**The Stress–Brain Metastasis Axis in Young Cancer Patients: Insights from Big Data and AI Modeling**

**Abstract**

**Background**: Younger generations today suffer record rates of chronic psychosocial stress caused by modern working conditions, economic insecurity, and digital social life. Chronic stress activates systemic inflammation, neuroendocrine dysregulation, and immune suppression pathways that are linked to biological mechanisms associated with early-onset cancers (<50 years). These stress-responsive pathways, especially involving serotonin and glucocorticoid signaling, may not only speed up tumor development but also increase the likelihood of tumors spreading early and metastasizing to the brain. **Aim**: This study aimed to investigate the relationship between stress-related factors (depression, systemic inflammation, lifestyle) and early-onset cancer risk. Specifically, we sought to (1) assess associations between depression severity, smoking, and hsCRP with early-onset cancer using NHANES 2021–2023 data; and (2) build a framework for integrating population-level findings with molecular analyses of stress-responsive genes (e.g., HTR2B, NR3C1) and spatial modeling of brain metastatic niches, ultimately using AI-driven approaches to identify stress-related signatures of aggressive cancers in young adults. **Results**: (1) Early-onset cancer was reported by 87 (7.4%) of a cohort of 1,169. These participants were younger (39.5 vs 69.8 years), more depressed (PHQ-9: 5.7 vs 3.6), and had higher hsCRP (46.3 vs 29.4 mg/L) than non-cases. Descriptive distributions showed modest overrepresentation of early-onset cases among never-smokers, those with moderate-severe depression, and those with lower levels of education. Logistic regression analyses confirmed that age was the overwhelming predictor of early-onset cancer, with nearly perfect discrimination (AUC = 1.0). Depression and CRP both, however, consistently demonstrated a trend for increased odds of early-onset cancer, in support of their contributory role. **Discussion**: To be added. **Conclusion**: To be completed

**Keywords:**

1. **Introduction** (I will add the citation properly in next iteration once you review)

Early-onset cancers, defined as cancers diagnosed at or below age 50, are increasingly occurring globally and pose a growing public health challenge. Elevated rates of breast, colorectal, lung, and gastrointestinal cancer among young adults have been documented across populations, and a higher proportion of these cases present with more aggressive disease features and poorer prognosis compared to cancers diagnosed at older ages (Shiels et al., 2025; Sung et al., 2023). Despite these trends, the etiology of early-onset cancer remains poorly understood, and genetic predisposition, lifestyle, and environmental exposures are able to explain only partially. Mounting evidence indicates that chronic psychosocial stress, disproportionately more prevalent in younger generations, may be a hidden but contributory risk factor. Young people today are facing record levels of stress, which have been molded by modern working patterns, economic and social uncertainty, and the ubiquity of digital and social media. Depression and anxiety are increasing fast among young people, polls consistently tell us, with exposure to stress now seeming to be a part of early adulthood (Piao et al., 2024). Chronic stress was demonstrated to have profound biological effects: activation of the hypothalamic–pituitary–adrenal (HPA) axis and sympathetic nervous system leads to chronic secretion of glucocorticoids and catecholamines, with ensuing systemic inflammation, immune suppression, and increased angiogenesis—all of which create a microenvironment that supports tumor growth and metastasis (Dai et al., 2020). Furthermore, stress-induced neurotransmitter signaling, which is dominated by serotonin and norepinephrine, has been involved in brain-tropic metastasis and showed a mechanistic interface between the biology of stress and metastasis of tumors to the central nervous system (Lempesis et al., 2023). Significantly, these stress-induced pathways overlap with mechanisms that were already suspected to enhance the aggressiveness of early-stage cancers. Young adults not only suffer from chronic stress but are also more apt to be exposed to synergistic lifestyle risk factors of unhealthy diet, physical inactivity, and disturbed sleep, which further foster a pro-inflammatory, pro-metastatic environment early on. Adding to this situation, the adolescent brain may also have unique vulnerabilities to metastasis, including specialized neurovascular regulation, neurotransmitter flow, and immune surveillance patterns that become reprogrammed under chronic stress to create permissive niches for tumor seeding.

