ELSEVIER

Journal of Clinical Epidemiology 63 (2010) 1205−1215

Journal of Clinical Epidemiology

ORIGINAL ARTICLES

Science mapping analysis characterizes 235 biases in biomedical research

David Chavalarias[a,b], John P.A. Ioannidis[c,d,*]

[a]Centre de Recherche en Épistémologie Appliquée, École Polytechnique - CNRS, 32 Bd Victor, 75015 Paris, France
[b]Institut des Systèmes Complexes de Paris Ile-de-France, 57−59 rue Lhomond, 75005, Paris, France
[c]Department of Hygiene and Epidemiology, University of Ioannina School of Medicine and Biomedical Research Institute, Foundation for Research and Technology-Hellas, Ioannina 45110, Greece
[d]Tufts Clinical and Translational Science Institute and Institute for Clinical Research and Health Policy Studies, Tufts Medical Center and Department of Medicine, Tufts University School of Medicine, Boston, MA 02111, USA

Accepted 22 December 2009

Abstract

Objective: Many different types of bias have been described. Some biases may tend to coexist or be associated with specific research settings, fields, and types of studies. We aimed to map systematically the terminology of bias across biomedical research.

Study Design and Setting: We used advanced text-mining and clustering techniques to evaluate 17,265,924 items from PubMed (1958−2008). We considered 235 bias terms and 103 other terms that appear commonly in articles dealing with bias.

Results: Forty bias terms were used in the title or abstract of more than 100 articles each. Pseudo-inclusion clustering identified 252 clusters of terms. The clusters were organized into macroscopic maps that cover a continuum of research fields. The resulting maps highlight which types of biases tend to co-occur and may need to be considered together and what biases are commonly encountered and discussed in specific fields. Most of the common bias terms have had continuous use over time since their introduction, and some (in particular *confounding*, *selection bias*, *response bias*, and *publication bias*) show increased usage through time.

Conclusion: This systematic mapping offers a dynamic classification of biases in biomedical investigation and related fields and can offer insights for the multifaceted aspects of bias.

Keywords: Bias; Mapping; Clustering; Directed clique; Text-mining; Biomedical literature

1. Introduction

"Bias" is a popular word in the research literature and beyond. Biases entail deviations that are beyond chance [1]. Understanding the sources of bias, its impact, and how it has been handled and hopefully avoided is important for grading evidence [2]. Over time, investigators have described a large number of different biases. New terms have been coined, cumulatively creating an extensive dictionary of bias nomenclature. Some biases are relevant to a wide spectrum of research designs, studies, and settings, whereas others are specific to special situations.

The wide diversity in this nomenclature makes categorization difficult. In a classic paper in 1979 [3], David Sackett cataloged different types of biases that may arise in clinical research in particular. However, many of the proposed terms were largely or completely abandoned, whereas many new terms were later introduced. It would be useful to create a systematic map of the different types of biases that exist in biomedical research at large. This map should address the overlapping terms and synonyms, to avoid redundancy. Moreover, one would wish to know whether some types of biases tend to coexist. Then, investigators who find or are apprehensive about the presence of one type of bias would be encouraged to consider whether other biases that tend to coexist may also be involved in the same study. Furthermore, it would be useful to map whether specific biases have been particularly invoked in particular domains. One could then routinely consider these potential biases when relevant studies are involved. Finally, one could evaluate whether

Both authors contributed to the conception of the project and the study design. D.C. developed the pseudo-inclusion clustering methods and performed all the clustering and mapping analyses. Both authors appraised and interpreted the results. J.P.A.I. wrote the first draft of the manuscript and D.C. critically revised it and contributed the methodological sections. This project started from the meeting of the authors at the Summer Workshop "Science and the Web" at the Mediterranean Institute for Life Sciences in Split in July 2008 to which both authors had been invited faculty.

* Corresponding author. Tel: + 30-265-1007807; fax: + 30-265-1007867.
E-mail address: jioannid@cc.uoi.gr (J.P.A. Ioannidis).

0895-4356/$ - see front matter
doi: 10.1016/j.jclinepi.2009.12.011

particular types of biases have acquired increasing prominence in the literature in recent years.

Here, we have used an intensive computational text-mining approach to create a clustering map of bias issues across the whole biomedical literature. The map offers a wide view of the different types of bias in biomedical research, their overlap, co-occurrence, and specificity; the terms they are associated with; and how they have evolved over time.

2. Materials and methods

2.1. Selected terms

We aimed to generate a systematic, comprehensive list of terms that include the word ‘‘bias’’ and one or two other words preceding ‘‘bias.’’ We did not consider terms that may signify biases but do not contain the word ‘‘bias’’ per se, because their identification would be subjective and potentially erratic. We also aimed to identify terms that are frequently encountered in articles that deal with bias.

