WHO EMRO | Development of an Eastern Mediterranean Region search strategy for biomedical citations

PDF version

Ghazi O. Tadmouri[1], Ahmed Mandil[2] and Arash Rashidian[2]

تطوير استراتيجية لإقليم شرق المتوسط للبحث عن الاستشهادات المنقولة عن البحوث الطبية البيولوجية المفهرسة في قاعدة بيانات PubMed

غازي تدمري، أحمد منديل، آرش رشيديان

"الخلاصة": لقد استخدمت قاعدة البيانات PubMed ، وهي قاعدة بيانات ببلوغرافية " مفتوحة " تغطي التخصصات الطبية والصحية، في تحديد عدد كبير من المؤشرات التي تساعد في تحليل الاتجاهات العالمية في إنتاج البحوث الطبية البيولوجية . وتمثِّل الدراسة الحالية محاولة أولية لتطوير بنية بحث خاصّة بإقليم شرق المتوسط وتحسينها على النحو الأمثل في PubMed لتمهيد الطريق لإجراء تحليلات وصفية لاحقة . وتتضمَّن الاستراتيجية المثالية للبحث لإقليم شرق المتوسط تركيب تفصيلي للجُمل مما يسهّل التحكم في عملية البحث ويحافظ على التوازن السليم بين حساسية النتائج التي عثر عليها ودقتها . وقد جرى تحرٍ لبيانات الاستشهادات الخاصة بالبلاد يدوياً للكشف عن النتائج الإيجابية الكاذبة . وتشير نتائجنا إلى أن إنتاج البحوث المنشورة قد زاد بنحو خمسة أضعاف في إقليم شرق المتوسط خلال الفترة ما بين 2004-2013. ووُجد أن خمسة بلدان فقط (هي جمهورية إيران الإسلامية ومصر والمملكة العربية السعودية وتونس وباكستان) قد ساهمت بمقدار 80 ٪ من جميع منشورات إقليم شرق المتوسط خلال هذه الفترة . وساهمت كل من البلدان السبعة عشر المتبقية في إقليم شرق المتوسط بما يقل عن 4 %. ونحن نعتقد أن المنهجية المعروضة في هذه الدراسة يمكن استخدامها جنباً إلى جنب مع مقاييس أخرى لاستخراج مؤشرات لا تقدَّر بثمن لوصف نُظُم البحوث الصحية في الإقليم .

ABSTRACT PubMed, a ‘barrier-free’ bibliographic database covering biomedical and health disciplines, has been successfully used to identify a multitude of indicators that assist in analyzing global trends for biomedical research productivity. The current study represents an original attempt to develop and optimize an Eastern Mediterranean Region (EMR) search strategy in PubMed to pave the way for subsequent descriptive analyses. The refined EMR search strategy contains elaborate syntaxes which facilitate controlling the search process and maintaining a proper balance between sensitivity and precision of the obtained results. Country-specific citation data were manually scanned for false positive publications. Our results indicate that publication productivity increased nearly five-fold in the EMR from 2004 to 2013.

1 / 11

Five countries (Islamic Republic of Iran, Egypt, Saudi Arabia, Tunisia and Pakistan; in order of total publications) contributed to 80% of all EMR publications during this period. Each of the remaining 17 EMR countries contributed less than 4%. We believe that the methodology presented in this study can be used in conjunction with other metrics to extract invaluable indicators to describe EMR health research systems.

Mise au point d'une stratégie de recherche pour la Région de la Méditerranée orientale en matière de citations biomédicales indexées dans PubMed

