

Evaluating the Impact of NCI Funding on Diversity and Innovation of Principal Investigators in HPV-Related Cancer Research: 2015 - 2021

Siqi Wang

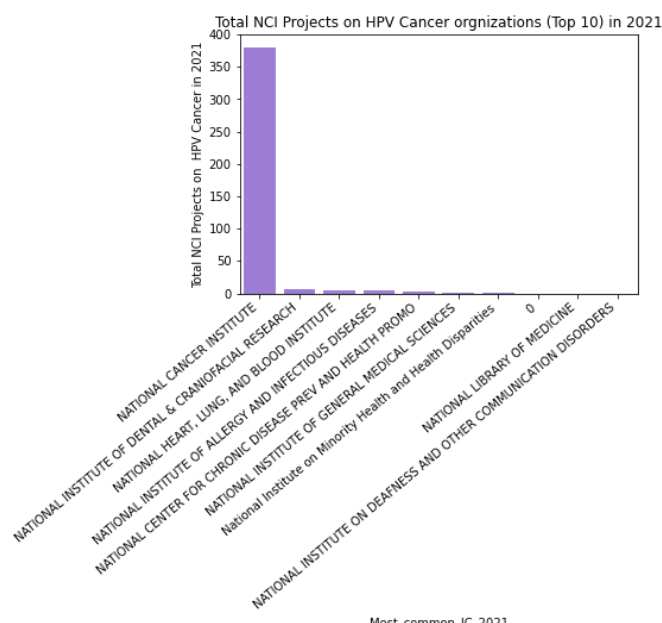
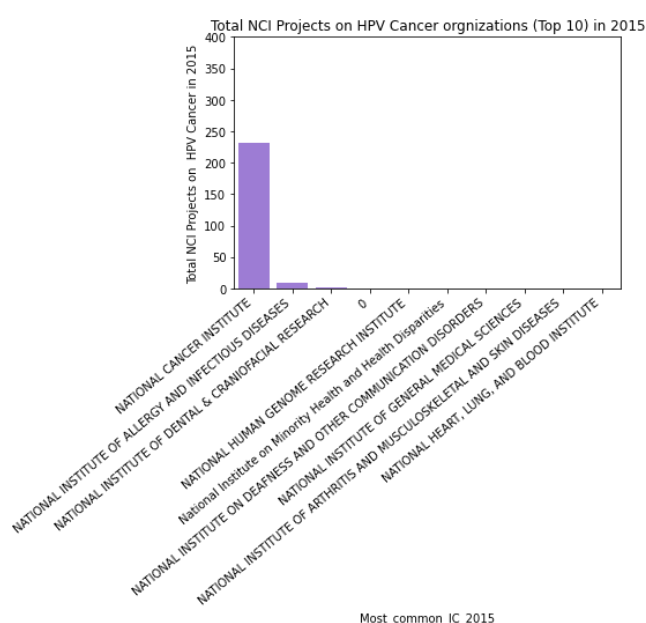
INTRODUCTION

This study explores the impact of research funding on the behaviors and diversity of Principal Investigators (PIs) in the field of Human Papillomavirus-related (HPV-related) cancer research, a significant area of study given the global prevalence of HPV-related cancers. Specifically, it examines how funding from the National Cancer Institute (NCI), strategically enhanced by the Cancer Moonshot initiative to foster innovative cancer research, contributes to attracting new and more diverse PIs. This study aims to highlight the wider impact of targeted financial support on researcher diversity and investigative approaches within the scientific community, particularly in HPV-related cancer research, through an equity lens.

MOTIVATION

HPV is one of the most common sexually transmitted infections and a leading cause of cervical cancer, which is a significant contributor to mortality among women (Office of the Assistant Secretary for Health, 2022). Worldwide, high-risk HPVs cause about 5% of all cancers. In the United States, high-risk HPVs cause 3% of all cancers in women and 2% of all cancers in men (NIH, 2023).

In the United States, the National Institutes of Health (NIH) provides the majority of federal funding for research on HPV-related cancer. From 2012 to 2022, NIH spent over 448 million on HPV and/or Cervical Cancer Vaccines, and in 2022, it allocated 61 million on it, increasing by 32.6% from the previous year 2021(NIH, 2023). NCI division has received the largest amount of funds for HPV-related cancer research. As the bar charts show, the NCI has the highest number of HPV-related cancer research projects among all NIH organizations.