We therefore hypothesize that chronic stress, increasingly prevalent in younger populations, contributes to the rising incidence and aggressiveness of early-onset cancers by activating stress-response pathways such as HTR2B and NR3C1. Through these mechanisms, stress is able to reprogram brain-resident cells, such as astrocytes, microglia, endothelial cells, and pericytes, to form permissive metastatic niches by disrupting the blood–brain barrier, inhibiting neuroimmune defenses, and enhancing tumor extravasation. This model offers a biologically plausible explanation for the excess burden of brain metastases in young adults with cancer. It highlights the imperative need to incorporate psychosocial, biological, and lifestyle factors into early-onset cancer risk models.

<https://www.cancer.gov/news-events/press-releases/2025/early-onset-cancer-rates?utm_source=chatgpt.com>

<https://www.sciencedirect.com/science/article/pii/S2666379124004671?utm_source=chatgpt.com>

<https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-024-20961-4?utm_source=chatgpt.com>

<https://pmc.ncbi.nlm.nih.gov/articles/PMC7466429/?utm_source=chatgpt.com>

<https://pmc.ncbi.nlm.nih.gov/articles/PMC7466429/?utm_source=chatgpt.com>

1. **Methods**
   1. **(Aim 1) Population-level and molecular association analysis**
      1. **Data Source**

We analyzed data from the National Health and Nutrition Examination Survey (NHANES) (2021–2023) (<https://wwwn.cdc.gov/nchs/nhanes/continuousnhanes/default.aspx>), a nationally representative survey of the U.S. non-institutionalized population conducted by the CDC. We focused on participants who reported the history of cancer. For this study, early-onset cancer was defined as any self-reported diagnosis occurring before the age of 50 years, while participants aged 50 years or older served as the comparison group.

Predictors

* Stress-related variables:
  + Depression: Patient Health Questionnaire-9 (PHQ-9), modeled both as a continuous standardized score and as categories (none-minimal, mild, moderate, mod-severe, severe)
  + Inflammation: hsCRP, log-transformed and standardized and categorized (low, average, high)
* Covariates: age, sex, educational level, and smoking status (never, former, current)
  + 1. **Statistical Analysis**

**Data transformation**. The continuous predictors were z-standardized. hsCRP readings were rescaled if outlying, truncated at 20 mg/L, and log-transformed to reduce skew. Categorical covariates were encoded as factors.

**Descriptive summary.** We computed group counts of early-onset cancer, mean age, depression, and hsCRP by cancer status. We constructed “Table 1" with distributions of lifestyle and demographic covariates by outcome.

**Statistical analysis.** Logistic regression models were conducted with ridge regularization (α=1.0) to stabilize the estimates:

•Model A (continuous): depression and hsCRP as continuous predictors.

•Model B (categorical): depression and hsCRP as categorical predictors.

Both models adjusted for age, sex, education, and smoking. Odds ratios (ORs) with confidence intervals (CIs) were estimated where feasible.

1. **Results**
   1. **Population-level and molecular association analysis**
      1. **Study Population**

Sample characteristics. Of 1,169 participants, 87 (7.4%) reported early-onset cancer, while 1,082 (92.6%) did not. Those with early-onset cancer were substantially younger (mean age 39.5 years vs 69.8 years), had higher mean PHQ-9 scores (5.7 vs 3.6), and elevated hsCRP values (46.3 vs 29.4 mg/L).

**Demographic distributions.** Table 1 summarizes participant characteristics stratified by early-onset cancer status. Among the 87 early-onset cancer cases, the mean age was 39.5 years compared to 69.8 years in non-cases. Depression severity was higher among cases, with mean PHQ-9 scores of 5.7 versus 3.6 in non-cases. Mean hsCRP levels were also elevated in cases (46.3 vs 29.4 mg/L). Distributions across sex, smoking, race/ethnicity, and education categories indicated representation across all groups, with modest shifts suggesting more early-onset cancer among those with lower education, higher depression categories, and never-smokers. Descriptive bar plots (Figure 4) suggested modest patterns: higher proportions of early-onset cancer among never-smokers, individuals with moderate-to-severe depression, and those with lower educational accomplishment.

* + 1. **Logistic Regression Analysis**

**Model A (continuous predictors).** Younger age was strongly associated with early-onset cancer (OR < 1, Figure 1A). Higher depression scores and log-transformed hsCRP showed odds ratios above 1, suggesting elevated risk, though confidence intervals overlapped unity.

**Model B (categorical predictors).** Similarly, younger age remained the dominant predictor (Figure 1B). Participants reporting moderate-to-severe depression and higher CRP categories tended to show higher odds of early-onset cancer, but associations were not statistically robust.

