mozambique					1	0.3					1	0.0
Nigeria									1	0.2	1	0.0
Norway									1	0.2	1	0.0
Poland									1	0.2	1	0.0
Sweden							1	0.2			1	0.0
Ukraine									1	0.2	1	0.0
Total	351	100.0	471	100.0	371	100.0	490	100.0	548	100.0	2.231	100.0

*ISI/Thomson Scientific (Web of Science), May/2006

lished by the *Revista* can be ascertained. Over the period from 1982 to 2005, there were a total of 2,241 publications, of which 2,231 have valid records of their country of origin. By dividing this period into five-year periods, the information presented in Table 1 is obtained. Similar information can also be obtained from SciELO for the period 1997-2005, but preference was given to the Web of Science because of its broader coverage.

Reach and repercussion

Also from the Web of Science records, the behavioral patterns of citations of articles published by the *Revista* can be ascertained. Among the 593 journals in the field of "General Social Sciences", within which the *Revista de Saúde Pública* is indexed, it is ranked 238th in numbers of citations, although it is 577th in numbers of citations per article. These statistics re-

late to May 2006 and are available under the heading "Essential Science Indicators", in which data relating to the previous ten years plus the current year are analyzed, with updating every two months. The Web of Science presents information processed as five-year periods, as shown by the data in Table 2.

An examination of the annual citation rates for the articles published since 1982 suggests a tendency towards each article resulting in two citations, as shown in Figure 4.

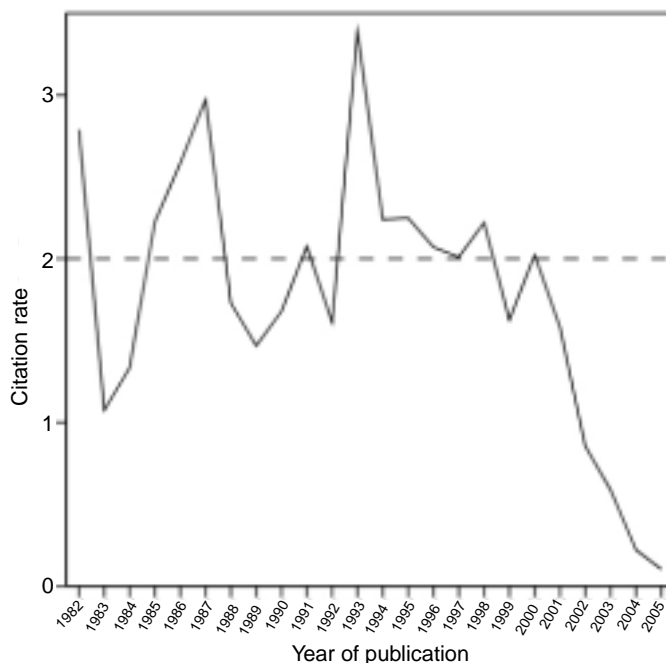
By examining Figure 2, from the most recent years to the earlier years, it can be seen that this mean rate of two citations per article is achieved after around five to seven years after publication. For the last ten years and the year 2006, the consultation of the Web of Science in May 2006 registered 1,312 citations for 1,044 articles, this giving a ten-year rate of 1.26 citations per article.

Table 2 - Reproduction of the impact measurements for the Revista de Saúde Pública presented by the Essential Science Indicators of the ISI/Thomson Scientific database, May 2006.

Citation data in 5-year intervals	1996-2000	1997-2001	1998-2002	1999-2003	2000-2004	2001-2005	2002-2006
# of papers	427	441	456	486	539	581	636
Times cited	173	203	216	268	323	387	350
Citations per paper	0.40	0.46	0.47	0.55	0.59	0.66	0.55

Recently, with the implementation of the “Essential Science Indicators”, the ISI/Thomson Scientific database has restricted the calculation of the impact factor (number of citations of the articles published over the last two-year period divided by the number of articles published during this two-year period) to the so-called Science Edition of the Journal Citation Reports (JCR). Until then, the impact factor for journals in the Web of Science - Social Sciences database appeared in the JCR, and the *Revista de Saúde Pública* registered an impact factor of around 0.35 citations per article.