The Cancer Moonshot initiative, launched in 2016, aims to accelerate scientific discovery in cancer research, foster greater collaboration, and improve the sharing of cancer data. This initiative has likely played a pivotal role in advancing HPV-related research. Research from Ohio State University Comprehensive Cancer Center (OSUCCC) has demonstrated significant growth in several key areas due to increased funding. This includes the recruitment of new faculty members, a rise in direct cancer-related funding, a higher number of published journals, and enhanced collaboration (NIH, 2021). However, there remains a notable gap in understanding how such increased funding specifically impacts PIs conducting HPV-related cancer research. Particularly, the effects on the amount and diversity of PIs in this field are yet to be explored.

Diversity of PIs

Diversity at all levels from the kinds of science to the regions in which it is conducted to the backgrounds of the people conducting it contributes to excellence in research training environments and strengthens the research enterprise.

— NCI, 2023

METHODOLOGY

LOGIC MODEL

The Logic Model outlines the inputs, activities, outputs, and outcomes, providing a comprehensive view of the study's scope and focus. The bolded texts correspond to this study's research questions, which are detailed in the following section.

Research Design and Assumption: The foundational assumption of this study is that the increase in funding for HPV-related cancer research can be attributed exclusively to the Cancer Moonshot initiative. Ideally, a difference-in-differences research design would be employed to compare the actual scenario with a hypothetical situation where the Cancer Moonshot initiative did not exist. However, due to time and resource constraints, this study will focus solely on the intervention group, representing the reality in which the Cancer Moonshot initiative did occur.

Timeframe: Financial Year 2015 is selected as the baseline year for the study, as it is the closest year before the Cancer Moonshot, capturing the research of interest before the increased funding was introduced. Fiscal Year 2021 is chosen as the comparison year, being the most recent year available following the start of the initiative.

Inputs	Activities	Outputs	Outcomes
NCI Research Funding	Recruitment of PIs	Increased number of PIs	New publications on HPV-related cancer research
Access to HPV-related cancer research data and research facilities	Conducting HPV-related cancer research	Enhanced diversity of PIs <ul style="list-style-type: none"> • Gender • Race and ethnicity • Affiliated organization • Geographic distribution (state level) 	New patents for HPV-related cancer research
Collaboration with medical institutes, universities, and other research centers	Facilitating meetings and coordination among researchers	Development of new research topics <ul style="list-style-type: none"> • Keyword frequency 	More cooperation on HPV-related cancer research
Facilities and equipments	Implementing strategies to attract PIs from underrepresented groups	...	Research represents a more diverse range of beneficiaries in HPV-related cancer research
...

RESEARCH QUESTION

The research question, "Does the increased NCI funding from 2015 to 2021, attributed to the Cancer Moonshot initiative, correlate with the emergence of new and diverse PIs conducting HPV-cancer research?" seeks to assess whether the Cancer Moonshot initiative has promoted a more varied and inclusive research community in this domain. Specifically, the study addresses the following three sub-research questions:

Q1: Are there any new PIs entering this field, meaning, are there PIs who did not have a new project in 2015 but started one in 2021?

Q2: What is the geographical distribution of the new PIs' affiliated organizations in 2021, and how does it compare to the distribution of PIs in 2015?

Q3: For PIs who initiated new research projects in 2021 compared to those in 2015, are there any notable changes in the research topics?

DATA

The datasets for this study were sourced from the NIH Reporter's project and abstract data for the financial years 2015 and 2021. The project data comprises research identifiers and numerical statistics, such as the number of PIs and their projects. The abstract data includes research identifiers and research abstracts, which provide insights into the research domains. As the datasets do not contain demographic information about the PIs, the geographic location of the PI's organization is utilized to analyze changes in the origins of the PIs.

Record Linkage

Considering that multiple PIs can work on a single project, the PI names were initially separated to create a list of unique PI names. This cleaned list was then merged with the original data to develop measures at the PI level.

The study identified the total number of PIs and those specifically working on HPV-related cancer research projects for each year. By linking project data with PI information, it was possible to track the emergence of new PIs who did not have projects in 2015 but began working on projects by 2021.

Text Analysis

In this research, any project abstract containing the terms “ ‘HPV’ or ‘Human Papillomavirus’ and ‘Cancer’ ” was categorized as related to HPV-cancer research. Missing data were excluded as imputing textual data was not feasible.

