Peer-Review comments and author responses

Reviewer 1

1. Methods:

-**Comment:** How do you define low, middle and high income?

Response*: Income categorization: The categorization of income levels (low, middle, high) was based on the family income-to-poverty ratio as provided in the NHANES dataset documentation, however we noticed these categories were not consistent with the data contained in this specific dataset given that there were binary categories originally mixed in the same categorization of the variable. Given this, we are grateful for your input, and have decided to change the variable to the Income Ratio to Poverty, which refers to a measure of the economic well-being of individuals in relation to the federal poverty threshold. With this, we can have a more accurate and comparable assessment of economic status across participants. Using the ratio as a continuous variable avoids the loss of information that can occur with categorization and provides a better visualization of the relationship between socioeconomic status and the outcome of interest, by reflecting how income relative to poverty influences the likelihood of cancer diagnosis. We believe this approach contributes to the robustness of the analysis by maintaining the variability in income data and reducing potential biases associated with arbitrary income thresholds.*

-**Comment:** I am missing the reason for the cut for the alcohol consumption.

Response*: Alcohol consumption cutoff: The cutoff of ≤2 and >2 drinks/day was chosen based on established public health guidelines according to the NIAAA, defining moderate versus heavy alcohol consumption. This rationale, along with an appropriate reference, has been included in the revised manuscript.*

-**Comment:** You refer to "Two models were developed: a basic model that included only exercise status (Figure 1)", but this Figure does not display the model; it's the STATA output of Table 1, just showing the descriptive statistics. It is, therefore, repetitive to have the same information in a figure and in a table.

Response*: Figure/Table overlap: Figure 1, which previously replicated the content of Table 1, has been removed to avoid redundancy. A new Figure 1 was generated, to represent our crude model.*

2. Results:

-**Comment:** It is mentioned that 39% of those included exercise at least more than 300 min of exercise. However, you never introduced these variables anywhere.

Response*: Exercise variable: We have clarified where and how the variables were introduced and incorporated them explicitly in the methods and results sections to enhance transparency.*

-**Comment:** You could have referred to Table 2 in the results section for the values you displayed there.

***Response**: Table 2: Cross-references to Table 2 have been added in the results section to guide readers to relevant data.*

-**Comment:** You refer to a sensitivity analysis, but all I read is not much new compared to the two models. Moreover, you said you did one for age, but it is included in the adjusted model. Same for BMI. The only new one includes alcohol consumption (figure 3!)

***Response**: Sensitivity analyses: We have clarified the distinction between sensitivity analyses and the adjusted model, emphasizing the unique findings of each. For example, alcohol consumption, included as a covariate, has been highlighted as a potential confounder impacting the association between exercise and cancer.*

-**Comment:** The last paragraph is already a discussion, not a simple stating facts results part.

***Response**: Discussion in the results: The last paragraph of the results section has been deleted, and the information has been incorporated in the discussion to maintain a clear separation between findings and their interpretation.*

3. Discussion:

-**Comment:** Pay attention. It's pretty repetitive; you don't have to state the whole results again, just the main message of the different parts.

***Response:** Repetitiveness: We have streamlined the discussion to focus on the main findings without repeating the results.*

-**Comment:** You stated, "The evidence suggests that regular exercise can help to reduce the risk of several types of cancer, including breast, colon, and prostate cancers." Where is the evidence for these different cancer types? You just analyzed cancer as a total.

***Response** Specific cancers: As our analysis aggregates all cancer types, we revised the discussion to reflect this limitation. References to studies analyzing specific cancer types were added for context, but we clarified that our study focuses on overall cancer risk.*

-**Comment:** The following sentence: "It's thought that this is down to things like reducing inflammation, improving immune function, and regulating hormones" is nice to know, as you stated in the introduction. But where is the link to your results that you want to discuss here?

Response: *Mechanisms: The discussion now explicitly connects our findings to the existing literature on the mechanisms through which physical activity influences cancer risk, with relevant citations.*

-**Comment:** The second last sentence is much too long to follow what you really want to mention here.

Response*: Simplification: Sentences have been restructured for clarity and readability, including the overly long sentence noted.*

-**Comment:** In general I miss in this part the real discussion where you set your findings into context of known results. Moreover, there are already analyses like yours in different NHANES datasets; why not use them for context? Like the following:
Liu W, An J, Jiao C, Zhi L, Guo J, Sun L. The Association Between Physical Activity and risk for breast cancer in US female adults: A cross- sectional study based on NHANES 2011-2020. Eur J Surg Oncol.
Published online August 29, 2024. doi:10.1016/j.ejso.2024.108647
Albanes D, Blair A, Taylor PR. Physical activity and risk of cancer in the NHANES I population. Am J Public Health. 1989;79(6):744-750. doi:10.2105/ajph.79.6.744