However, the impact factor for the *Revista* can be gauged from the SciELO database, which was implemented with pioneering participation from the *Revista*. There, in a consultation in May 2006, the current impact factor found was 0.1038, using a two-year calculation basis, and 0.1333, for a three-year basis. Taking the two-year impact factors in a historical series, figures ranging from 0.2 to 0.4 were found. It was noted that, although the impact rates originated from different data collections, they converged towards similar figures, thereby suggesting the strong inductive inference that this is a fair measurement of the *Revista’s* performance.

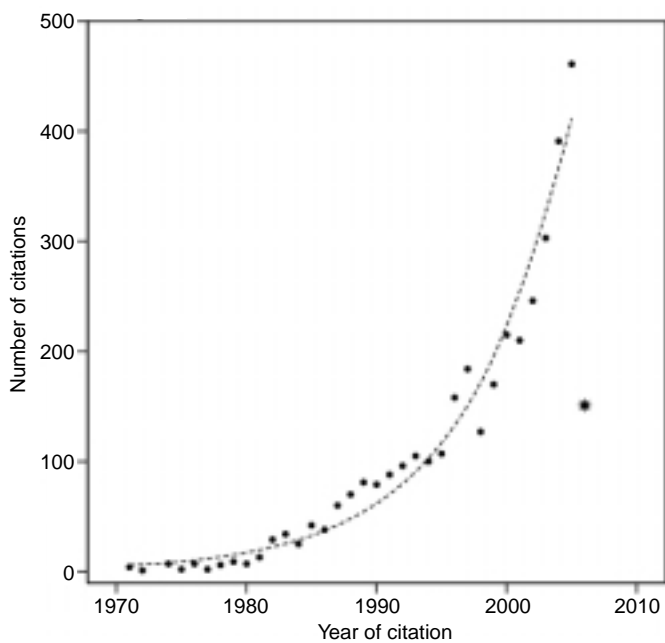


*Number of citations / Number of articles published, ISI/Thomson Scientific (Web of Science), May 2006

Figure 4 - Annual citation rate* for the *Revista de Saúde Pública*, according to records from ISI/Thomson Scientific, 1982 to 2005.

In SciELO, in the same way as previously done by JCR, there is the additional information of calculations of the half-lives of citations (the time taken for 50% of the total number of citations to occur, or the median of the age of the citations). At present, the half-life is 6.38 and, from examination of the historical series, it was observed that it ranged from four to eight years. Again, this information is consistent with the citations register in the Web of Science, which was analyzed in Figure 4, and this suggests that after around ten years, all the citations expected have been concluded.

In addition to now offering statistics on 158 indexed journals with the same quality as in the Web of Science, SciELO also offers free access to each of them in digital files. Any edition of the *Revista de Saúde Pública* can be freely consulted, from 1977 up to the present date, and PubMed offers a link for such consultations. This, together with the growing share of articles in the English language (Figure 3), establishes the conditions needed for an increase in its reach and progressive growth in its citations. Just as in the production, the numbers of citations have also grown exponentially, as suggested by Figure 5.



Source: ISI/Thomson Scientific (Web-of-Science), May 2006

Figure 5 - Numbers of citations of the *Revista de Saúde Pública*, by year of occurrence of the citation, and curve fitted to the data.

By fitting the function suggested by Figure 5 to the data (note that 2006, which is still incomplete, has been disregarded because it is aberrant), an excellent fit is obtained ($R^2 = 0.97$) for an exponential function of the following form:

$$n = 5,52.e^{0,127(\text{ano} - 1971)} \quad (2)$$

in which the parameters, as in the case of the production, indicate respectively:

- 5.52: The *Revista* started with around five citations per year;
- 0.127: The numbers have been increasing geometrically at a mean rate of 12.7% per year;
- 1971: The *Revista* started to be cited in 1971.

It is worth noting that the growth rate for citations is more than twice the growth rate for production. This suggests that, in addition to vegetative growth resulting only from the increase in the numbers of articles published and their citation yield (two per article published), the growth in citations can be attributed to other factors. Among the candidates for such explanatory variables are both the progressive increase in texts in English and the greater ease of access provided by SciELO.

The sources of citation of the *Revista* today differ from those of its early years. The first five-year period with records in the Web of Science (1982-1986) allows the citation pattern to be ascertained: there were 553 citations distributed among 112 journals. Over the last five-year period (2002 to 2006), because of the long half-life there were only 238 citations, but these were distributed among 136 journals. The final five-year period has around half the number of citations seen in the first five-year period, but still has practically the same number of citing journals: in the first five-year period there were 4.9 citations per citing journal, and 1.75 in the final period. This degree of concentration would be greatest if equal to the number of citations (all citations in the same journal) and least if equal to one (each citation in a different journal). From the first to the last five-year period there was a reduction in concentration of 64.3%, such that the last five-year period is only 35.7% of the first, i.e. 2.8 times lower in concentration or, said in another way, 2.8 times greater outreach.