RÉSUMÉ PubMed, une base de données bibliographiques en libre accès, couvrant les disciplines des sciences biomédicales et de la santé, est utilisée avec succès pour identifier de nombreux indicateurs qui permettent d'analyser les tendances mondiales en matière de productivité de la recherche biomédicale. La présente étude constitue une tentative originale en vue de définir et d'optimiser une stratégie de recherche pour la Région de la Méditerranée orientale dans PubMed afin d'ouvrir la voie à des analyses descriptives ultérieures. Cette stratégie optimisée comporte des syntaxes complexes qui facilitent le contrôle du processus de recherche et assurent un juste équilibre entre la sensibilité et la précision des résultats obtenus. Les données de citation spécifiques à chaque pays ont été soumises à un examen manuel afin de repérer les fausses publications positives. Les résultats que nous avons obtenus indiquent que la productivité des publications a quasiment été multipliée par cinq dans la Région de la Méditerranée orientale entre 2004 et 2013. L'étude a montré que seuls cinq pays (la République islamique d'Iran, l'Égypte, l'Arabie saoudite, le Pakistan et la Tunisie ; par ordre de nombre de publications totales) ont contribué à 80 % de toutes les publications pour la Région de la Méditerranée orientale pendant cette période. La contribution de chacun des 17 pays restants de cette Région a été inférieure à 4 %. Nous pensons qu'il est possible d'utiliser la méthodologie présentée dans cette étude en association avec d'autres mesures afin de définir des indicateurs précieux permettant de décrire les systèmes de recherche en santé au sein de la Région de la Méditerranée orientale.

[1]Faculty of Public Health, Jinan University, Tripoli, Lebanon (Correspondence to: Ghazi Tadmouri: ph@jinan.edu.lb). [2]Department of Information, Evidence and Research, WHO Regional Office for the Eastern Mediterranean, Cairo, Egypt.

Introduction

Health research is an important aspect of health care delivery, since it plays a significant role in the global economic growth and contributes to improving the living standards and quality of life (1). Assessing the quantity and quality of the scientific output in a country or region is a key

indicator to understand and improve its research system.

Bibliometric analysis of scientific publications involves the application of mathematics and statistical methods to determine the extent and quality of research in a given territory (2,3). In the biomedical and health fields, scientists utilize many high-quality citation databases to search for published literature. Of these databases, PubMed has been acknowledged to be the most significant ‘barrier-free’ biomedical resource available on the World Wide Web (4).

PubMed citations come from the MEDLINE© (Medical Literature Analysis and Retrieval System Online, or MEDLARS Online) bibliographic database, manuscripts deposited in the PubMed Central (PMC) free digital repository, and freely accessible citations from biomedical books published in the National Center for Biotechnology Information (NCBI) Bookshelf (5). In fact, PubMed provides a strong health discipline indexing coverage, and currently catalogues over 26 million biomedical articles that were published in more than 44 000 journals in 37 languages (6). PubMed depends on a text-based search that uses an indexing system for rapid retrieval of information. Effective search strategies provide references that are more specific for the intended topic compared with other popular search engines. In PubMed, the citation information is broken into index fields (e.g., journal name, author name, title, primary author’s address, language of publication, and others). The power of PubMed search could then be further enhanced by the use of search rules, syntaxes, and qualifying terms in combination with search field abbreviations. PubMed has been used with success to perform bibliometric studies aiming at the assessment of various aspects of research outputs (7).

The free access to PubMed and its friendly interface have led to the development of elaborate techniques to analyze global trends for biomedical research productivity and provide objective and useful tools to evaluate the results of scientific activity in different locations worldwide (7,8–10). To the best of our knowledge, most of these studies required manual checks at the data collection phase for the quality of the data collected. There are inherent difficulties when the search strategies are intended towards certain geographical locations, as most validated search strategies are not limited by geographical concerns. No single study has been undertaken to consider the comprehensive collection of biomedical and health research outputs in different countries of the WHO Eastern Mediterranean Region (EMR; Figure 1). While a previous study used a simple search strategy to identify “access to medicine” studies related to EMR (11), there is a need for a search strategy that is tested and applicable to different research needs. It is for this reason we aimed in the present study at the development and optimization of the EMR search strategy in PubMed to pave the way for subsequent analyses that could be based on the automated monitoring of biomedical research outcomes in the Region, and the subsequent instantaneous forecasting of research activities in the region as a prerequisite for proper policy development.