Response: *Contextualization: We integrated findings from similar analyses of NHANES datasets into the discussion to provide a comprehensive perspective.*

4. Figures and Tables:

-**Comment:** Figure 1 and Table 1 are actually the same

Response: *Figure 1 was removed, as it was redundant with Table 1.*

-**Comment:** I miss any information regarding Figure 4 and 5

Response: *Figures 4 and 5 have been delated, and the figures have been changed to provide a more comprehensive visualization of the developed models.*

-**Comment:** Table 2: You have a crude OR value for age and non-hispanic white, but the crude or basic model does not include them"

Response: *The crude and adjusted OR values now accurately reflect the variables included in each model.*

Reviewer 2

General Comment:

"Overall, the idea of the article is interesting, given the relevance of the topic and the gap in the need to investigate how regular physical activity acts as a protective factor for the development of cancer. Conducting the evaluation using a real-world dataset is relevant. Below are some comments and suggestions:

1. Introduction

-***Comment:*** The introduction section is well-written and structured. "*Cancer remains a significant global health issue, accounting for millions of new cases and deaths annually*." - I suggest including a citation with objective data on the number of deaths and new cases worldwide.

Response*: Global cancer statistics: A citation providing the number of global cancer cases and deaths has been added.*

Comment: "Investigating the association between physical activity and cancer is crucial in clinical research, as it offers insights into preventive strategies that could lessen the burden of cancer worldwide." - In my opinion, it would be helpful to includef public health strategies or how physical activity can reduce the global burden of cancer.

Response: *Public health strategies: Examples of public health strategies promoting physical activity have been included to contextualize its role in cancer prevention.*

-**Comment:** Preclinical models and clinical-epidemiological studies have shown that exercise can reduce tumor growth and cancer risk." - In this sentence, it would be useful to specify how these reductions occurred and which mechanisms were involved.

Response: *Mechanisms: Specific mechanisms through which physical activity may reduce cancer risk have been described in greater detail, with supporting references.*

-**Comment:** Moreover, there is a need for more sensitivity analyses to account for potential residual confounders, such as diet and genetic predispositions,..." - In this sentence, the authors directly present the gap and mention what remains unanswered in the literature. However, in my opinion, it would be interesting to explain how this study will address this gap. As it is currently written, it seems that the study will add to the many others that do not resolve this gap. In fact, due to the nature of the data, it is complex to address this specific gap concerning the control of the mentioned variables. It would be interesting to directly acknowledge this limitation, highlighting the positive aspect of using and evaluating data from a public dataset. In this regard, it is also relevant to explain what the NHANES 2017-18 is, with at least a brief introduction that clarifies its nature and the type of data it contains.

Responses: *We acknowledged the limitations regarding dietary factors and clarified that while our study cannot fully address these gaps, it leverages NHANES data to provide valuable insights.*

2. Methods

-**Comment:** "This survey collects health-related data from a nationally representative sample of the non-institutionalized U.S. population, providing detailed information about participants' demographic..." - I suggest changing "this" to "the NHANES 2017-18 survey." Additionally, the sentence is in the present tense but should preferably be in the past tense, as the survey has already been administered. The details about the number of questions, the application format, and a reference for the survey results should be mentioned, preferably in a dedicated paragraph providing more detail on the instrument used in the assessment.

Response: *NHANES survey description: We revised the terminology and provided additional details about the NHANES 2017-18 survey, including the number of questions and its administration format.*

Comment: *The authors mention Figure 1 and Table 1 in the Methods section. I suggest that tables and figures should be presented only in the Results section.*

-**Comment:** Regarding the research question: "Does exercise influence cancer cases?" - I believe it would be better to reformulate this question, as the data are cross- sectional and do not include a longitudinal assessment of the physical activity variable. Additionally, it would be better to use the term "regular physical activity," and for the authors to define a cutoff for the number of days per week or hours of exercise per week, mentioning this criterion in the Methods section.

Response: *The research question was reframed to reflect the cross-sectional nature of the data and avoid implying causality.*

-**Comment:** "*The analysis was adjusted for several covariates.*" *- This sentence should be mentioned after explaining which type of analysis was conducted and when the adjustment was made. In this case, I suggest describing the variables used in this paragraph, including the variable type and whether any transformations were* applied. In this context, I believe using the term **multivariable** rather than **multivariate** analysis is more appropriate, as the models are used for one dependent variable.

Response: *Variable transformations: Detailed descriptions of variable transformations and their rationale were added. We have also corrected the definition of the model as a multivariable regression, rather than a multivariate, given that we are comparing to a singular*

outcome, and adjusting to various confounders.

Comment: "Sensitivity analyses were conducted to test the reliability of the findings, examining how other health-related factors might influence the results." - I suggest that the authors explain how this analysis was performed.