Text analysis techniques were used to analyze abstract datasets from 2015 and 2021. This process involved tokenizing the abstracts to create a corpus “bag of words”, which is a list of cleaned words (without punctuation) along with the frequency of each word and its features. Stop words were removed using Natural Language Toolkit (NLTK) and a custom list, which included terms such as ‘HPV’, ‘Human Papillomavirus’, ‘Cancer’, and ‘research’ to concentrate on significant and distinctive topics. The Latent Dirichlet Allocation (LDA) model was applied to group the abstracts into rough categories, revealing the top 8 keywords for each topic. Additionally, the Term Frequency-Inverse Document Frequency (TF-IDF) method was used to re-weight the words, allowing a focus on words unique to a document rather than those commonly occurring across the corpus. Lastly, the weighted topics for each project were merged with the abstract data and arranged in descending order by the total cost of research projects. This helped identify the topics that received the most funding in both 2015 and 2021.

FINDINGS AND DISCUSSION

Q1: Are there any new PIs entering this field, in other words, are there PIs who did not have a new project in 2015 but started one in 2021?

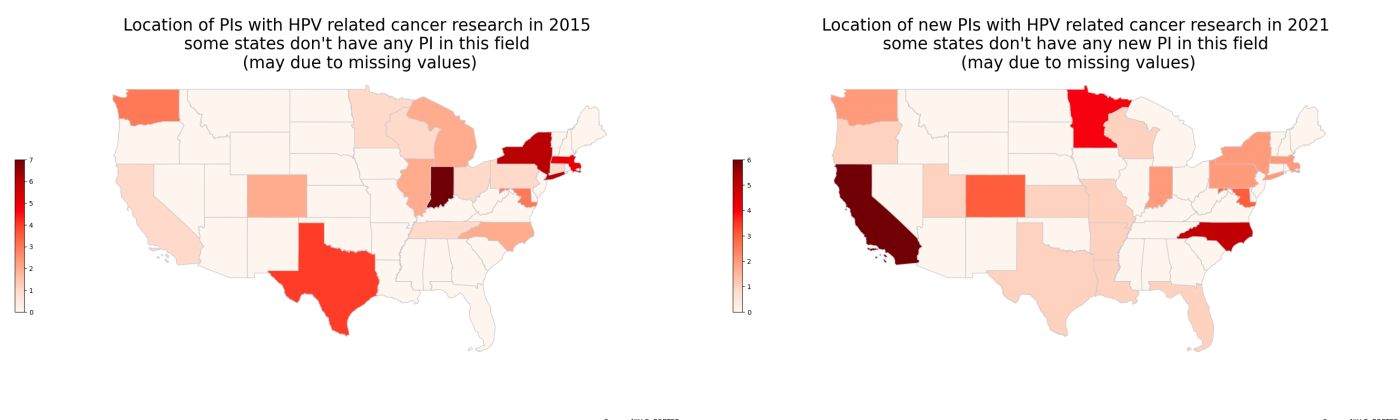
The study showed that 49 PIs who had a new HPV research project in 2021 did not have one in 2015. This indicates an increase in the number of researchers conducting HPV research for HPV-related projects after the launch of the Cancer Moonshot.

The entry of new PIs into the field of HPV research potentially introduces diverse research perspectives. However, it is vital to assess whether these new researchers represent a variety of backgrounds, such as different genders, ethnicities, geographical origins, and academic disciplines. This diversity is not only reflective of equitable resource allocation but also indicative of the accessibility of research opportunities. Questions 2 and 3 in our study specifically address the geographic distribution of these researchers and the topics they are focusing on.

Q2: What is the geographical distribution of the new PIs' affiliated organizations in 2021, and how does it compare to the distribution of PIs in 2015?

The analysis revealed a significant divergence in the geographical distribution of organizations associated with new PIs in HPV-related cancer research. The data showed an expansion in the distribution of these organizations across a broader range of states. According to the heatmaps, in 2021, California led in the number of PIs starting HPV-related cancer research projects, followed by North Carolina and Minnesota.

In contrast, Texas experienced a decline in the number of new research projects by PIs compared to 2015, but an uptick was noted in the neighboring states of Texas. Additionally, the Northeast and Great Lakes regions saw a decrease in the number of PIs starting projects in 2021, while an increase was observed in the Southeast regions. Furthermore, certain states in the Rocky Mountain region, Midwest, and South maintained consistently low numbers of new PIs in both 2015 and 2021.



Therefore, when viewed through a health equity lens, it becomes evident that while the Cancer Moonshot initiative has diversified the range of organizations where PIs conduct HPV-cancer research, there remains a significant need for the NCI to address disparities in PI accessibility in certain regions. This is crucial to ensure equitable opportunities for emerging PIs across different geographic areas and to avoid the loss of potential new contributors in regions where representation is consistently low. By fostering inclusivity and diversity in these areas, the NCI can contribute to a more comprehensive and equitable approach to cancer

research, ultimately leading to better health outcomes and advancements in cancer care that benefit all segments of the population.