Methods

We conducted a comparative analysis of three types of search strategies at PubMed (http://www. pubmed.com) to obtain a collection of biomedical and health research citations published between 2004 and 2013, by principal researchers affiliated to institutions from any country in the EMR. The main outcome of this comparative analysis is to develop an optimized EMR Search Strategy that can be used in performing automated collections of biomedical citations from the Region, which could be used in various analysis or for forecasting research activities in the Region. The search strategies that were implemented included: a “classical” search strategy, a “pitfall” search strategy, and a suggested EMR search strategy (Supplement 1 [online]) that combines the benefits of the former strategies.

The “classical” search strategy was based on simple queries using standard Medical Subject Heading (MeSH) terms corresponding to EMR countries and the [AD] tag which has the function of collecting all published articles carrying the requested country name in the affiliation (or address) field (e.g. Iran [AD], Egypt [AD], Saudi Arabia [AD]). The “pitfall” search strategy involved the use of elaborate syntaxes previously developed by Tadmouri and Bissar-Tadmouri (12–14). A previously developed search strategy for EMR (Pre-EMR) contained more sophisticated syntaxes that allowed further control over the search processes. In certain cases, the search was either “specific”, based on excluding false-positive results using the Boolean operator NOT; or “sensitive”, based on the use of the inclusive Boolean operator OR, and/or the “star” wildcard character (*); or “optimized”, based on a mixed use of the “star” wildcard character and the OR and NOT Boolean operators (Supplement 1 [online]).

For some countries, variant names in several world languages were also incorporated in the search syntaxes to cover non-English citations (14). All types of search strategies were directed to PubMed within a limit of few hours on 30 July, 2015. Since address-based searches on PubMed automatically exclude letters to the editors and commentary articles, citations that were investigated in this study included reviews, original journal articles, and case reports. In addition, address-based searches also restricted results to papers in which only the principal investigators are affiliated to institutions located in the EMR.

The resulting country-specific citation data were collected in an offline file of Extensible Markup Language (XML) database format and then converted into Excel tables. Subsequently, the address field in citation records were manually scanned for false positives or inconsistent addresses (12). In many instances scientific papers suffer from a variety of inconsistencies including the absence of uniformity in reporting addresses at the level of city, institution, faculty, and department names (15), the transliteration of addresses from native languages into English or French (12) and the use of misspelled names or abbreviations to express names of

universities or research centers (14).

The leading author in this study undertook the task of deleting false-positive entries and building up an exclusion dictionary to formulate a new enhanced EMR search strategy comprising optimized search strategies for citations from EMR countries (Supplement 1 [online]). With the exception of Djibouti and Pakistan, retrieving data from PubMed for the rest of the EMR countries required careful implementation of elaborate queries that are specific or sensitive due to the occurrence of false-positive entries. This task becomes more imperative because the National Library of Medicine ceased, at the beginning of October 2013, performing quality control review and editing of the author affiliation field in citations indexed in PubMed and started to rely on data supplied directly by journal publishers (16). Therefore, the method used in this study is not recommended to be used for papers published since 2014.

Following data collection and control reviews, specific-country databases were built using text-file systems that included raw Medline Format citations in the health disciplines and published by first authors affiliated to EMR countries during years 2004–2013. Data from these raw Medline files were then transferred, according to the need, to other flat-file database containers to conduct various types of descriptive analyses and draw relevant statistics, especially those related to the geographic distribution of health research in the EMR (Table 1).

Results

During the period from 2004 to 2013, biomedical and health publication productivity increased nearly five-fold in the EMR region (Table 2). A few countries represent the majority of publications from the Region. This study indicated a dominance of the biomedical research publications by Iranian scholars, who contributed to a sizeable portion—39.3% of all EMR publications during the study period (2004–2013). In the EMR, the share from the Islamic Republic of Iran is almost equivalent to the combined productivities of researchers from the following four countries: Egypt (14.1%), Saudi Arabia (10.6%), Tunisia (8.1%), and Pakistan (7.8%). The remaining 17 EMR countries had each a contribution of less than 4% and an overall EMR share of about 20% (Table 1).

Using the EMR search strategy approach (described above), 140 911 citations were found to be indexed in PubMed for all EMR countries during the study period 2004–2013 (Table 1). This overall figure is not very different from overall figures obtained using the “classical” and “pitfall” search strategies. By looking at the details, however, each of the search strategies exhibited some peculiar characteristics for a number of EMR countries, and had significant limitations as compared with the EMR search strategy as described below.