Comment In my opinion, a paragraph addressing the inclusion and exclusion criteria is necessary in this section, as they are only mentioned in the "Results" section.

Response: *Inclusion/exclusion criteria: A paragraph outlining the inclusion and exclusion criteria has been added to the methods section.*

Comment *The abbreviation "BMI" should be defined and explained.*

Response: *The abbreviation and its relevance have been clarified.*

3. **Results**:

-**Comment:** "Overall, 39% of participants engaged in regular exercise, completing at least 300 minutes of moderate physical activity." - I suggest citing the absolute number corresponding to the 39% (n = ?) and explaining why 300 minutes was chosen as the cutoff. Is there any regulation on this? Cite the source.

Response: *Exercise statistic: Thank you for this valuable feedback. We would like to clarify that the percentages reported reflect participants who engaged in moderate-intensity physical activity, as defined by activities requiring ≥3.0 METs, rather than implying they met the 300-minute threshold. This clarification has been incorporated into the revised manuscript. Specifically, 47% of participants without a history of cancer (n = 2,340) and 37% of participants with a history of cancer (n = 218) reported engaging in moderate physical activity. To avoid confusion, the sentence has been updated as follows: "Among the participants, 2,340 individuals (47%) without a history of cancer engaged in moderate physical activity, compared to 218 individuals (37%) with a history of cancer." Additionally, we have removed the reference to 300 minutes and explicitly defined moderate physical activity based on the World Health Organization (WHO) guidelines, which classify moderate activity as requiring an energy expenditure of ≥3.0 METs. The updated text now reflects these changes and cites the WHO guidelines appropriately.*

-**Comment:** "This shows that people who exercise regularly are about 36% less likely to get cancer than those who don't." - Given the nature of the data, I do not recommend making this claim. In my understanding, the temporal relationship between physical activity and the prevalence/incidence of cancer needs to be considered. From my interpretation, the physical activity data were collected at the time of the survey (cross-sectional data), while the cancer incidence data were collected in the past (cancer

history). By making this statement, the authors assume that individuals were followed prospectively to assess cancer incidence. However, this was a cross-sectional assessment of cancer prevalence with or without exercise. To clarify this issue, I suggest that the authors specify how the question was asked in the survey and how the physical activity variable was transformed for the regression (this should be described in the Methods section).

Response: *Claims regarding causality have been removed, and the results have been framed to reflect the cross-sectional nature of the data.*

-**Comment:** "In the adjusted model, we also looked at variables like age, gender, race/ethnicity, education level, and annual family income as potential factors that could affect the results." - I suggest that the authors provide details on exactly which variables were included in the model, which ones showed correlation, and what criteria were used for inclusion in the multivariable model (this should be mentioned in the Methods section).

Response: *Covariates in adjusted model: Variables included in the adjusted model are explicitly listed, and the criteria for their inclusion are explained.*

-**Comment:** "The results of these sensitivity analyses show that the protective effect of physical activity on cancer risk is probably not only attributable to other lifestyle or demographic factors. That said, it's worth noting that there might be other factors at play, like..." - this paragraph should be included in the Discussion section.

Response: *Statements regarding sensitivity analyses have been moved to the discussion.*

4. **Discussion and Final Comments**:

-***Comment:*** In my opinion, the authors should clarify the exact question used in the survey to make the conclusions easier to evaluate in the study. This description also applies to how "regular physical activity" was defined (citing the source and how the question was made in the survey). Since these are secondary data from the completion of the NHANES survey, the association between regular physical activity and cancer incidence assumes that the patients were followed over time, i.e., longitudinal data, and assessed concerning cancer incidence, which does not seem to be the case. If cancer data are available, I suggest that the authors include a description of the types of cancer in the study population, as risk factors vary depending on the histological subtype. In the introduction, the authors mention that the relationship between physical exercise and cancer incidence is well-established, but the confounder's evaluation of diet remains a gap, and they aim to address this gap. However, there is no description of this adjustment for dietary factors, which I understand is complex. Therefore, I suggest that the authors more precisely define the main focus of the study without mentioning the adjustment for dietary factors. The statistical analysis is well-described and relevant. I recommend that the discussion focus on findings from the literature that support the study's result, how the NHANES dataset differs from other subject pools in the literature, and the importance of this finding in advancing knowledge.

Response: *Survey question: Details of the question on cancer diagnosis and physical activity were added to clarify the basis of our conclusions. Cancer types: Acknowledged the limitation of aggregating all cancer types and discussed its potential impact*

> -***Comment:*** Finally, I believe the study's methodological structure was well conducted, and the rationale for its development is plausible."

Response: *The discussion now emphasizes our unique contributions and situates findings within the broader literature.*