Research paper

A marginal structural model analysis for the effect modification by education on the association between cancer diagnosis history and major depressive symptoms: Findings from Midlife Development in the U. S. (MIDUS)

Lumin Liu a,1, Junjie Lu b,1, Jiarui Yang c, Yiyue Dong a, Ping Yin a,* , Yuelai Chen a,*

a LongHua Hospital Shanghai University of Traditional Chinese Medicine, Shanghai, China
b Department of Social and Behavioral Sciences, Harvard University T.H. Chan School of Public Health, Boston, USA
c Department of Biomedical Engineering, Boston University, Boston, USA

ARTICLE INFO

Keywords:
Cancer diagnosis history
Major depressive disorder (MDD)
Education level
Longitudinal study

ABSTRACT

Background: Limited research has employed a longitudinal approach to investigate the role of education level as an effect modifier on the relationship between cancer diagnosis history and the experience of major depressive disorder (MDD) with a nationally representative sample.

Methods: We harnessed data from three installments of the MIDUS Longitudinal study (n = 7108). A Marginal Structural Model facilitated the investigation of associations between a history of cancer diagnosis, MDD, and potential modifying effects of education level. Inverse probability weighting helped manage confounding factors.

Results: Findings indicated that a cancer diagnosis made one year prior was linked with 3.741 times greater odds of experiencing MDD (95 % CI: 1.411–9.918, p < 0.01). This connection was absent for diagnoses made two years earlier. Among individuals with education up to high school, a recent cancer diagnosis significantly increased the likelihood of MDD in the subsequent wave by 3.45 times (95 % CI: 1.31–9.08, p < 0.05). This pattern was not apparent among better-educated individuals.

Limitations: As the exposure variable was dependent on self-reported questionnaires, recall bias could be a potential limitation. Moreover, unaccounted variables like genetic factors could introduce confounding.

Conclusions: A recent cancer diagnosis, particularly among less educated individuals, correlated with an increased probability of MDD, while the impact was not observed for older diagnoses. These findings emphasize that the timing of a cancer diagnosis and education level need consideration in the mental health assessment of cancer survivors.

1. Introduction

1.1. Cancer related major depressive disorder (MDD)

Cancer poses a significant challenge to life expectancy and quality of life across the globe. According to the 2020 Global Cancer Observatory (GLOBOCAN) database, which covers 36 cancer types in 185 countries, approximately 19.3 million new cancer cases and nearly 10.0 million cancer-related deaths were reported (Sung et al., 2021). Although survival rates for cancer patients are relatively high in most countries, the profound psychological impact of the disease has emerged as a critical clinical and public health concern (Pilevarzadeh et al., 2019).

The well-documented adverse effects of cancer on mental health of patients indicated increasing prevalence of post-diagnosis mental disorders (Krebber et al., 2014; Linden et al., 2012; Y.-H. Wang et al., 2020). Current estimates suggest that depression affects roughly 32.2 % of cancer survivors (Krebber et al., 2014). Furthermore, a heightened risk of mental disorders may negatively impact cancer recurrence, mortality, and overall quality of life (X. Wang et al., 2020). Numerous studies have identified associations between cancer diagnoses and depression, with factors including individual characteristics, social and contextual elements, prior psychological factors, psychological

* Corresponding authors.
E-mail addresses: junjielu@hsph.harvard.edu (J. Lu), jryang@bu.edu (J. Yang), bingxue616@163.com (P. Yin), chenyuelai@163.com (Y. Chen).
1 Co-first authors.

https://doi.org/10.1016/j.jad.2023.08.123
Received 7 June 2023; Received in revised form 23 August 2023; Accepted 24 August 2023
Available online 26 August 2023
0165-0327/© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
responses to diagnosis, and cancer treatment characteristics (Linden et al., 2012; McDaniel et al., 1995; Niedzwiedz et al., 2019).