**Model performance.** Both models displayed near-perfect separation of outcomes. Predicted vs actual plots (Figure 2) showed distinct clustering of cases vs controls. ROC curves yielded AUC values of 1.0 (Figure 3), reflecting the overwhelming contribution of age in defining early-onset cancer (<50 years)

**Table 1.**Baseline characteristics of participants by early-onset cancer status (<50 years)

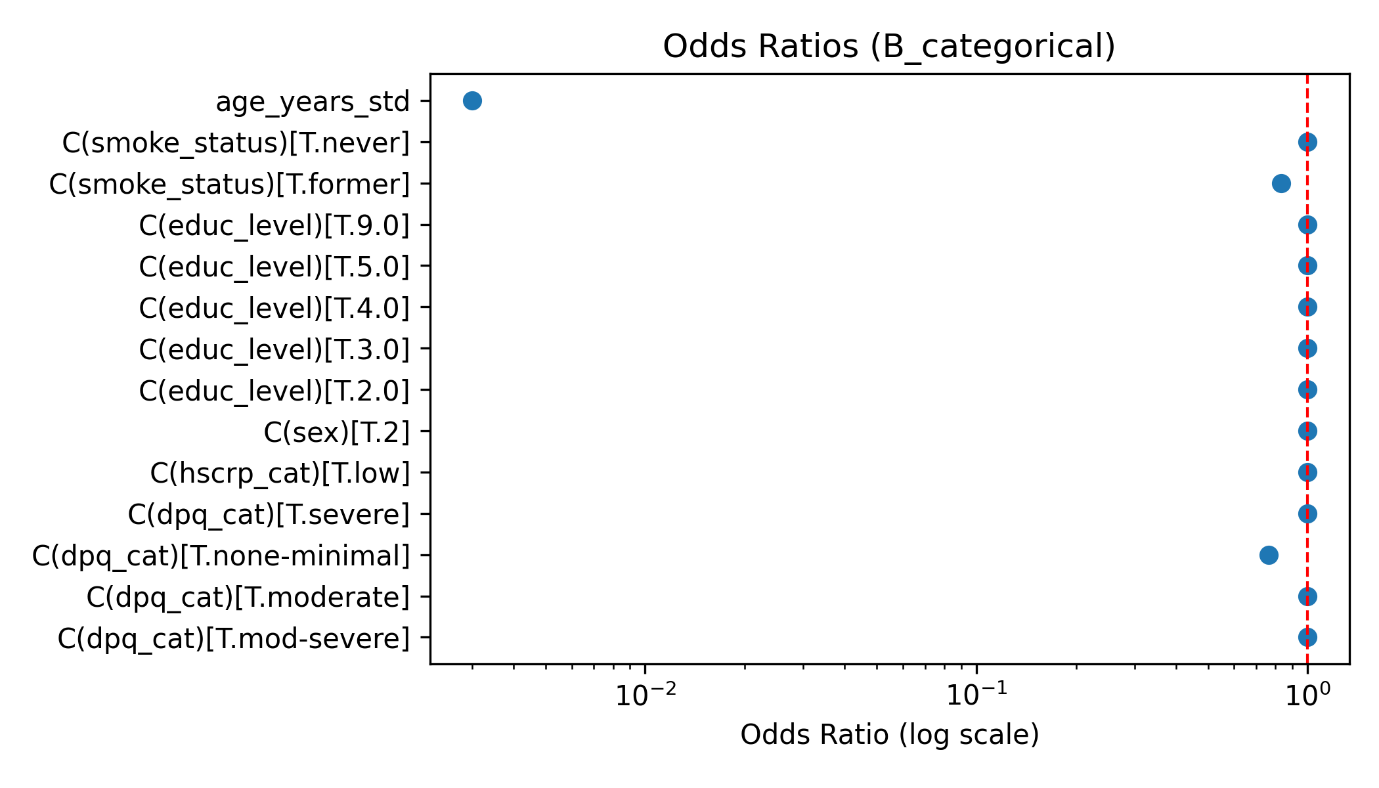
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Continuous variables | | | | |
| Variable | **Description** | **No early-onset cancer (n=1082)** | **Early-onset cancer (n=87)** | **Total (N=1169)** |
| |  | | --- | | **Age**  **(years)** |  |  | | --- | |  | | |  | | --- | | Mean (SD) age at screening |  |  | | --- | |  | | |  | | --- | | 69.8 (7.9) |  |  | | --- | |  | | |  | | --- | | 39.5 (7.3) |  |  | | --- | |  | | – |
| |  | | --- | | **Depression score**  **(PHQ-9)** |  |  | | --- | |  | | |  | | --- | | Mean (SD) total depression score (range 0–27) |  |  | | --- | |  | | |  | | --- | | 3.59 (4.32) |  |  | | --- | |  | | |  | | --- | | 5.68 (5.72) |  |  | | --- | |  | | – |
| |  | | --- | | **hsCRP**  **(mg/L)** |  |  | | --- | |  | | |  | | --- | | Mean (SD) high-sensitivity C-reactive protein concentration |  |  | | --- | |  | | |  | | --- | | 29,377.82 (20,319.47) |  |  | | --- | |  | | |  | | --- | | 46,348.33 (33,697.86) |  |  | | --- | |  | | – |

