

Analyzing Funding Disparities in Cancer Research

A DS 4002 Case Study by Maansi Taori

Cancer research remains a cornerstone of medical innovation and investment, with substantial resources allocated to advancing our understanding, prevention, and treatment of this pervasive disease. In 2022 alone, the National Cancer Institute (NCI) managed a budget exceeding \$6.8 billion, underscoring the critical priority placed on combating cancer. By 2024, the NCI awarded more than 1,600 new grants dedicated to cancer-specific research, funding a diverse array of projects spanning basic science, clinical trials, and cutting-edge experimental therapies.

To ensure transparency and accountability within the scientific community, the National Institutes of Health (NIH) meticulously collects and shares data on funded projects across its affiliated organizations. This open data promotes collaboration, enhances public trust, and informs stakeholders about ongoing advancements in cancer research. For researchers, understanding the factors that contribute to successful grant applications is essential. Insights into grant selection criteria and trends can improve the quality of proposals, align research objectives with funding priorities, and ultimately drive innovation in cancer treatment and care. This data not only serves as a tool for navigating the competitive grant application process but also reflects the evolving focus and direction of cancer research at large.

You are a data scientist tasked with analyzing funding data in order to uncover trends and disparities in research support for different cancer types and treatment methods. By examining currently approved projects, you can identify relationships between cancer types, methodologies, and the total funding/support that is received. Furthermore, you are tasked with analyzing whether funding disparities exist based on the methodologies used in these research projects, such as basic science versus clinical research, or traditional treatments versus innovative approaches, such as immunotherapy, to help potentially guide funding agencies to distribute resources more equitably. You will be applying a keyword analysis model and topic modeling on submitted abstracts awarded grant funding to group the research abstracts into different topics that represent common research methodologies and subsequently quantify the frequency of cancer types and methodologies in each abstract in the dataset. After, a statistical analysis will be conducted to determine if the observed differences are statistically significant.

“Home | RePORT,” *report.nih.gov*. <https://report.nih.gov/>

A. Schlafly and R. Sebro, “Does NIH funding differ between medical specialties? A longitudinal analysis of NIH grant data by specialty and type of grant, 2011–2020,” *BMJ Open*, vol. 12, no. 12, pp. e058191–e058191, Dec. 2022, doi: <https://doi.org/10.1136/bmjopen-2021-058191>.