Using the classical search strategy, data extracted for Afghanistan, Bahrain, Djibouti, Egypt, Iraq, Islamic Republic of Iran, Kuwait, Libya, Oman, Pakistan, and Saudi Arabia (half of the countries in the Region) remained very close to those obtained using the refined EMR search strategy (95–100%; Table 1).

In the case of Qatar, Sudan, and Yemen 6–7% of the citations were excluded when applying the EMR search strategy because some of the papers' authors affiliated to the institutions in these countries were not principal investigators. This observation was further noted in the cases of Jordan and Lebanon that included 21% and 45% false-positive citations, respectively. Part of this deviation is blamed on the fact that many of the excluded citations included Jordanian and Lebanese authors who were not principal investigators. Yet, a larger part of this deviation is explained by the fact that Jordan and Lebanon could also refer to several cities and neighborhoods in Great Britain, Northern Ireland, and the United States (12). Because of all reasons mentioned above, the classical search strategy proved to be less sensitive for countries like the Syrian Arab Republic (8% less sensitive), the United Arab Emirates (20% less sensitive), and Palestine (29% less sensitive). The extreme positions of Tunisia (53% less sensitive) and Morocco (66% less sensitive) is mainly due to the fact that the classical search strategy is not capable of detecting address details when written in French, as many authors from these countries publish in the French language.

Using the "pitfall" search strategy, citation data for most of the EMR countries came closer to the results obtained by the application of the refined EMR search strategy (95–105%; Table 1). Data for Libya, Qatar, Sudan, Syrian Arab Republic and Yemen contained non-specific citations for the reason that in approximately 6–7% of those papers; authors affiliated to institutions in these countries were not principal investigators.

The pre-EMR search strategy was formulated with the aim to increase the sensitivity of the "pitfall" search strategy mainly by the proper use of the wildcard character "*" (e.g. Syrian Arab Republic), the use of country name abbreviations (e.g. United Arab Emirates) or polymorphs (e.g. Palestine), and the use of language-specific country name variants (e.g. Morocco, Tunisia, and Lebanon). This model worked well for many of the EMR countries, but returned false-positives especially in the case of Jordan (21%), Palestine (26%), Lebanon (47%), and Somalia (78%). This is mostly because those country names intersect with various world locations and also because many citations did not include first authors affiliated to these countries. This observation becomes more obvious when EMRSS:Pre-EMRSS ratios are computed for the top most health research productive countries for each year of the studied period (Table 2). In the majority of the cases, most of the deviation occurred in data extracted for the year 2013. The least affected country is the Islamic Republic of Iran with a ratio of 0.98

for the year 2013 compared to a 0.99 overall ratio. Egypt, Tunisia, Pakistan, and Morocco were moderately affected (2013 ratios: 0.94–9.96 vs. overall ratio: 0.99). Saudi Arabia and Kuwait were significantly affected (2013 ratios: 0.90–9.93 vs. overall ratio: 0.99). In the cases of Lebanon and Jordan, deviating ratios were not restricted to the year 2013 and occurred throughout the studied period (Table 2).

Discussion

Our study provides systematic evidence of important errors and pitfalls that may occur if sub-optimal search strategies are used for the identification of geographically bound publications. More importantly, we demonstrated that such errors are not equally distributed among the countries, and different countries are affected with different types of searching errors and by varying states. As such, our proposed refined EMR search strategy is a useful tool for future studies that look for EMR publications. It also provides a platform for future research on how to assess and test alternative search strategies concerned with other bibliographic databases, different regions or different time-frames.

Biomedical and health publication productivity increased nearly five-fold in the EMR region in the study period. Five countries in the region represented 80% of biomedical and health research publications indexed in the PubMed (2004-2013): Islamic Republic of Iran (39.3%), Egypt (14.1%), Saudi Arabia (10.6%), Tunisia (8.1%), and Pakistan (7.8%).

