THEMES AND DEBATES

Mainstreaming social determinants of health (SDH) terms into the PubMed database would facilitate research into SDH

Yegnanarayana Iyer Saraswathy Sivan, PhD

Abstract

Research into social determinants of health (SDH) is a vital link that bridges the social sciences and public health sciences. The effective assessment of SDH requires scrutiny of specific social risk factors for diseases. A major handicap for SDH research is the lack of a strong and common citation platform for researchers from both the social sciences and public health sciences. Improving the research infrastructure with a built-in interface between these sciences is a prerequisite for facilitating evidence-based research. This includes unfettered access to searchable databases. A few databases offer an interface between the social and public health sciences; of these, PubMed is one of the few freely accessible databases with global representation. The SDH need to be mainstreamed within PubMed to support comprehensive, evidence-based assessment of social factors in health and disease. In addition, the search and retrieval strategies of the database should be improved to better extract knowledge in the broad domain of SDH. This paper suggests steps to strengthen the PubMed database by mainstream-

ing SDH terms into its current features, Automatic Term Mapping and Medical Subject Headings (MeSH), and the creation of an exclusive subset on SDH. Adding more SDH terms to the PubMed database and making them more accessible to researchers would facilitate research into SDH and the creation of an evidence base on SDH without much financial investment. This is the first time that such a proposal for mainstreaming SDH terms into the PubMed database is being offered. When implemented, its results are expected to be not only positive but also far-reaching across the disciplines involved in the study of SDH.

Background

Research into the social determinants of health (SDH) is a vital link that bridges the social sciences and public health sciences. Dandona, et al. found that, among 4,876 articles from India published in 2002 and indexed in the PubMed database, 216 were in public health.[1] Of these 216, only 38 (0.8% of all articles) fell into the social/environmental subcategory. This means that less than 1% of Indian articles indexed in PubMed are classified as involving "social sciences." This statistic illustrates the dismal state of the social science–public health interface, and is one indication of limited research infrastructure available for the effective scrutiny of SDH.

Improving the research infrastructure is a prerequisite for facilitating research into SDH. Access to searchable databases of published articles is an integral part of that infrastructure. The social science–public health interface is strengthened when researchers have equally strong and easily accessible

Yegnanarayana Iyer Saraswathy Sivan, PhD
Associate Professor in Social Research
Department of Community Medicine
PSG Institute of Medical Sciences & Research
Coimbatore, India

Email: yssivan@gmail.com

databases for the mutual exchange of research information on SDH.

Health sciences and social sciences have sophisticated databases of published articles. SDH researchers from a spectrum of disciplines can look for literature from health sciences databases (e.g., PubMed, IndMED, Bioline International) and social science databases (e.g., Sociological Abstracts). While prominent health sciences databases (most notably PubMed) are accessible free of charge from anywhere in the world, access to their social sciences counterpart, Sociological Abstracts, is subscription-based. Making the latter freely accessible is not currently feasible, as it is commercially marketed. PubMed, in contrast, is government-owned. One of the available solutions would be forming a consortium of funders, including governments, to make Sociological Abstracts freely accessible through a subsidy fund.

However, a more feasible and easily achievable solution is to strengthen the social science content of the PubMed database. This would not only facilitate effective scrutiny of SDH but also improve the social science–public health interface.

A better evidence base

The study of SDH requires scrutiny of specific social risk factors for diseases. Strengthening the often-used search databases by mainstreaming (i.e., routinely including) SDH terms and concepts into them would not only keep them comprehensive, but also keep them current, providing the evidence base for further research. Concepts could be contributed by all relevant disciplines for the scrutiny of SDH.

PubMed uses a controlled vocabulary called Medical Subject Headings (MeSH) for indexing articles. MeSH already incorporates some SDH terms.[2] Adding more and specific SDH terms would facilitate the understanding of specific underlying causes and, thereby, effective solutions.

Steps to mainstream SDH terms into the PubMed database

The existing features and algorithms used by PubMed offer a few opportunities for mainstreaming SDH terms into the PubMed database.

