The knowledge base evolution: the case of lifestyle research in medicine

by Elena Cefis | Nicola Cornali | Matteo Tubiana | Università degli studi di Bergamo | Università degli studi di Bergamo | Università degli studi di Bergamo

Abstract ID: 54
Inviato: 06/05/2021
Evento: XIX Workshop Annuale SIEPI
Argomento: XIX Workshop Annuale SIEPI
Parole chiave: Lifestyle research, health behaviour, knowledge base, knowledge recombination, statistical methods, text mining

Science is the basement of innovative activities and economic growth. It evolves reacting to how scientists interact and establish new habits and norms. In this article, we investigate how the statistical knowledge base exploited by scientists to reach empirically sound conclusions evolves in time and space. The analysis focuses on lifestyle research in medicine, an increasingly central topic in medical research and health management and policy, given the determinant role misbehaviours have on health-related issues in developed countries. We set up an algorithm mining the text of a large sample of PubMed Central research articles, looking for expressions of statistical methodologies. Embracing the tradition of the economics of knowledge, we investigate how the set of statistical methods used in lifestyle research evolve and how these methods recombine and establish the field's statistical knowledge base. Furthermore, we narrow the investigation on a novel, rapidly expanding class: *Computer intensive methods*. Are they percolating lifestyle studies? How do they recombine with more established methods? Do they define a methodological niche? And, do they generate a "publication premium"? We find that the statistical knowledge base rests upon a few, very diffused methods (*Descriptive statistics*, *Hypothesis tests* and *Epidemiology 2*). Most articles recombine three to four methods, but some methods emerge only combined with others (complementarity in use). The distribution of methods is geographically concentrated, even though the use of statistics is increasingly popular. *Computer intensive methods* appear as widely applicable: they do not cluster in a niche, instead recombine with many other classes without apparent privileges. Articles that employ them experience, on average, a *publication premium*, i.e. a publication in a higher than average Impact Factor journals. The same happens for *Survival analysis*, *Epidemiology 1* and *Missing data*.