Assessing biomedical and health research outputs in a specific region is key for the evaluation and improvement of its research productivity and direction. In the study period regional publications increased nearly five-fold. Despite their many imperfections, indicators of biomedical bibliometrics offer a means by which countries and regions of varying geographic sizes and socio-economic development may be monitored and compared (17,18).

Using inappropriate search strategies on PubMed may lead to conclusions that are contradictory to the realities with regards to biomedical research activity in the region (19). For this reason we aimed in this study to develop a validated EMR Search Strategy that is both sensitive and specific, while attempting to retrieve all possible biomedical research citations produced in EMR countries. The implementation of the EMRSS would help in collecting citations missed by non-sensitive strategies and to automatically avoid false-positive citations for EMR countries that have names overlapping with several world geographical locations.

By comparing the number of citations obtained using the “classical” search strategy, which

depends on the use of standardized MeSH terms for EMR countries, versus our optimized EMRSS search strategy, results demonstrated that our structured approach allowed us to enhance the sensitivity and/or specificity of representative data for several countries in the Region. However, the NLM's recent policy to include the affiliation data for all authors citations indexed in PubMed caused serious deviations in the results, especially for data for the year 2013. If a selective search method to assemble citations based on the country of affiliation of the first authors is not devised, it will be too difficult to conduct similar analyses beyond the year 2014.

To overcome this obstacle, we carried out manual checking procedures to remove false-positive citations in which EMR country names were not associated with first/principal authors. Interestingly, the highest volumes of false-positive citations that were eliminated were from Saudi Arabia and Kuwait. This is an indication that, in many cases, researchers in these countries occur as secondary and not as primary authors in most published research. This phenomenon is less prominent in Egypt, Tunisia, Pakistan, and Morocco. In the Islamic Republic of Iran, only a minute fraction of false-positive citations were encountered, indicating the dominant role of Iranian researchers in biomedical research publications collected using the inclusion criteria of our study (Table 2).

Our study also suffers from important limitations and its findings should be interpreted in light of these. First, PubMed does not represent all scientific and biomedical journals published. This database consists largely of English-language journals, therefore possibly contributing to selection bias due to language barriers. With respect to the EMR, WHO has its own Index Medicus for the Eastern Mediterranean Region (IMEMR), which includes some 600 journals produced in the Region, some of which, but not all, are included in PubMed. Subsequent similar studies should include such important database (20,21).

Second, we limited the present research to reviews, original journal articles, and case reports published by principal investigators affiliated to institutions located in the EMR. Papers with main authors being from EMR institutions reflect a prominent role in the design or execution of the reported health research, provides a fair representation of health research directions in the region as reflected by research cited in PubMed, and ascertains that each citation is assigned to a unique country. This latter reason would better serve better our future analytical purposes and help avoiding overlapping results.

Additionally, the change in the NLM's policy from indexing the affiliation of the first author only and to include affiliations of all authors in every citation in the PubMed database for citations indexed after October 2013, was a major reason to limit our analysis to the period up to year

2013. This change required the implementation of manual checking procedures to remove false-positive citations with address affiliations referring to countries of the EMR, but not belonging to first-authors. In several instances, this also revealed the presence of false positives in citations published during the years 2012, 2011, and, to a lesser extent, during year 2010.

Despite the above-mentioned limitations, the comprehensiveness and breadth of PubMed databases are enough reasons for us to believe that the results presented in this study are a bona fide representation of the overall biomedical and health research outputs from the EMR. Additionally, the application of appropriate metrics on the extracted data would certainly help in the formulation of invaluable indicators. This could be used to assess various aspects of research systems in the EMR, and to assist decision-makers in designing policies to improve research activities and to align them with health priorities in the region.

The refined EMR search strategy allows the task of citation data collection to be automated and may be used as on-the-spot monitoring systems that are used for the forecasting of biomedical research activities in the region. With the combination of diverse and versatile bibliometric indicators, scholars in the field could analyze research systems and could advise policymakers on designing policies to improve research productivy, adjust research objectives according to the time- or location-specific health priority requirements, and consequently rationally use often scarce resources in most EMR nations, allocated to scientific research in general, health and biomedical research in specific.

Acknowledgments

Facilities of the Faculty of Public Health at Jinan University, Tripoli, Lebanon, were utilized.