From the Web of Science records, it can also be ascertained which journals made these citations of the *Revista de Saúde Pública*. To find out about these while taking into account the passage of time, the period beginning in 1971 was divided into decades up to the present date, with the fourth decade starting in 2000. Since it is not possible to list all the journals that have already cited the *Revista de Saúde Pública*, a selection method has to be determined. Taking into consideration that, during the first decade, six journals registered at least two citations of the *Revista de Saúde Pública*, it was established that these six would be identified. For the subsequent periods, the quantity of journals registering at least two citations of the *Revista* grew prohibitively, which called for a different cutoff method that still respected the cutoff rationale established for the first decade. Given that, according to the number of citations, the hierarchy of citing journals had a Pareto distribution, it was defined that the number of citing journals to be presented for each period would be proportional to the dispersion parameter of a Pareto function fitted to the data in each period. Thus, independent of the number of journals, those with the greatest contribution were selected for highlighting, according to the characteristics of the period. The Pareto probability density function is a continuous asymmetrical function that can roughly be imagined as the right half of a normal distribution curve, and it has the following form:

For the purposes of its application here, it can be simplified to the following description:

$$f(x) = \frac{\alpha\beta^\alpha}{x^{\alpha+1}}$$

in which the parameters c and a represent, respectively:

$$f(x) = \frac{c}{x^a}$$

- c : frequency level that indicates the first place among citing journals;
- a : shape of the distribution, describing its decay rate or dispersion. For $a=0$, the distribution curve results in a horizontal straight line from the initial level, while for $a=8$, the curve results in a straight

Table 3 - Descriptive statistics on the citations during the periods indicated and for characterizing each period using a Pareto distribution.

Statistics	Periods			
	1971 - 1981	1981 - 1991	1991 - 2001	2001 - maio 2006
Number of citations	45	471	1,350	1,762
Number of citing journals	29	71	253	509
Fit of the Pareto function	$R^2=0.93$	$R^2=0.99$	$R^2=0.99$	$R^2=0.98$
Starting level: proportion of citations in the first citing journal	0.16	0.52	0.43	0.21
Decay rate: dispersion	0.72	1.42	2.15	1.48

Source of the adjusted data: ISI Web of Science, May 2006.

line that descends abruptly before reaching the second rank in the hierarchy. Between these values, the curve has a fairly smooth shape reminiscent of the right half of a normal curve.

Consequent to defining this methodological strategy, Table 3 shows the statistics on the citations of the *Revista de Saúde Pública* and Table 4 lists the journals highlighted in the four periods.

Authorship

For the 2,540 articles registered in the PubMed database, 4,575 authors were found, who wrote between one and 128 articles for the *Revista*. In bibliometry, the number of authors for an article is an indirect measurement of its complexity, if conditioning factors of another order such as possible authorship due to comradeship are ignored in principle. Over this study period of almost 40 years, the *Revista* has shown

a year-by-year progressive increase in the mean number of authors per article, consistent with what is a universal trend of growing complexity in scientific investigations (Figure 6).

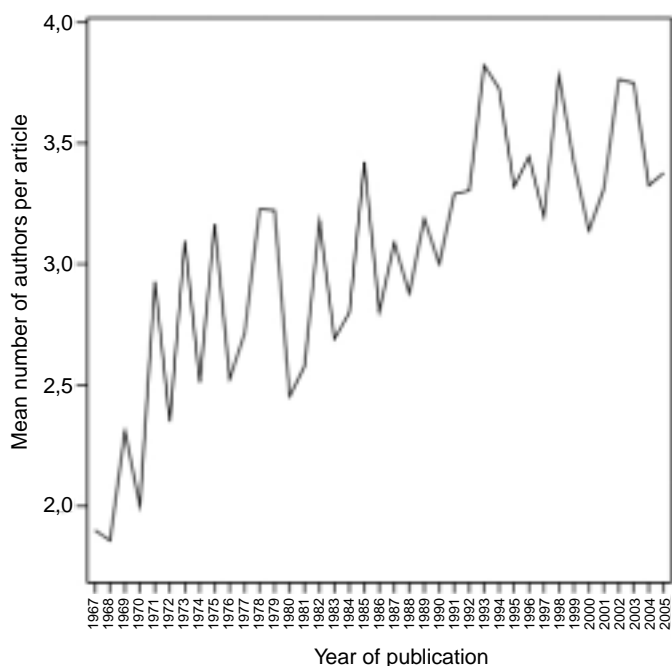
Since each article may have more than one author, the 4,575 authors gave rise to 8,104 authorships. The canons of scientometry state that the distribution of authorships among authors within a given field of scientific production follows Lotka's Law.³ According to this theory, which is also called the inverse-square law, 60% of authors would have only one authorship and this proportion would decrease strongly, inversely to the square of the number of authorships. Thus, authors with one authorship would represent 60%, with two authorships 15%, and so on, in accordance with the following function:

$$\text{Proporção de autores com } n \text{ autorias} = \frac{0,6}{n^2}$$

Table 4 - Journals citing the *Revista de Saúde Pública*, by period indicated.

Period	Journals highlighted	n° of citations
1971 – 1981	Transactions of the Royal Society of Tropical Medicine and Hygiene	7
	Boletín de la Oficina Sanitaria Panamericana	5
	American Journal of Tropical Medicine and Hygiene	4
	Journal of Medical Entomology	2
	Bulletin of Entomological Research	2
1981 – 1991	Revista Brasileira de Pesquisas Médicas e Biológicas	2
	Revista de Saúde Pública	268
	Revista do Instituto de Medicina Tropical de São Paulo	46
	Transactions of the Royal Society of Tropical Medicine and Hygiene	15
	Memórias do Instituto Oswaldo Cruz	14
	Journal of the American Mosquito Control Association	11
	International Journal of Epidemiology	6
	Brazilian Journal of Medical And Biological Research	6
	Interciência	6
	Journal of Medical Entomology	5
1991 – 2001	American Journal of Tropical Medicine and Hygiene	5
	Journal of Tropical Pediatrics	4
	Revista de Saúde Pública	579
	Memórias do Instituto Oswaldo Cruz	138
	Revista do Instituto de Medicina Tropical de São Paulo	24
	American Journal of Tropical Medicine and Hygiene	24
	Journal of Medical Entomology	20
	Archivos Latinoamericanos de Nutricion	20
	International Journal of Epidemiology	18
	Transactions of The Royal Society of Tropical Medicine and Hygiene	17
	Journal of The American Mosquito Control Association	16
	Social Science & Medicine	15
	Salud Pública de México	14
	Brazilian Journal of Medical and Biological Research	11
	Bulletin of The World Health Organization	11
Annals of Tropical Medicine and Parasitology	11	
2001 – maio 2006	Arquivos de Neuro-Psiquiatria	8
	Medical and Veterinary Entomology	8
	Paediatric and Perinatal Epidemiology	8
	International Journal of Geriatric Psychiatry	7
	Revista de Saúde Pública	375
	Memórias do Instituto Oswaldo Cruz	126
	Revista da Sociedade Brasileira de Medicina Tropical	46
	Revista Panamericana de Salud Pública-Pan American Journal of Public Health	43
	Journal of Medical Entomology	35
	Arquivos de Neuro-Psiquiatria	26
	American Journal of Tropical Medicine and Hygiene	20
	Archivos Latinoamericanos de Nutricion	20
	International Journal of Epidemiology	16
Salud Pública de México	16	
Medical and Veterinary Entomology	13	
Community Dentistry and Oral Epidemiology	13	

Source: ISI Web of Science, May 2006.



Source: PubMed, May 2006

Figure 6 - Mean number of authors per item published, for each year of PubMed records.

Figure 7 shows the relative frequencies of the authors according to the number of authorships and a data curve adjusted to Lotka's Law.