We screened electronically the PubMed database from 1958 to 2008 (17,265,924 references) for articles that included the word ‘‘bias’’ in their title, so as to focus on an enriched sample of papers where bias was a key consideration. The titles of the 6,405 retrieved abstracts were electronically text-mined to identify two-grams (two words in sequence) with ‘‘bias’’ as last word that appeared in at least three titles.

Moreover, the abstracts of the 6,405 retrieved papers were also text mined to identify all n-grams with $n \leq 3$ appearing in at least $100^{(1/n)}$ abstracts. Thus, we retained all words that appeared at least 100 times; all two-grams that appeared at least 10 times; and all three-grams that appeared at least 4 times. We then cleaned the terms with standard linguistic treatment. Specifically, we removed all n-grams ending with ‘‘ly,’’ ‘‘ful,’’ ‘‘ary,’’ ‘‘ory,’’ ‘‘al,’’ ‘‘able,’’ ‘‘ed,’’ or ‘‘ic’’; uninformative stop words such as ‘‘and,’’ ‘‘of,’’ ‘‘the,’’ and so forth; all two-grams and three-grams including the word ‘‘bias’’ as a first word (e.g., ‘‘bias correction’’ or ‘‘bias is significant’’); all three-grams including the word ‘‘bias’’ as middle word (e.g., ‘‘large bias is’’); and two-grams or three-grams where ‘‘bias’’ was the last word, but the preceding words were only a verb, an adjective, or adverb and that did not characterize a specific type of bias (e.g., ‘‘reducing bias,’’ ‘‘large bias,’’ ‘‘significantly bias’’). Moreover, we merged synonyms with different spelling (e.g., ‘‘meta-analysis’’ and ‘‘metaanalysis,’’ terms with English versus American spelling, and singular and plural of same word). Eventually, the final list of terms, thereafter referred as L, had 338 entries, of which 235 were bias terms. We counted the number of occurrences and pair-wise co-occurrences for all 338 entries across the entire PubMed (titles and abstracts).

2.2. Mapping: general principles

The general principle of our analysis is to use automated methods to extract meaningful sets of terms related to each other and occurring together in the literature of specific themes. We define a proximity measure between terms that reflects the way terms are associated with similar themes and extract coherent sets of terms delimiting a domain. These sets of terms (clusters) are then displayed as connected nodes of a graph, where connections indicate the proximity between domains of investigation. The resulting maps highlight which types of biases tend to co-occur and may be considered together and along with what other terms.

2.3. Pseudo-inclusion measure

We choose here a variant of the paradigmatic proximity P_α proposed by Chavalarias and Cointet [4], thereafter called *pseudo-inclusion measure*. This measure has the advantage to convey information about the degree of specificity of a term: given two terms i and j, it conveys whether one is more specific than the other, i.e., tends to be used by a subcommunity of the community using the other. See the Appendix for the exact formula to calculate the pseudo-inclusion. P_α is asymmetric and α is a focus parameter that determines the direction and the strength of the asymmetry. For small values (<1) of α, the higher $P_\alpha(i,j)$ the more specific is j relative to i. Asymmetry means that $P_\alpha(i,j)=P_{1/\alpha}(j,i)$ so if j is specific to i, i is generic to j. For example, ‘‘citation bias’’ is specific to the term ‘‘publication,’’ whereas ‘‘publication’’ is generic to ‘‘citation bias.’’ Two terms that are specific to each other are used roughly by the same communities, e.g., ‘‘emotional bias’’ and ‘‘affective bias.’’ Two terms in which both $P_\alpha(i,j)$ and $P_\alpha(j,i)$ are low are considered as irrelevant to each other, e.g., ‘‘accessibility bias’’ and ‘‘myocardial infarction.’’ In the presented analyses we used $\alpha = 0.1$.

2.4. Clustering

$P_{0.1}$ transforms the co-occurrence matrix into a proximity matrix. This matrix defines a directed weighted graph G on the set L that can be further analyzed with clustering methods. In our context, clusters should represent domains of investigation defined by sets of strongly related terms that contextualize each other, some being more specific and others more generic. This notion fits well with the concept of cliques from graph theory. A clique is a subgraph of G such that all vertices are linked to all others in the subgraph. Because our proximity measure is asymmetric, we choose to define clusters as directed cliques [5]; this makes it possible to take into account the directionality of edges. We also pruned the graph G, keeping for each node only its 20 strongest connections. This operation makes it possible both to render the clique detection algorithms computationally tractable on large networks and avoid very common (and less informative) terms linked to many others to be over represented.