Funding: The first author was financially supported by the World Health Organization Regional Office for the Eastern Mediterranean, Cairo, Egypt.

Competing interests: None declared.

References

1. Durieux V, Gevenois PA. Bibliometric indicators: quality measurements of scientific publication. Radiology. 2010 May;255(2):342–51. PMID:20413749
2. Prichard A. Statistical bibliography or bibliometrics. J Doc. 1969;25:348–9.

3. Wallin JA. Bibliometric methods: pitfalls and possibilities. Basic Clin Pharmacol Toxicol. 2005 Nov;97(5):261–75. PMID:16236137

4. Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2008 Jan;36(Database issue):D13–21. PMID:18045790

5. National Library of Medicine. MEDLINE, PubMed, and PMC (PubMed Central): How are they different? 2016 (https://www.nlm.nih.gov/pubs/factsheets/dif_med_pub.html, accessed 27 November 2016)

6. PubMed. PubMed – NCBI (https://www.ncbi.nlm.nih.gov/pubmed, accessed 5 February 2017).

7. Tadmouri GO, Bissar-Tadmouri N. Biomedical publications in an unstable region: the Arab world, 1988-2002. Lancet. 2003 Nov 22;362(9397):1766. PMID:14643139

8. Thompson DF. Geography of U.S. biomedical publications, 1990 to 1997. N Engl J Med. 1999 Mar 11;340(10):817–8. PMID:10075537

9. Hefler L, Tempfer C, Kainz C. Geography of biomedical publications in the European Union, 1990-98. Lancet. 1999 May 29;353(9167):1856. PMID:10359422

10. Uthman OA, Uthman MB. Geography of Africa biomedical publications: an analysis of 1996-2005 PubMed papers. Int J Health Geogr. 2007 10 10;6:46. PMID:17927837

11. Rashidian A, Jahanmehr N, Jabbour S, Zaidi S, Soleimani F, Bigdeli M. Bibliographic review of research publications on access to and use of medicines in low-income and middle-income countries in the Eastern Mediterranean Region: identifying the research gaps. BMJ Open 2013;3:10 e003332.

12. Tadmouri GO, Bissar-Tadmouri N. A major pitfall in the search strategy on PubMed. Saudi Med J. 2004 Jan;25(1):7–10. PMID:14758370

13. Tadmouri GO. Biomedical bibliometrics of a country with multiple identities: the case of Palestine. Ann AlQuds Med. 2006;2:63–8.

14. Tadmouri NB, Tadmouri GO. Bibliometric analyses of biomedical research outputs in Lebanon and the United Arab Emirates (1988-2007). Saudi Med J. 2009;30(1):130-9.

15. Tadmouri GO, Tadmouri NB. Biomedical research in the Kingdom of Saudi Arabia (1982-2000). Saudi Med J. 2002 Jan;23(1):20–4. PMID:11938358

16. National Library of Medicine. Changes Coming to Author Affiliations. NLM Tech Bull. 2013; (394):b4.

17. Cooper ID. Bibliometrics basics. J Med Libr Assoc. 2015 Oct;103(4):217–8. PMID:26512226

18. Agarwal A, Durairajanayagam D, Tatagari S, Esteves SC, Harlev A, Henkel R, et al. Bibliometrics: tracking research impact by selecting the appropriate metrics. Asian J Androl. 2016 Mar-Apr;18(2):296–309. PMID:26806079

19. Shaban SF, Abu-Zidan FM. A quantitative analysis of medical publications from Arab countries. Saudi Med J. 2003 Mar;24(3):294–6. PMID:12704508

20. Mandil A, Chaaya M, Saab D. Health status, epidemiological profile and prospects: Eastern Mediterranean Region. Int J Epidemiol. 2013 Apr;42(2):616–26. PMID:23505252

21. Saleh S, Alameddine M, Mourad Y, Natafgi N. Quality of care in primary health care settings in the Eastern Mediterranean region: a systematic review of the literature. Int J Qual Health Care. 2015 Apr;27(2):79–88. PMID:25574040

Thursday 11th of July 2024 08:15:57 AM