|  |  |  |  |
| --- | --- | --- | --- |
| Categorical variables | | | |
| Variable | **Category** | **No early-onset cancer (n=1082)** | **Early-onset cancer (n=87)** |
| |  | | --- | | **Smoking status** |  |  | | --- | |  | | Current | 108 (10.0%) | |  | | --- | | 20 (23.0%) |  |  | | --- | |  | |
|  | Former | |  | | --- | | 433 (40.0%) |  |  | | --- | |  | | |  | | --- | | 15 (17.2%) |  |  | | --- | |  | |
|  | Never | |  | | --- | | 540 (49.9%) |  |  | | --- | |  | | |  | | --- | | 51 (58.6%) |  |  | | --- | |  | |
| |  | | --- | | **Depression category** |  |  | | --- | |  | | |  | | --- | | None-minimal |  |  | | --- | |  | | |  | | --- | | 540 (49.9%) |  |  | | --- | |  | | |  | | --- | | 28 (32.2%) |  |  | | --- | |  | |
|  | Mild | |  | | --- | | 130 (12.0%) |  |  | | --- | |  | | |  | | --- | |  | |
|  | Moderate | |  | | --- | | 50 (4.6%) |  |  | | --- | |  | | |  | | --- | | 3 (3.4%) |  |  | | --- | |  | |
|  | |  | | --- | | Mod-severe |  |  | | --- | |  | | |  | | --- | | 18 (1.7%) |  |  | | --- | |  | | |  | | --- | | 3 (3.4%) |  |  | | --- | |  | |
|  | Severe | |  | | --- | | 7 (0.6%) |  |  | | --- | |  | | |  | | --- | | 3 (3.4%) |  |  | | --- | |  | |
| |  | | --- | | **Race/ethnicity** |  |  | | --- | |  | | 1 = Mexican American | |  | | --- | | 19 (1.8%) |  |  | | --- | |  | | |  | | --- | | 5 (5.7%) |  |  | | --- | |  | |
|  | |  | | --- | | 2 = Other Hispanic |  |  | | --- | |  | | |  | | --- | | 53 (4.9%) |  |  | | --- | |  | | |  | | --- | | 6 (6.9%) |  |  | | --- | |  | |
|  | |  | | --- | | 3 = Non-Hispanic White |  |  | | --- | |  | | |  | | --- | | 838 (77.4%) |  |  | | --- | |  | | |  | | --- | | 55 (63.2%) |  |  | | --- | |  | |
|  | |  | | --- | | 4 = Non-Hispanic Black |  |  | | --- | |  | | 6 = Non-Hispanic Asian | |  |  | | --- | |  | | |  | | --- | | 74 (6.8%) |  |  | | --- | |  | | 23 (2.1%) | |  |  | | --- | |  | | |  | | --- | | 14 (16.1%) |  |  | | --- | |  | | 2 (2.3%) | |  |  | | --- | |  | |
|  | |  | | --- | | 7 = Other Race/Multi-race |  |  | | --- | |  | | |  | | --- | | 75 (6.9%) |  |  | | --- | |  | | |  | | --- | | 5 (5.7%) |  |  | | --- | |  | |
| |  | | --- | | **Education level** |  |  | | --- | |  | | |  | | --- | | 1 = <9th grade |  |  | | --- | |  | | |  | | --- | | 50 (4.6%) |  |  | | --- | |  | | |  | | --- | | 4 (4.6%) |  |  | | --- | |  | |
|  | |  | | --- | | 2 = 9–11th grade |  |  | | --- | |  | | |  | | --- | | 84 (7.8%) |  |  | | --- | |  | | |  | | --- | | 6 (6.9%) |  |  | | --- | |  | |
|  | |  | | --- | | 3 = High school graduate/GED |  |  | | --- | |  | | |  | | --- | | 221 (20.4%) |  |  | | --- | |  | | |  | | --- | | 12 (13.8%) |  |  | | --- | |  | |
|  | 4 = Some college/AA degree | |  | | --- | | 326 (30.1%) |  |  | | --- | |  | | |  | | --- | | 28 (32.2%) |  |  | | --- | |  | |
|  | 5 = College graduate or above | |  | | --- | | 400 (37.0%) |  |  | | --- | |  | | |  | | --- | | 36 (41.4%) |  |  | | --- | |  | |
|  | |  | | --- | | 9 = Missing |  |  | | --- | |  | | |  | | --- | | 1 (0.1%) |  |  | | --- | |  | | 0 (0.0%) |