To achieve these goals, one potential way is to enact and promote research projects that encourage PIs from underrepresented communities to apply for funding, especially because of the informational gap enlarged by socioeconomic factors that have already been borne. Another recommendation is to keep track of the diversity of the organizations and demographic profiles of PIs, so these datasets can better inform future research funding.

Q3: For PIs who initiated new research projects in 2021 compared to those in 2015, are there any notable changes in the research topics?

Analyzing the funding allocation for HPV-related cancer research projects in 2015 and 2021 provides insightful trends in research focus and priorities.

In 2015, the top-funded project predominantly centered around topics like "cervical, women, intervention, test, screening, sexual, current, clinic," with the highest funding of \$3,473,975, emphasizing screening, prevention, and clinical approaches. Other significant topics included treatment efficacy, cell and antigen research, DNA sequencing, and HPV risk factors.

In 2015, PIs received the most funding for projects on the following topics.		
Top	Cost	Topic
1	3,473,975	cervic, women, intervent, test, screen, sexual, current, clinic
2	3,435,389	treatment, group, 16, specif, efficaci, e6, tumor, effect
3	1,544,914	via, antigen, cell, pap smear, adapt, malign, smear, agent
4	1,165,850	dna, sequenc, genom, somat, associ, bacteri, integr, character
5	803,364	dna, cervic, invas, high risk, current, risk infect, progress cervic, risk type

In 2021, the topics of the highest funded project, receiving \$7,142,707, emphasized "cell, radiation, tumor, DNA," indicating a focus on molecular and oncological aspects of HPV-related cancers. This is followed by projects focusing on healthcare systems and communication, immunology and tumor research, and screening in low and middle-income countries (LMICs). The increase in funding for molecular and oncological studies could be attributed to advancements in technology and a deeper understanding of cancer biology. This progress potentially paves the way for developing more effective treatments and personalized medicine approaches.

In 2021, PIs received the most funding for projects on the following topics.

Top	Cost	Topic
1	7,142,707	cell, radiat, tumor, dna, dr, cervic, caus, oncolo
2	4,436,020	aat, communic, healthcar system, primari care, vaccin communic, vaccin, care team, healthcar
3	3,982,916	cell, immun, tumor, rt, model, hnscc, protein, inhibit
4	2,653,679	cervic, screen, women, cervic screen, hrhpv, lmhc, low, countri
5	2,292,610	vaccin, aim, multilevel, outcom, tailor, level, 3, structur

There are some insights from the 2021 research topic concerning equity issues.

First, the increased focus in 2021 on screening in low- and middle-income countries (LMICs) represents a step towards addressing global health inequities. HPV-related cancers disproportionately affect women in LMICs due to limited access to screening and vaccination. Future research can evaluate the impact of these projects on progress toward accessible and affordable interventions for low- and middle-income regions. Considering the varying levels of healthcare infrastructure and literacy across different populations, future research can also explore strategies to improve implementation in underserved communities.

Second, the emphasis on vaccination in 2021 research highlights the critical role of vaccines in preventing HPV-related cancers. However, significant disparities in vaccine uptake, often related to socioeconomic status, cultural beliefs, and misinformation, persist. Future research should address these disparities and develop targeted strategies to increase vaccination rates, particularly in communities with low uptake.

Finally, organizations such as the NCI and NIH should continue to implement, assess, and enhance efforts to diversify the scientific workforce in cancer research. Examples include the Exploratory Grant Award to Promote Workforce Diversity in Basic Cancer Research (R21 Clinical Trial Not Allowed), the NCI Cancer Moonshot Scholars Diversity Program (CMSDP), and Training and Support for Early-Career Scientists.

CAVEATS AND FUTURE RESEARCH

Data Source Bias: The data sourced from NIH Reporter may have a bias towards more leading research organizations, potentially skewing the results. Additionally, the absence of demographic information on PIs in the datasets limits the study's ability to thoroughly analyze aspects of equity among researchers. Future studies may consider integrating data from a wider range of sources to obtain a more representative sample and include demographic profiles of PIs to provide a more comprehensive understanding of equity in research funding and participation.