The Automatic Term Mapping (ATM) algorithm

The automatic term mapping (ATM) algorithm[3] is a feature used to search *PubMed*.[4] In a "phrase search," more than one term is typed between double quotes in the PubMed search box (ATM "off" mode). The advantage of a phrase search is that only the specific phrase exactly as it is typed in the search box is retrieved and other combinations or sequences of terms (as in the ATM "on" mode or PubMed's default search strategy) are not included in the search output. The number of citations retrieved will be smaller compared to those in a non-phrase search. However, in order to do a phrase search, the phrase needs to be indexed under the ATM algorithm.

Most phrases used in health sciences (e.g., "kidney allograft," "smoking cessation") are retrievable using the ATM "off" mode. However, commonly used SDH terms in the health research literature are not indexed for ATM "off" searches in PubMed. For example, an important phrase, "social determinants of health," was not indexed until recently when the author of this article took up the matter with the U.S. National Library of Medicine (NLM). Unless important SDH terms are indexed under the ATM algorithm, search output will be poor, especially when SDH terms are searched for in conjunction with other (biomedical) terms using Boolean operators ("AND," "OR," "NOT").

There are many such social science phrases representing social risk factors for diseases, or underlying social factors, etc., that may be brought under the ATM algorithm to enable them to be searched in conjunction with relevant disease conditions with the help of Boolean operators. One example would be a search for "tuberculosis" in conjunction with terms such as "social exclusion," "social capital," "female-headed household," "feminization of poverty," "double deprivation," "social stigma," "gender," "empowerment," "poverty," "socioeconomic status," "tradition," "attitude," "behavior," "social rehabilitation," "self-help group," and "social inclusion." Although "socioeconomic status" covers education, occupation, and income, these three subheadings should be listed individually in addition to the broader category. A search in conjunction with one

or more SDH terms contributes to a better evidence base and makes the scrutiny more holistic.

The Medical Subject Heading (MeSH) algorithm

The MeSH database,[2] a feature of PubMed, is a controlled vocabulary system with a more "intelligent" subject search facility compared to a basic key word search. While a basic "key word search" retrieves an abstract even with a casual citation of the term searched for in it (and even when filters such as for language and date are activated), a "subject search" using the MeSH algorithm provides the opportunity to select from more relevant medical subject heading(s) and sub-heading(s) and thereby helps in retrieving more subject-specific and relevant output from the database.

At present, some SDH terms are indexed in the MeSH database; many more need to be added. Furthermore, there are very few or no SDH terms indexed as sub-headings in MeSH. Including more specific SDH terms in the MeSH database would allow a more relevant and, therefore, more efficacious subject search for researchers studying the association between SDH and specific disease conditions. That in turn would contribute to strengthening the evidence base on SDH.

The MeSH database allows both horizontal and vertical searching. In a horizontal search, disease conditions (and concepts) and SDH concepts, as MeSH subject headings or sub-headings, could be chosen jointly from the MeSH database; the results would provide articles that mention both concepts. In a vertical search, the disease conditions (or concepts) would be selected from the MeSH database and the search would then retrieve SDH concepts associated with this condition (or vice versa); this typically provides more focused results. For example, for a vertical search under the MeSH subject heading "obesity," could bring up relevant SDH subheadings such as "socioeconomic status," "social class," "culture," "acculturation," etc.

Creating a subset on SDH

A suggested landmark innovation in the PubMed database for facilitating better scrutiny of SDH would be the creation of a "subset" on SDH. Ac-

cording to the NLM, "[s]ubsets allow searchers to limit their search to a particular portion of the PubMed database."[5] At present, the PubMed database has 10 subsets,[6] including HIV/AIDS, cancer, complementary medicine, and dietary supplements, among others. A subset on SDH would incorporate as many specific SDH terms as possible, drawn from all relevant disciplines contributing to the scrutiny of SDH, most notably the social sciences (and, among them, sociology). The subset should have broad categories, such as social determinants of communicable diseases, social determinants of non-communicable diseases, social determinants of violence and injuries, and the health system. Creating a subset on SDH is a challenging but achievable task. A systematic process of identifying social determinants for each of the conditions needs to be developed based on an SDH framework. The present author and colleagues have looked at the identification of core SDH elements for teaching health topics.[7] Similar discovery processes could help in identifying core SDH terms denoting social risk factors and allied concepts for inclusion in the SDH subset. The database would need to be updated continually to keep it current.

