The proportion of cancer-related entries in PubMed has increased considerably; is cancer truly “The Emperor of All Maladies”?

Constantino Carlos Reyes-Aldasoro

School of Mathematics, Computer Science and Engineering, City, University of London, London, United Kingdom

Background

This work explored the presence of Cancer-related publications in PubMed. The database MEDLINE of the United States National Library of Medicine (NLM) and its search engine PubMed (https://www.ncbi.nlm.nih.gov/pubmed) have grown to include over 26 million entries out of which more than 3 million entries correspond to Cancer, which correspond roughly to 12% of the total entries.

Methods

The public database of biomedical literature PubMed was mined systematically using queries with combinations of keywords: Cancer-related, organ, funding and year restrictions. In addition, the relationships with DNA, Computing and Mathematics, were performed to explore the impact of these scientific advances on Cancer Research. All queries and figures were generated with the software platform Matlab® and the files are freely available.

Results

The proportion of Cancer-related entries per year in PubMed has risen from around 6% in 1950 to more than 16% in 2016. This increase is not shared by other conditions such as AIDS, Malaria, Tuberculosis, Diabetes, Cardiovascular, Stroke and Infection some of which have, on the contrary, decreased as a proportion of the total entries per year. Interestingly, the proportion of Cancer-related entries that contain “DNA”, “Computational” or “Mathematical” have increased, which suggests that the impact of these scientific advances on Cancer has been stronger than in other conditions.

Conclusions

The sharp increase of Cancer Research as testified by the number of entries in PubMed may be due to the strong impact of the scientific advances in the areas of Genetics, Computing and Mathematics, which have had a stronger influence in Cancer than other areas like cardiovascular disease. It is important to highlight that the results obtained with a data mining approach and thus are limited to the presence or absence of the keywords on a single, yet extensive, database.

Reference:

Reyes-Aldasoro CC (2017) The proportion of cancer-related entries in PubMed has increased considerably; is cancer truly “The Emperor of All Maladies”? *PLOS ONE* 12(3): e0173671. https://doi.org/10.1371/journal.pone.0173671