1.2. The effect of education level in cancer related MDD

The role of education merits particular attention for the relationship between cancer diagnosis and MDD. However, existing findings are inconsistent. Some studies suggest that higher level of education serves as a protective factor against depression, owing to increased self-regulation and access to treatment (Lopes et al., 2022; Mols et al., 2018). Some studies conducted among cancer patients found that low level of education was associated with depression (Alcalar et al., 2012; Mehnert and Koch, 2008; Morrill et al., 2008). Conversely, other studies have found no association between depression severity and education level, while some even posit that higher education may contribute to depression (Muzzatti et al., 2018). Potential explanations for this include greater psychological discrepancies in highly educated patients when faced with cancer and a subsequent decline in self-esteem (Xia et al., 2020). Moreover, one study revealed that highly educated cancer patients were more likely to delay treatment during the coronavirus disease 2019 (COVID-19) pandemic, potentially exacerbating depressive symptoms (He et al., 2022).

1.3. Rationale for a longitudinal approach

The long-term effects of a cancer diagnosis on MDD are of critical importance in understanding the psychological implications for cancer survivors. While the immediate impact of a cancer diagnosis can often lead to an onset of depression, it is the enduring consequences that may have a lasting effect on a survivor’s mental health and overall well-being (Yen et al., 2022). Studies have demonstrated that depression can persist for over five years post-diagnosis, indicating the sustained burden of MDD on cancer survivors (Lopes et al., 2022). These long-term effects may manifest in various ways, including reduced quality of life, impaired social functioning, and challenges in maintaining employment or sustaining personal relationships (Badr and Taylor, 2008; Feuerstein et al., 2010; Tan et al., 2002). Furthermore, most existing studies investigated cross-sectional designs, which often overlook the long-term psychological impact of cancer. By examining the long-term relationship between cancer diagnosis and MDD in a longitudinal manner, researchers can develop a more comprehensive understanding of the factors contributing to the persistence of depression in cancer survivors, ultimately informing targeted interventions to support individuals in managing their mental health throughout their recovery.

Consequently, this study aims to: a) investigate the effects of cancer diagnosis history on MDD over time, and b) explore the role of education level as an effect modifier in the relationship between cancer diagnosis and MDD. This study provides key insights into the long-term implications of cancer on mental health, considering the influence of education level, ultimately contributing to the development of more targeted and effective interventions for cancer patients. (3)

2. Methods

2.1. Study sample and population

The Midlife in the United States (MIDUS) Longitudinal study, supported by the John D. and Catherine T. MacArthur Foundation Research Network on Successful Midlife Development (MIDMAC), aimed to identify the key biomedical, psychological, and social factors that contributed to individuals achieving good health, psychological well-being, and social responsibility during their adult years (Ryff et al., 2010).

The baseline sample of MIDUS was derived from four sources: (1) a national random digit dialing (RDD) sample (n = 3487); (2) oversamples from five metropolitan areas in the U.S. (n = 757); (3) siblings of individuals from the RDD sample (n = 950); and (4) a national RDD sample of twin pairs (n = 1914).

The eligible participants consisted of non-institutionalized, English-speaking adults in the contiguous United States, aged between 25 and 74. (6a) (10)

MIDUS employed three waves of longitudinal data collection (1995–1996, 2004–2006, and 2013–2014). In each wave, participants were invited to complete a 30-minute phone interview and two 50-page Self-Administered Questionnaire (SAQ) instruments. The data from each wave were subsequently compiled into a single dataset. The first wave of the MIDUS study (Wave 1) amassed survey data from a total of 7108 participants. The present study utilized data from all three waves, with the sample characteristics displayed in Table 1. (5)

2.2. Measurements

2.2.1. Exposure: cancer diagnosis history

Cancer diagnosis history was assessed during the Wave 1 phone interview. Participants were asked, “Have you ever had cancer?” Responses were categorized as “Yes” or “No.” Cancer diagnosis was re-evaluated at Wave 2 and Wave 3. Participants without a cancer diagnosis history were considered the reference group.