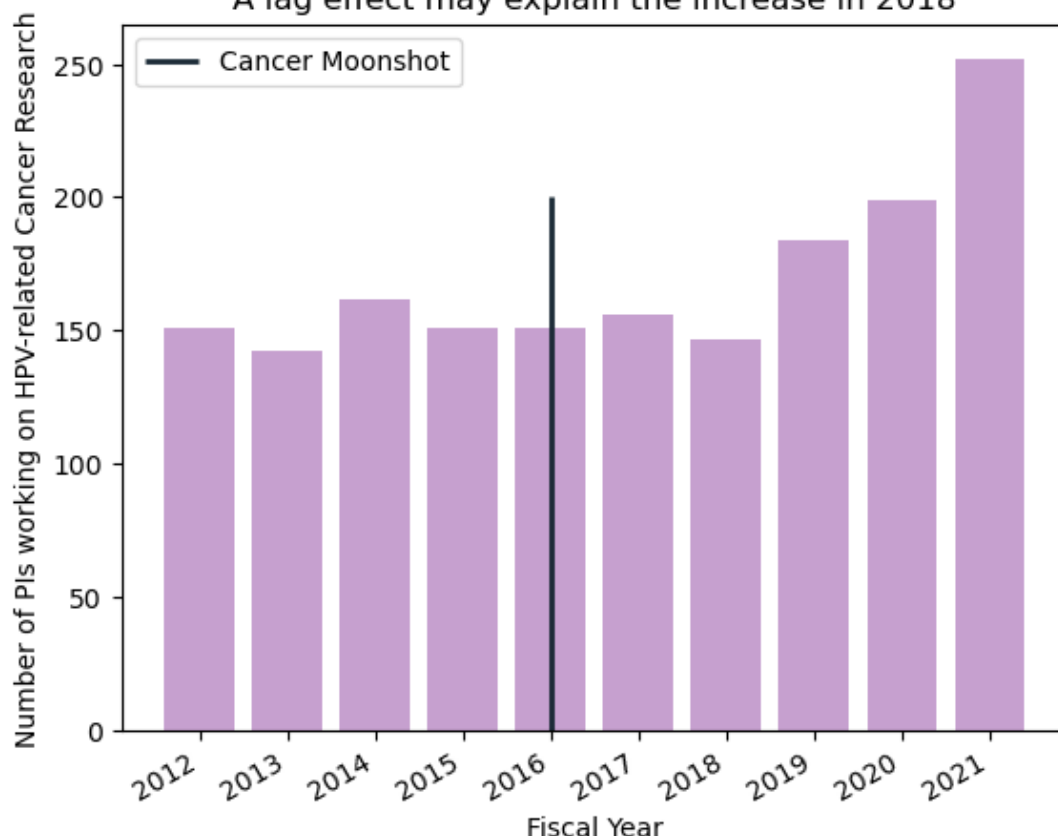
Linkage Bias: The linkage to PIs' unique names, rather than IDs, might have affected accuracy. Additionally, categorizing HPV-related cancer research solely based on keywords in project abstracts could have led to the inclusion of irrelevant projects or the exclusion of relevant ones. Future research should aim for a more robust model of identifying relevant research projects.

Observation of Lag Effect: The data showed a consistent number of PIs working on HPV research in 2018, followed by an increase, which might indicate a lag effect from the Cancer Moonshot initiative.

This lag effect observation warrants further investigation into the timing and impact of funding initiatives.

Research Assumption Limitation: The assumption that increased funding is directly attributable to the Cancer Moonshot initiative was not empirically tested for causality. Future research could focus on refining text analysis techniques and establishing a more concrete causal link between funding and research changes.

The Number of NCI Funded HPV Projects remained consistent until 2018 then increased
A lag effect may explain the increase in 2018



REFERENCES

- i Office on Women's Health. (n.d.). Human papillomavirus. OASH Office on Women's Health. Retrieved May 9, 2023, from <https://www.womenshealth.gov/a-z-topics/human-papillomavirus>
- ii National Cancer Institute. (n.d.). HPV and Cancer. Retrieved May 9, 2023, from <https://www.cancer.gov/about-cancer/causes-prevention/risk/infectious-agents/hpv-and-cancer>
- iii National Institutes of Health. (n.d.). Estimates of Funding for Various Research, Condition, and Disease Categories (RCDC). Retrieved May 9, 2023, from <https://report.nih.gov/funding/categorical-spending#/>
- iv Duncan, K., Abudu, R., & Klein, A. (2018, October 19). Using world report to measure how much HPV-related research is being funded throughout the world. National Cancer Institute. Retrieved May 9, 2023, from http://hpvroundtable.org/wp-content/uploads/2018/06/Cancer-Center-HPVConsensusStatement_FINAL_06.01.2018.pdf