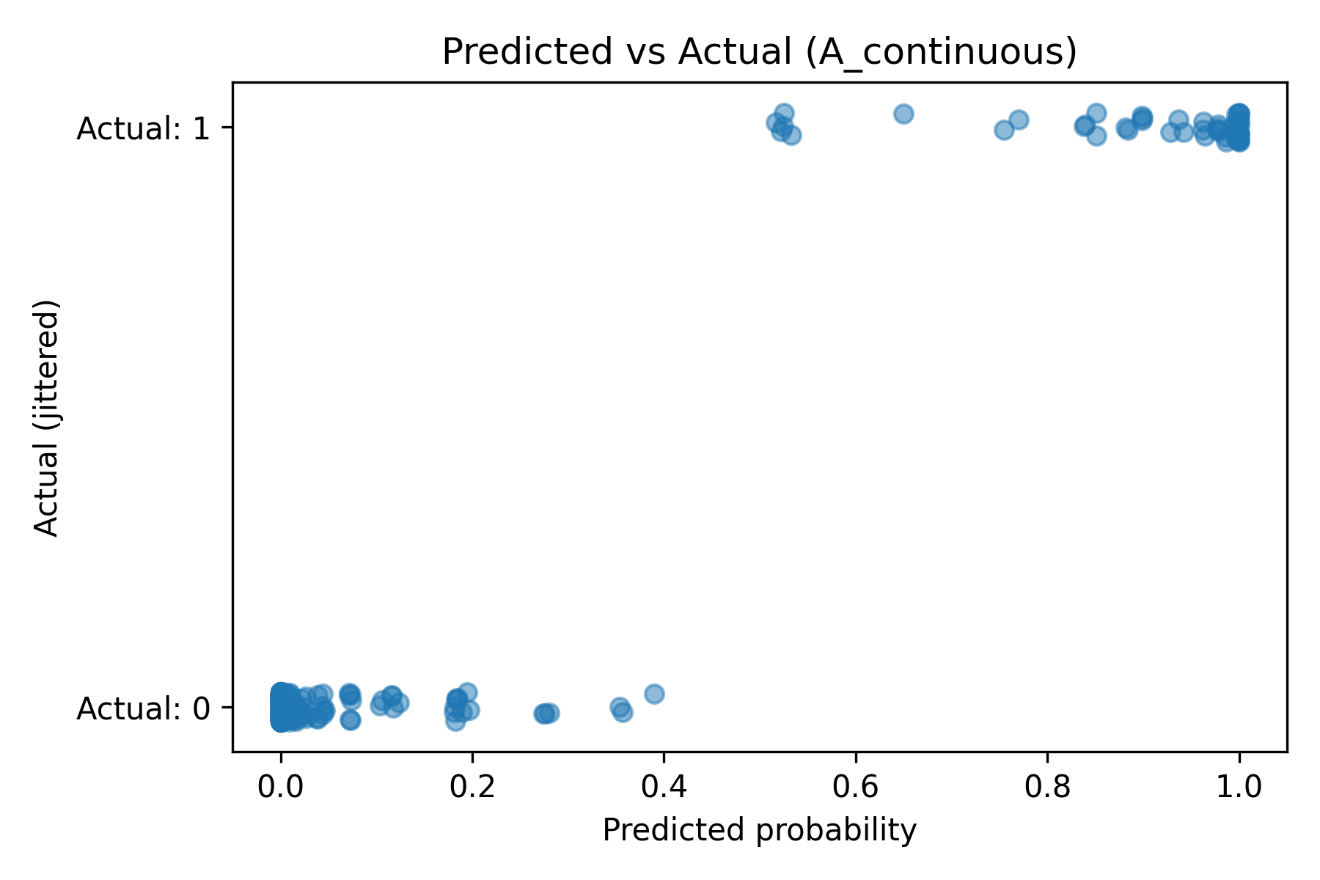
A graph with numbers and a number of objects