2.2.2. Outcome: major depressive disorders (MDD)

The primary outcome of this study is the presence of major depressive disorder (MDD) at Wave 3, operationalized as a binary variable provided by the MIDUS team. MDD diagnosis is based on the definitions and criteria outlined in the third edition-revised of the American Psychiatric Association’s (APA) Diagnostic and Statistical Manual of Mental Disorders (Spitzer et al., 1992). A diagnosis of MDD necessitates a minimum duration of two weeks characterized by either a depressed mood or anhedonia most of the day, nearly every day, and at least four additional associated symptoms commonly found in depression, such as disturbances in eating, sleeping, energy levels, concentration, self-worth, and suicidal ideation or actions (Kessler and Walters, 1998).

MDD was assessed using the screening versions of the World Health Organization’s (WHO) Composite International Diagnostic Interview, Version 10 (CIDI) (Haro et al., 2006; Kessler et al., 1998). The reliability and clinical validity of CIDI diagnoses have been demonstrated through WHO Field Trials and other methodological studies (Blazer et al., 1994; Wittchen, 1994). Although MDD was measured across all three waves, the study’s primary outcome was designated as MDD at Wave 3 to facilitate a longitudinal analysis of the effects of cancer diagnosis on MDD at Wave 3. Participants without MDD at Wave 3 served as the reference group. As a sensitivity analysis, MDD was also treated as a continuous variable based on CIDI scores provided in the MIDUS database. The results of this analysis are provided in the Supplementary Tables 2 and 3. (12e)

Given the strong association between MDD and sleep disturbances, sleep problems at Wave 3 were designated as a secondary outcome. Participants were asked, “In the past 12 months, have you experienced or been treated for chronic sleep problems?” Responses were categorized as either “Yes” or “No.” The reference group for this secondary outcome consisted of individuals without sleep problems at Wave 3.

2.2.3. Effect modifier

The study sought to determine whether the education level served as a potential effect modifier on the association between previous cancer diagnosis (Wave 1, Wave 2, Wave 3) and subsequent sleep problems (Wave 3). Participants’ education levels were assessed only at Wave 1.
MIDUS researchers classified education levels into four categories: some grade school to some high school, General Education Diploma (GED) to graduated high school, some college (no bachelor’s degree), and college graduate to doctorate or professional degree. People with the least level of education were set as the reference group.

### Covariates

#### Biological sex assigned at birth

Participants were asked about their sex during the phone interview. Responses were categorized as Female or Male. Sex was assessed only at Wave 1 and treated as a binary variable, with female as the reference group.

#### Age

Participants’ ages were assessed during the phone interview. For those missing age information, it was updated based on the SAQ responses. Age was measured at Wave 1 and treated as a continuous variable.

#### Race

Participants were classified into five categories based on phone interviews and SAQs: White, Black and/or African American, Native American or Aleutian Islander/Eskimo, Asian or Pacific Islander, and Multiracial. Race was assessed at Wave 1, and treated as a categorical variable, with White as the reference group.

#### Current smoke

Participants were asked about their current smoking behavior during the phone interview, specifically, “Do you smoke cigarettes regularly now?” Responses were categorized as “Yes” or “No.” This variable was assessed in all three waves of the study. Responses were binarized, with non-smokers as the reference group.

#### Alcohol problem

MIDUS utilized the Alcohol Screening Test (AST) to measure alcohol problems. AST is a 5-item scale, with representative questions such as, “Were you ever, during the past 12 months, under the effects of alcohol or feeling its after-effects in a situation which increased your chances of getting hurt - such as when driving a car or boat, or using knives or guns or machinery?” and “Did you ever, during the past 12 months, have any emotional or psychological problems from using alcohol – such as feeling depressed, being suspicious of people, or having strange ideas?” The alpha of AST was 0.68 (Conigrave et al., 2002).

A dichotomous variable, “No alcohol problems,” was constructed: 1 for no alcohol problems (alcohol problems = 0) or 0 for having alcohol problems (alcohol problems > 0). MIDUS provided the calculated variable for Wave 2 and Wave 3, which were used in the analysis. We applied the same scaling method to calculate the AST and generate the variable at Wave 1. Alcohol problem data from all three waves were used in the analysis, with participants without alcohol problems serving as the reference group.

