

Is there a relationship between article downloaded and scientific production? A case study: BIBLIOSAN - The Italian biomedical libraries system

Authors:

Moreno Curti (1),
Angela Aceti (2),
Massimo Casciello (6)
Luisa Fruttini (3),
Luisa Garau (4),
Saba Motta (5),
Renato Piccinin (6),
Roberta Querini (7),
Daniela Simone (8),
Franco Toni (9)

Affiliations:

- (1) San Matteo Hospital Foundation, IRCCS, Pavia, Italy
- (2) The Italian Authority for Occupational Health and Safety (ISPESL), Rome, Italy
- (3) Umbria and Marche Regions Veterinary Public Health and Research Institute, Perugia, Italy
- (4) Lombardia and Emilia Romagna Regions Veterinary Public Health and Research Institute, Brescia, Italy
- (5) C. Besta Neurological Hospital Foundation, IRCCS, Milan, Italy
- (6) Directorate General of Science and Technology Research, Ministry of Health, Rome, Italy
- (7) Santa Lucia Hospital Foundation, IRCCS, Rome, Italy
- (8) 'Giovanni Paolo II' Oncological Hospital, IRCCS, Bari, Italy
- (9) The Italian National Institute of Health (ISS), Rome, Italy

Correspondence:

ftoni@iss.it

General Framework:

Bibliosan is the library system financed by the Italian Ministry of Health, which is composed of 55 biomedical research institutes. From 2007 a network was activated which permits the possibility of sharing electronic resources and allows one to retrieve more than 7000 titles available of which 5000 journals are subscribed

Purpose of the Study:

To verify if the availability for electronic documentation allows:

1. An increase in the scientific production of the institutes, which belongs to the Bibliosan Network
2. An improvement in the scientific quality of the articles produced by the Institutes

The answer to the first question involves a **quantitative analysis** of document flow while the second one is **qualitative**

Quantitative analysis:

- One must consider the number of resources available, how it is used and the number of articles or documents produced
- Substantially this means analysing the relationship between the quantity of articles downloaded by the researchers and the quantity of their scientific productivity → number of articles downloaded / number of articles published
- It seems like a relatively simple operation but instead it consists of solving a series of problems of major importance with regard to resource usage and scientific production

Problems related to resource usage:

Obtain the data compatibility through COUNTER statistic usage
Not all publishers adopt COUNTER standard
The importance of activating SUSHI (Standardized Usage Statistics Harvesting Initiative) in an automated collection and standardization of data supplied by editors
The lack in the majority of OA publications of statistical data about the number of accesses or downloads of articles (Only BioMed Central furnishes statistics.)

The quantity of the OA databases (journals, repositories) utilized is determined only by activating a TLA recognition (Transaction Log Analysis), with its existing difficulties

Problems associated with scientific productivity:

The type of publication which must be taken into consideration (e.g. including or not proceedings, posters, reports, etc. which are not published in scientific journals)
Time Gap (over many years) beginning from when a resource is downloaded up until the publication time of the article causes a distortion of statistical data

Qualitative analysis:

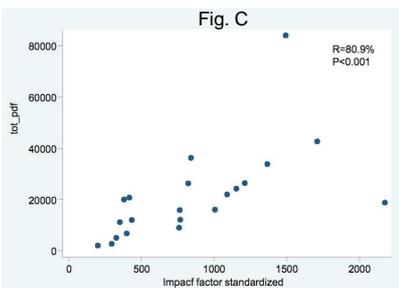
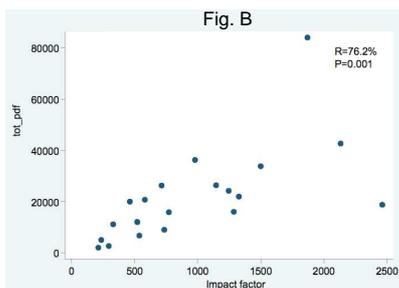
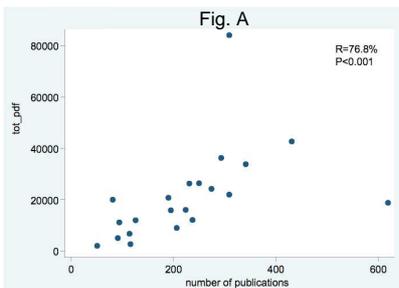
2 aspects to consider:

1st aspect:

The research benefits which can be obtained from the quality and type of documentation made available to researchers.
This type of analysis can be conducted for example by verifying how many of those cited publications researchers retrieved through the journal collections made available by the Network.
The difficulty lies in the fact that this is not an investigation which can be done using an automated data processing system and in many cases the researcher does not acknowledge in the bibliographical references if the information is obtained from a digital resource.
However, it can be useful to do a sampling of a limited number of selected units based on the disciplinary nature and by using non-systematic recognition. This way one can verify if there is a discrepancy in the results of the different scientific sectors

2nd aspect:

The Citation impact, that is to verify if, by improving the quality offered, increases the quantity of articles published in impacted journals, the number of citations for authors, the H-Index for single authors.
In this case problems which can arise are those typically related to IF use: need to precisely separate the disciplinary areas in order to obtain reliable values, presence of self-citations, distinction between positive and negative citations, etc...
This kind of recognition needs to start from research conducted on selected samples and in specific disciplinary fields.



Data Analysis:

The number of total articles downloaded in 2007 by the 21 Research Institutes examined has been related with the following models:

Fig. A: Number of articles published in 2007 on journals with IF

Fig. B: Total of IF articles

Fig. C: IF 2007 standardized on the basis of Ministry of Health criteria

Fig. D: FTE

Fig. E: Bedspace

Figures 1-3 show as for every model (A, B, C) there is a significant relationship between the results of research activity (evaluated by the number of articles and IF) and the use of electronic resources.

Less evident is the correlation between downloads and bedspace while is absent that between downloads and FTE.

These results show as the size of the Institutes don't have influence on the use of scientific documentation.

So it has been applied a bivariate analysis with a RRM (Robust Regression Model) in order to evaluate the presence of a relationship between standardized IF and downloads eliminating the influence of bedspace. The obtained results (R= 73% and P= 0,002) are very significant.