The Lotka function fitted to the *Revista's* production data ($R^2 = 0.99$) takes on the following form:

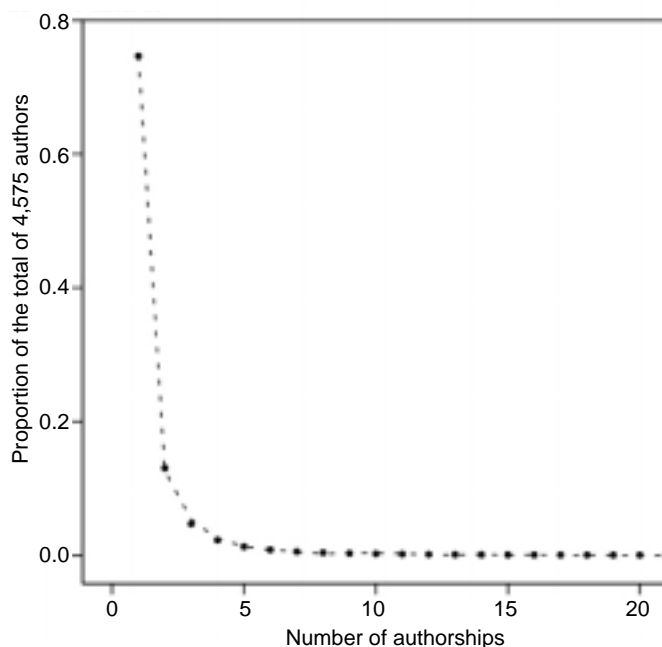
$$P(\text{autor com } n \text{ autorias}) = \frac{0,75}{n^{2,5}} \quad (3)$$

This configuration of the distribution of authorships suggests that the *Revista's* community of authors is eclectic, given the large proportion of authors with only one article published (75%, or 25% greater than the 60% expected by Lotka's Law) and the narrow variability (2.5, or also 25% greater than the 2 expected by Lotka's Law: the greater the exponent is, the lower the variability and the flatter the curve are).

Ninety-three percent of the authors of the *Revista de Saúde Pública* are authors with only one, two or three contributions. Although this large proportion of authors with a small number of publications suggests a lack of any parochialism in the *Revista*, this must also be considered in the light of the field of knowledge that it serves. Pedrosa,*

studying the scientific production of post-graduate lecturers in a sample from all the programs accredited by *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* (Capes - Coordination for the improvement of higher education personnel), found that a Lotka function fitted to these data also suggested a large concentration of authors with few publications: 69% of the authors recognized through the Web of Science records had only one authorship and the dispersion was less than expected (the exponent of the denominator of the Lotka function was 2.24).

By consulting the Web of Science database, the institutional authors that have contributed most to the *Revista* can be determined (Table 5). It is worth noting that there has been a progressive reduction in the proportional participation by the *Universidade de São Paulo* (USP - which hosts the *Revista*), giving room for other institutions, such as the *Universidade Federal de São Paulo* (Unifesp), which had tripled its participation by 2005. The *Universidade Estadual de Campinas* (Unicamp) had a regular participation in the three periods; the *Universidade Federal de Pelotas* (UFPeL) started to appear in the 1990s and proportionally in-



Source: PubMed, May 2006

Figure 7 - Proportion of authors of the *Revista de Saúde Pública*, according to the number of authorships and adjusted Lotka function.

*Pedrosa FM. Caracterização da saúde coletiva no Brasil segundo sua produção científica registrada no ISI [Master's dissertation]. São Paulo: Faculdade de Saúde Pública da USP; 2005.

creased its participation in the 2000s; the State Health Department of São Paulo and its research institutes dropped out of the top ten institutions in terms of proportional participation in the 2000s, while the *Universidade Federal da Bahia* (UFBA) recovered its position that had been lost in the 1990s.

Topics of interest

To classify the topic of interest of a publication, the Web of Science allocates each journal registered in its database to a certain field of scientific knowledge. This is a rough classification that fits each item published by a journal into a large field, in accordance with the scope of topics of the journal. The *Revista de Saúde Pública* is classified in the field of "Public, Environmental & Occupational Health". This type of classification is greatly used in scientometric studies and allows profiles to be determined for scientific production in given fields of interest. For example, a study on Brazilian scientific production over the period from 1981 to 1995⁴ enabled the identification of physics, biology and biochemistry, clinical medicine, engineering, chemistry and plant and animal sciences as the fields of knowledge that were most prominent within Brazilian science. For the present study, for which the field of interest was one journal, this classification was of no help for characterization purposes. For this reason, it was necessary to resort to the keywords of the articles published. Both the Web of Science and PubMed include keywords for each bibliographic item

registered in their databases, although they do this distinct ways that do not necessarily agree.