#### Body Mass Index (BMI)

BMI was calculated by dividing the weight in kilograms by the height in meters squared. Since height was recorded in inches on the questionnaire, MIDUS researchers converted it to meters by multiplying the inches by 0.0254. Similarly, since weight was recorded in pounds, researchers converted it to kilograms by multiplying the pounds by 0.4536. BMI was calculated for all three waves and treated as a continuous variable.

### Statistical analysis

We employed the Directed Acyclic Graph (DAG) to explicitly outline the assumptions regarding the association of variables of interest, as depicted in Fig. 1. We postulated that baseline variables C0 (sex, age, race) were associated with exposures A1, A2, A3 (cancer diagnosis at three waves), the outcome Y (MDD at Wave 3), and covariates V1, V2, V3 (smoking behavior, alcohol problems, and BMI at all three waves). Additionally, we assumed that both exposures (A1, A2, A3) and covariates (V1, V2, V3) were time-varying and interrelated. The descriptive
analysis of time-varying exposures and covariates was shown in Supplementary Figs. 1 to 4. Given that traditional regression analysis may result in biased outcomes when dealing with time-varying exposures and covariates, as noted by VanderWeele et al., we employed a marginal structural model (MSM) to investigate the impact of cancer diagnosis history on MDD and the potential effect modification by education status (VanderWeele et al., 2011).

### 2.3.1. Confounding weight

The initial step in fitting an MSM involved calculating the inverse probability of exposure weights for each individual. Specifically, the weight represented the inverse probability of an individual being exposed to cancer diagnosis at Wave 1, Wave 2, and Wave 3 as a function of the covariates associated with the exposure. The weights took the form:

\[
W_{\text{wave 1}} = \frac{1}{Pr(A1 | C0, V1)} \quad (1)
\]

\[
W_{\text{wave 2}} = \frac{1}{Pr(A2 | C0, A1, V1, V2)} \quad (2)
\]

\[
W_{\text{wave 3}} = \frac{1}{Pr(A3 | C0, A1, V1, A2, V2)} \quad (3)
\]

However, the numerator in Eqs. (1), (2), and (3) was 1, which leads to unstable weights (Hernan and Robins, n.d.). As a result, we stabilized the weight by using the proportion of exposure at each wave as the numerator instead. The stabilized weights took the form:

\[
SW_{\text{wave 1}} = \frac{Pr(A1)}{Pr(A1 | C0, V1)} \quad (4)
\]

\[
SW_{\text{wave 2}} = \frac{Pr(A2)}{Pr(A2 | C0, V1, A1, V1, V2)} \quad (5)
\]

\[
SW_{\text{wave 3}} = \frac{Pr(A3)}{Pr(A3 | C0, A1, V1, A2, V2)} \quad (6)
\]

To further investigate the effect modification by education level on the association of interest, we stabilized the weight by including the education level as a variable in the numerator. The education level-stabilized weights took the form:

\[
SW_{\text{Education}}^{\text{wave 1}} = \frac{Pr(A1 | Education)}{Pr(A1 | C0, V1)} \quad (7)
\]

\[
SW_{\text{Education}}^{\text{wave 2}} = \frac{Pr(A2 | Education)}{Pr(A2 | C0, A1, V1, V2)} \quad (8)
\]

\[
SW_{\text{Education}}^{\text{wave 3}} = \frac{Pr(A3 | Education)}{Pr(A3 | C0, A1, V1, A2, V2)} \quad (9)
\]

The final confounding weight took the form:

\[
SW_{\text{Education}} = SW_{\text{Education}}^{\text{wave 1}} \times SW_{\text{Education}}^{\text{wave 2}} \times SW_{\text{Education}}^{\text{wave 3}}
\]

\[
= \frac{Pr(A1 | Education)}{Pr(A1 | C0, V1)} \times \frac{Pr(A2 | Education)}{Pr(A2 | C0, A1, V1, V2)} \times \frac{Pr(A3 | Education)}{Pr(A3 | C0, A1, V1, A2, V2)}
\]

(10)

Utilizing the confounding weight, we created a pseudo-population in which there was no association between the covariates and the exposure.