Additional avenues

A meta-search facility could be introduced in PubMed to allow users to retrieve information from other databases indexing social science publications related to health. [Formerly, the NLM Gateway could have been used for this purpose.][8] PubMed could also index more social science journals carrying health-related articles and index relevant meeting abstracts. MEDLINE could require that journals ask authors to supply both MeSH and non-MeSH SDH terms as key words.

Resources

PubMed already has a system in place to solicit comments and suggestions, e.g., suggestions for new MeSH terms or modifications to existing ones.[9] Through this routine channel and by calling for suggestions through an introductory article in relevant journals, NLM might attract SDH researchers to contribute to enriching the database with SDH

terms. The number of such terms is not going to be very high. Although funds may be required for quality checks before the SDH terms are accepted, engaging SDH researchers and experts through such a massive consultation exercise is likely to bring down the cost of adding the new terms.

Impact

The benefits accrued from the proposed initiative in terms of building the SDH evidence base in PubMed would seem to far exceed the investment. The creation of a better and stronger evidence base on SDH will support research and decision making. Research output will find its way into research protocols and papers, and subsequently into policy and program documents, thereby making research into SDH more effective. The impact of mainstreaming SDH terms into the PubMed database is expected to have a positive influence on other databases and on public health at large.

Conclusion

Social determinants play a significant role in both mortality and morbidity. Mainstreaming SDH terms into the PubMed database will contribute to the creation of a stronger evidence base for the scrutiny of SDH. By adding more SDH content systematically and philosophically (reflecting the biopsychosocial model), PubMed will not lose its identity as a "biomedical" database; instead, incorporation of SDH terms will result in the enrichment of the database's biomedical content and thereby increase its contribution to the betterment of health.

Acknowledgements

References

1. Dandona L, Sivan YS, Jyothi MN et al. The lack of public health research output from India. BMC Public Health. 2004;4:55.
2. MeSH [Internet]. Bethesda (MD): National Center for Biotechnology Information, U.S. National Library of Medicine; 2005 [cited 2014 Nov 13]. Searching PubMed: Searching for a phrase. Available from: http://www.ncbi.nlm.nih.gov/sites/entrez?db=mesh.
3. PubMed Help [Internet]. Bethesda (MD): National Center for Biotechnology Information, U.S. National Library of Medicine; 2005 [cited 2014 Nov 1]. Appendices: How PubMed works: automatic term mapping. Available from: http://www.ncbi.nlm.nih.gov/books/NBK3827/#pubmedhelp.How_PubMed_works_aut.
4. PubMed Help [Internet]. Bethesda (MD): National Center for Biotechnology Information, U.S. National Library of Medicine; 2005 [cited 2014 Nov 13]. Searching PubMed: Searching for a phrase. Available from: http://www.ncbi.nlm.nih.gov/books/NBK3827/#pubmedhelp.Searching_for_a_phra.
5. Nahin A. PubMed: Searching using subsets. NLM Technical Bulletin [Internet]. 1999 Jan-Feb [cited 2014 Nov 13]; 306. Available from: http://www.nlm.nih.gov/pubs/techbull/jf99/jf99_subset.html.
6. MEDLINE®/PubMed® Resources Guide [Internet]. Bethesda (MD): U.S. National Library of Medicine, National Institutes of Health; 2001 Jan 1 [updated 2014 May 27; cited 2014 Nov 13]. PubMed subject filters. Available from: http://www.nlm.nih.gov/bsd/pubmed_subsets.html.
7. Sivan YS, Chacko TV, Subramaniyan P. Identification of social determinants of health components for teaching: a pilot process. Med Educ. 2010;44:1137-8.
8. NLM Gateway [Internet]. Bethesda (MD): U.S. National Institutes of Health. [updated 2014 May 27; cited 2014 Nov 13]. Available from: http://gateway.nlm.nih.gov/gw/Cmd/about.jsp [Accessed: 2013 Jun 29].
9. Medical Subject Headings [Internet]. Bethesda (MD): U.S. National Library of Medicine, National Institutes of Health; 2001 Nov 20 [updated 2013 Oct 21; cited 2014 Nov 13]. Suggestions for medical subject headings change. Available from: https://www.nlm.nih.gov/mesh/meshsugg.html.