Analysis of keywords is recognized as inadequate for characterizing publications. It finds better application for use in database searches with the purpose of identifying articles dealing with similar subjects. Even for this purpose, its performance is modest. Over the last few years there has been some progress towards overcoming these limitations by means of conceptual structure recognition techniques in scientific texts. One initiative of this type, the technique from Collexis5, a German company that provides services for the World Health Organization (WHO - Alliance for Health Policy and Systems Research), and for some journals such as those from the publisher Elsevier, have been the subject of investigations and tests by BIREME/SciELO. This technique recognizes a "fingerprint" in every word of a given text: a root that concords with a Thesaurus of specialized terms, from which it classifies the degree of intensity with which this fingerprint is recognized in the text.

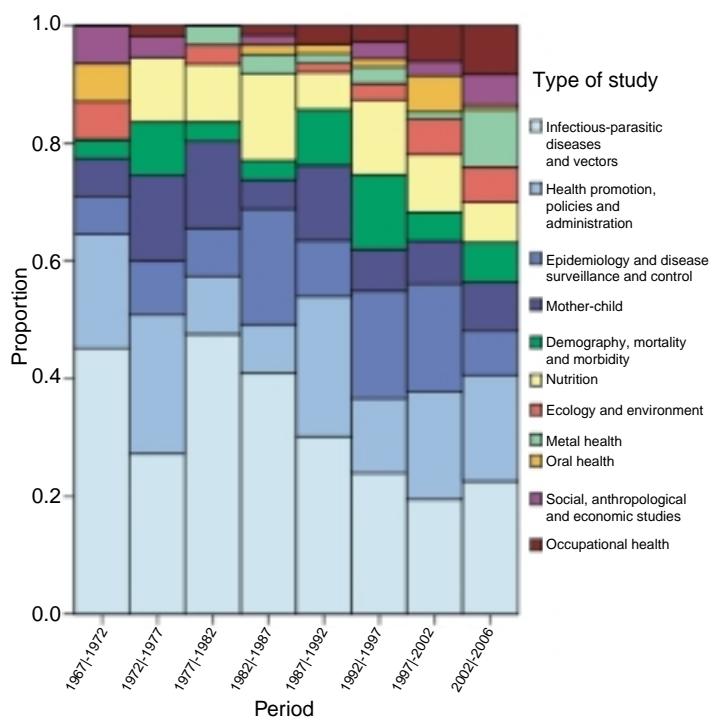
These reflections are necessary because, to analyze the topics of interest in the *Revista de Saúde Pública*, only the keywords can be made use of in analyses, among the data available, and caution is required in any interpretation of the information derived from these keywords.

To analyze the keywords in the articles published

Table 5 - Ten main institutional authors of the *Revista de Saúde Pública* over the periods indicated.

Year	Institutional author	N of articles	Total proportion of production during the period
80	Universidade de São Paulo	366	58.3%
	Universidade Estadual de Campinas	28	4.5%
	Universidade Federal de Pelotas	12	1.9%
	Universidade Federal de São Paulo	12	1.9%
	Secretaria da Saúde do Estado de São Paulo	10	1.6%
	Universidade Federal da Bahia	10	1.6%
	Instituto Butantan	9	1.4%
	Superintendência de Controle de Endemias	9	1.4%
	Universidade Federal do Maranhão	8	1.3%
	Universidade de Brasília	8	1.3%
90	Universidade de São Paulo	335	39.7%
	Universidade Estadual de Campinas	37	4.4%
	Universidade Federal de Pelotas	25	3%
	Instituto Adolfo Lutz	17	2%
	Universidade Estadual Paulista	17	2%
	Universidade Federal de Minas Gerais	17	2%
	Universidade Federal do Rio Grande do Sul	17	2%
	Universidade Federal de São Paulo	17	2%
	Escola Nacional de Saúde Pública	16	1.9%
	Universidade Estadual de Maringá	16	1.9%
2000	Universidade de São Paulo	235	29.8%
	Universidade Estadual de Campinas	47	6%
	Universidade Federal de São Paulo	46	5.8%
	Fundação Instituto Oswaldo Cruz - Mato Grosso do Sul	33	4.4%
	Universidade Federal de Minas Gerais	32	4.1%
	Universidade Federal de Pelotas	32	4.1%
	Universidade Federal do Rio Grande do Sul	26	3.3%
	Universidade Federal da Bahia	23	2.9%
	Universidade Estadual Paulista	18	2.3%
	Universidade Federal de Santa Catarina	16	2%

Source: ISI/Thomson Scientific (Web of Science), May 2006.



Source: PubMed, May 2006, MeSH Major Topics edited for study categories.