### 2.3.2. Missing values

We utilized Predictive Mean Matching (PMM) as the imputation method to address missing values in the dataset. We chose PMM for its ability to address missing values in the dataset. We chose PMM for its ability to maintain the distribution of the original dataset and minimize the bias. Additionally, PMM is less sensitive to the choice of regression model and is applicable to various data types, including continuous, ordinal, and categorical variables. In this study, we generated 5 imputed datasets with a maximum of 25 iterations for the Expectation-Maximization (EM) algorithm to converge during the imputation process.

### 2.3.3. Marginal structural model

After applying the final weight, the marginal structural model took the form:

\[
E \left[ Y_{a1,a2,a3} \mid Education, X = x, a1^{*}Education, a2^{*}Education, a3^{*}Education \right] =
\]

\[
\mu + \gamma X + \beta_1 a_1 + \beta_2 a_2 + \beta_3 a_3
\]

\[
+ i\beta_{1E} Edu high school + i\beta_{2E} Edu less than college + i\beta_{3E} Edu college
\]

\[
+ i\beta_{4E}(a_1^{*} Edu high school) + i\beta_{5E}(a_1^{*} Edu less than college) + i\beta_{6E}(a_1^{*} Edu college)
\]

\[
+ i\beta_{7E}(a_2^{*} Edu high school) + i\beta_{8E}(a_2^{*} Edu less than college) + i\beta_{9E}(a_2^{*} Edu college)
\]

\[
+ i\beta_{10E}(a_3^{*} Edu high school) + i\beta_{11E}(a_3^{*} Edu less than college) + i\beta_{12E}(a_3^{*} Edu college)
\]

Fig. 1. Directed Acyclic Graph (D.A.G.) for the study.
Note: The analysis was conditional on the Wave 1 MDD; adjustment for baseline and time-varying covariate was done by weighting; measures were changes in the cancer diagnosis history for the change in experiencing MDD. SE = standard error; CI = confidence interval; LL = lower limit; UL = upper limit.

\[ X = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \]

In this model, \( X_{a1+a2+a3} \) was referred to as a "counterfactual outcome", referring to MDD at Wave 3 for an individual that would have occurred under hypothetical cancer diagnoses set at Wave 1, Wave 2, and Wave 3 to \( a_1, a_2 \) and \( a_3 \) respectively. \( X \) denoted as the baseline MDD.

Furthermore, the joint effect of cancer diagnoses at Wave 1, Wave 2, and Wave 3 on MDD at Wave 3 were represented by \( \beta_1, \beta_2, \) and \( \beta_3 \) respectively, corresponding to the change in cancer diagnosis from no to yes. The terms \( \beta_2 \) to \( \beta_3 \) were employed to examine the potential effect modification by education level on a multiplicative scale. The confidence interval of \( \beta_1, \beta_2, \) and \( \beta_3 \) were calculated using the robust variance (Hernan and Robins, n.d.).

For the subgroup analysis, Generalized Estimating Equations (GEE) models were used to examine the association between cancer diagnosis history and MDD. GEE models are an extension of generalized linear models that account for the correlation between repeated measurements taken from the same individuals over time. This approach enables an examination of the association between cancer diagnosis history and MDD while considering the within-subject correlations that may exist due to the repeated measures design of the study. The analysis was conducted using R (version 4.2.2). (12b)

### 3. Results

Table 1 presents the sample characteristics. At baseline, 7108 participants completed the Wave 1 survey. In the sample, most subjects were white (92.14 %). We found that 6.91 % of participants were diagnosed with cancer, 13.25 % with MDD, and 12.01 % experienced sleep problems. A total of 4042 subjects participated in Wave 2, while 3295 subjects participated Wave 3. The changing trends of the percentages of participants with cancer and the trends of smoking behavior, alcohol problems, and BMI, are shown in Supplementary Figs. 1–4. (13a)

### 3.1. Primary outcome: MDD diagnosis

Table 2 presents the results for the effects of cancer diagnosis history on MDD as a binary variable over time. The analysis indicates that only the cancer diagnosis at Wave 2 has a statistically significant effect on MDD at Wave 3, as opposed to the cancer diagnoses at Wave 1 or Wave 3. Specifically, a cancer diagnosis 1 year prior to the MDD assessment was associated with 3.74 times the odds of having MDD compared to those without a cancer diagnosis at Wave 2, with a 95 % CI of 1.41–9.918 and \( p < 0.01 \). However, cancer diagnoses at Wave 1 and Wave 3 did not show a statistically significant association with MDD in the final assessment (Wave 3), with 95 % CI values of 0.170–1.788, \( p = 0.321 \), and 0.105–1.670, \( p = 0.218 \), respectively.