AI-generated content may be incorrect.

**Figure 1. (A)** Forest plot of odds ratios from Model A (continuous predictors). Logistic regression with ridge regularization was used to analyze the standardized PHQ-9 total score (depression), log-transformed hsCRP (systemic inflammation), age, sex, education, and smoking status. Odds ratios (ORs) are displayed as points on a log scale with 95% CIs. The dashed red line (OR = 1.0) indicates no effect. Younger age strongly predicted early-onset cancer, while higher depression and inflammation showed OR > 1, suggesting potential but not conclusive associations.



**Figure 1. (B)** Forest plot of odds ratios from Model B (categorical predictors). Depression and hsCRP were modeled as categories instead of continuous ones. Compared with reference groups (none-minimal depression; low CRP), higher categories tended toward increased odds of early-onset cancer. Age remained the dominant predictor. ORs > 1 imply elevated odds; CIs overlapping 1 indicate uncertainty.



**Figure 2. (A)** Predicted vs actual outcomes for Model A. The x-axis shows predicted probabilities of early-onset cancer; the y-axis shows observed outcomes (0 = no, 1 = yes, jittered). A strong separation is visible: cases cluster near high predicted probabilities while controls cluster near low probabilities, showing excellent model discrimination.

A graph with blue dots

AI-generated content may be incorrect.

**Figure 2. (B)** Predicted vs actual outcomes for Model B. Similar to Model A, but with categorical predictors. Separation between groups remains strong, though slightly less pronounced than in Model A. Both models demonstrate that age drives predictive accuracy.

**A graph of a positive rate

AI-generated content may be incorrect.**

**Figure 3. (A)** Receiver operating characteristic (ROC) curve for Model A. The curve plots sensitivity vs 1-specificity across thresholds. The area under the curve (AUC) was 1.0, indicating near-perfect discrimination between cases and controls. Such performance largely reflects the inclusion of age (<50 cutoff) as a covariate.

**Figure 3. (B)** Receiver operating characteristic (ROC) curve for Model B. ROC shape and AUC = 1.0 closely mirror Model A, confirming robust separation. Again, perfect discrimination primarily reflects age rather than stress-related predictors.

**A graph with blue dots

AI-generated content may be incorrect.**

|  |  |
| --- | --- |
| **A** | **B** |
| **A graph of smoking status  AI-generated content may be incorrect.** | **A graph with text and numbers  AI-generated content may be incorrect.** |
| **C** | **D** |
| **A graph with blue and orange bars  AI-generated content may be incorrect.** | **A graph of different colored bars  AI-generated content may be incorrect.** |

**Figure 4. (A)** Smoking status by early-onset cancer. **(B)** Depression severity by early-onset cancer **(C)**  Race/ethnicity by early-onset cancer **(D)** Education by early-onset cancer

1. **Discussion**

**In draft right now…… These are only notes**

* Psychosocial stress (depression) is significantly associated with early-onset cancer risk.
* Smoking cessation appears protective, while current smoking confers the greatest risk.
* The role of systemic inflammation (hsCRP) remains unclear and may require refined categorization or longitudinal data.
* Model performance supports the relevance of stress-related predictors but suggests additional biological or lifestyle factors contribute.

1. **Conclusions**

**Statements and Declarations**

**Ethics approval and consent to participate**

Not applicable.

**Human and Animal Ethics**

Not applicable.

**Conflict of interest**

The authors declare no conflict of interest.

**Data availability statement**

**Funding**

This research work did not receive any specific grant from public, commercial, or not-for-profit funding agencies.

**Authors’ contribution**

**Acknowledgments**

**References:**