Figure 8 - Revista de Saúde Pública: proportional shares of the different types of studies with statistically significant associations with the periods indicated.

by the *Revista*, those registered by PubMed were chosen, given that this database covers the whole period of existence of the *Revista*. Among these keywords, interest was restricted to the “MeSH Major Topics”, which are keywords at the highest level of aggregation that supposedly best represent the content of each article. “Supposedly”, because although

something like “*Creatinine/*urine*” would represent a high level of aggregation (a study on creatinine in the urine, without clarifying the purpose), it still denotes a specific aspect of the study, not a general characteristic. For 2,540 publication items that were recorded in PubMed for the *Revista*, 4,719 key words of “MeSH Major Topic” type were identified, which in PubMed are highlighted by an asterisk. These 4,719 keywords were used in 6,466 valid cataloging records (non-empty) for published items (a mean of 2.55 keywords per article: minimum of 0, maximum of 8, mode =2), of which only 842 had at least two occurrences per five-year period of existence of the *Revista*.

These 842 keywords accounted for 2,589 records of MeSH Major Topics, a volume that when distributed in a contingency table according to the *Revista*'s five-year periods showed a non-random distribution (Fisher exact test; $p=0.000$). Residual analysis⁵ on this contingency table allowed identification of 557 records of 508

different MeSH Major Topics that characterized the *Revista*'s five-year periods by statistically significant associations. Since this was still a very large volume for reporting in the present study, the MeSH Major Topics were recoded in a discretionary manner into 11 study categories, thus resulting in what is presented in Table 6 below, and summarized in Figure 8.

Table 6 - Numbers and proportions of MeSH Major Topics with statistical associations with periods of publication of the *Revista de Saúde Pública*, according to type of study and publication period.

Type of study	1967 – 1973		1972 – 1977		1977 – 1982		1982 – 1987		1987 – 1992		1992 – 1997		1997 – 2002		2002 – 2006		Total	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Infectious-parasitic diseases and vectors	14	45.2	15	27.3	29	47.5	25	41.0	19	30.2	17	23.9	16	19.5	30	22.6	165	29.6
Health promotion, policies and administration	6	19.4	13	23.6	6	9.8	5	8.2	15	23.8	9	12.7	15	18.3	24	18.0	93	16.7
Epidemiology and disease surveillance and control	2	6.5	5	9.1	5	8.2	12	19.7	6	9.5	13	18.3	15	18.3	10	7.5	68	12.2
Mother-child	2	6.5	8	14.5	9	14.8	3	4.9	8	12.7	5	7.0	6	7.3	11	8.3	52	9.3
Nutrition			6	10.9	6	9.8	9	14.8	4	6.3	9	12.7	8	9.8	9	6.8	51	9.2
Demography, mortality and morbidity	1	3.2	5	9.1	2	3.3	2	3.3	6	9.5	9	12.7	4	4.9	9	6.8	38	6.8
Occupational health			1	1.8			1	1.6	2	3.2	2	2.8	5	6.1	11	8.3	22	3.9
Mental health					2	3.3	2	3.3	1	1.6	2	2.8	1	1.2	13	9.8	21	3.8
Ecology and environment	2	6.5			2	3.3			1	1.6	2	2.8	5	6.1	8	6.0	20	3.6
Social, anthropological and economic studies	2	6.5	2	3.6			1	1.6			2	2.8	2	2.4	7	5.3	16	2.9
Oral health	2	6.5					1	1.6	1	1.6	1	1.4	5	6.1	1	0.8	11	2.0
Total	31	100	55	100	61	100	61	100	63	100	71	100	82	100	133	100	557	100

Source: PubMed, May 2006, MeSH Major Topics edited for study categories.

Placing the keywords into categories according to type of study for the present study involved arbitrary judgment, and readers' tolerance of this is requested, given that exhaustive description of the 557 "MeSH Major Topics" would be prohibitive, both because of their extensiveness and because of the difficulty in interpretation.

COMMENTS

Over the 40 years of the *Revista de Saúde Pública*, it has exhibited significant dynamism, as suggested well by functions (1) and (2). Following the criticism of the Malthusian model of exponential growth, the expectation that there has been for dynamic models describing natural phenomena is that, at some point, any growth would come up against a limit and that the intrinsic nature of the phenomenon would prohibit the surmounting of this limit. It would therefore be expected that, after the long period of 40 years, the growth in the annual number of publications and citations would already have reached its natural limit and that the function that described its dynamics would be asymptotic to this limit. However, the *Revista* still does not show any behavior suggestive of stabilization that would allow it to be argued that it might some day find a limit to its growth. The technological revolution that today allows articles to be published in the form of digital files that are accessible through the Internet has broken through the narrow limits of an economic and logistical nature that previously restricted the publishing of journals on paper. The *Revista de Saúde Pública* is still published in both forms, but today there are exclusively electronic journals and the future for journal publication seems to point in this direction. This may come to pass, and perhaps the growth limit for publications within a field of knowledge may one day come to be the exhaustion of that field's capacity to produce knowledge.

The growing complexity of the studies published (Figure 6), allied with the characteristics of function (3), suggests multidisciplinary work in the context of the articles published by the *Revista*. In fact, these characteristics suggest that authors from other fields of knowledge may be present in the *Revista's* articles as collaborators with public health authors: multidisciplinary teams that are needed to cope with public health studies that are progressively more complex. Even if the high concentration of authors with few studies published is considered to be a result from the characteristics of the field of public health, as duly signaled in the results, it cannot be neglected that the parameters of function (3) suggest that the

Revista has a greater concentration of authorship or, expressing this in another way, greater diversity of authors, than in the cited study by Pedrosa* on this field. The growth and diversification of citing journals (Tables 3 and 4) reinforces the idea of greater coverage, complexity and diversification.

The numbers of items published in English and Spanish have increased (Figure 3) and both the types of documents published (Figure 2) and the sources of citation of the *Revista* (Table 4) have diversified, although the same cannot be said regarding the types of studies. Looking at Figure 8, the impression is that the topics of infectious-parasitic diseases and vectors; health promotion, policies and administration; epidemiology and disease surveillance and control; mother-child health; demography, mortality and morbidity; and nutrition have notable participation in the *Revista* that is regular and stable, and which has varied little over the five-year periods analyzed. However, proportional growth can be highlighted for, on the one hand occupational health and social, anthropological and economic studies (with progressive increases over recent years) and, on the other hand, ecological and environmental studies (with initially erratic appearance but subsequent regular increases in proportional participation over more recent five-year periods).

As signaled in the text, this information on types of study requires caution in its interpretation. The factors limiting the faithfulness of the *Revista's* portrait that are highlighted are the restriction of the analysis only to the MeSH Major Topics, the arbitrary reduction of these to the ones with two or more presences per five-year period, the subsequent reduction of the latter to those with statistical associations with the five-year periods, and finally the arbitrary categorization by type of study. The only excuse for tolerating so many provisos is that, otherwise, any analysis would have had to be abandoned.

In other respects, the results have to be examined while taking their sources into account. The data analyzed originated from three distinct sources: PubMed, Web of Science and SciELO. These databases are independent and cover different periods and aspects of the classification of a publication. They are not necessarily concordant and may even diverge in the numbers of articles recorded for the *Revista* over a given period. In fact, some items cataloged by the Web of Science as "Item about an individual", "Book review", "Correction, addition", "Editorial material" and even "Article" are sometimes not found in PubMed, and

*Pedrosa FM. Caracterização da saúde coletiva no Brasil segundo sua produção científica registrada no ISI [Master's dissertation]. São Paulo: Faculdade de Saúde Pública da USP; 2005.

vice versa. The classification of documents differs from one bibliographic database to another: here, the PubMed database was utilized (Figure 2) because it has the greatest coverage, albeit without this ensuring greater faithfulness. With regard to the citations, it should be remembered that these relate exclusively

to the Web of Science database, which explains the absence of *Cadernos de Saúde Pública*, a journal published by *Escola Nacional de Saúde Pública* (ENSP - National Public Health School) and a sister publication of *Revista de Saúde Pública* in the family of Brazilian public health publications.

REFERENCES

1. Bradford SC. Sources of information on specific subjects. *Eng.* 1934;137:85-6.
2. Bassanezi RC. Modelagem matemática. São Paulo: Editora Contexto; 2002 p. 80-1.
3. Lotka AJ. The frequency distribution of scientific productivity. *J Wash Acad Sci.* 1926;16:317-23.
4. Pereira JCR, Escuder MML, Zanetta DMT. Brazilian sciences and government funding at the State of São Paulo. *Scientometrics* 1998;43:177-88.
5. Pereira JCR. Análise de dados qualitativos. 3ª ed. São Paulo: EDUSP; 2004.