The significance of the interaction term indicates effect modification on a multiplicative scale. Stratification analysis results are shown in Fig. 2. Among individuals with education level of some grade school to some high school, a cancer diagnosis at Wave 2 was significantly associated with 3.45 times the odds of having MDD in Wave 3, with a 95 % CI of 1.31–9.08 and \( p < 0.05 \). However, neither the cancer diagnosis at Wave 1 nor at Wave 3 showed a statistically significant association with sleep problems at Wave 3: OR: 0.57, 95 % CI: 0.18–1.80, \( p = 0.335 \); OR: 0.45, 95 % CI: 0.12–1.72, \( p = 0.241 \). (17)

Similar patterns were observed among individuals with education level of some college but no bachelor’s degree. Only cancer diagnosis at Wave 1 was statistically associated with 1.96 times the odds of MDD at Wave 3 (95 % CI: 1.04–3.67, \( p < 0.05 \)), but not the cancer diagnosis at Wave 2 (95 % CI: 0.35–1.62, \( p = 0.468 \)) or Wave 3 (95 % CI: 0.41–2.15, \( p = 0.893 \)). However, for people who had a GED to high school education level and graduated college or obtained higher degrees, no association was observed between the cancer diagnosis at the 3 waves and MDD at Wave 3. Additional information was shown in Supplementary Table 1.

In the sensitivity analysis, MDD diagnostic scores from the CIDI were alternatively treated as a continuous variable. The results from MSM suggested that akin to considering MDD as a binary variable, only a cancer diagnosis at Wave 2 exerted a statistically significant impact on

<table>
<thead>
<tr>
<th>Parameter</th>
<th>OR</th>
<th>95 % CI</th>
<th>LL</th>
<th>UL</th>
<th>Z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDD at Wave 1</td>
<td>2.147</td>
<td>1.640</td>
<td>2.810</td>
<td>30.883</td>
<td>&lt;0.001***</td>
<td></td>
</tr>
<tr>
<td>Cancer diagnosis at Wave 1</td>
<td>0.551</td>
<td>0.170</td>
<td>1.788</td>
<td>0.986</td>
<td>0.321</td>
<td></td>
</tr>
<tr>
<td>Cancer diagnosis at Wave 2</td>
<td>3.741</td>
<td>1.411</td>
<td>9.918</td>
<td>7.036</td>
<td>&lt;0.01**</td>
<td></td>
</tr>
<tr>
<td>Cancer diagnosis at Wave 3</td>
<td>0.419</td>
<td>0.105</td>
<td>1.670</td>
<td>1.520</td>
<td>0.218</td>
<td></td>
</tr>
<tr>
<td>Education level</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than high school</td>
<td>Reference</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GED to graduated high school</td>
<td>0.691</td>
<td>0.515</td>
<td>0.928</td>
<td>6.018</td>
<td>&lt;0.05*</td>
<td></td>
</tr>
<tr>
<td>Some college (no bachelor's degree)</td>
<td>0.748</td>
<td>0.561</td>
<td>0.996</td>
<td>3.940</td>
<td>&lt;0.05*</td>
<td></td>
</tr>
<tr>
<td>Graduated college and higher</td>
<td>0.653</td>
<td>0.484</td>
<td>0.881</td>
<td>7.756</td>
<td>&lt;0.01**</td>
<td></td>
</tr>
<tr>
<td>Interaction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cancer diagnosis (Wave 1) * Less than high school</td>
<td>Reference</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cancer diagnosis (Wave 1) * GED to graduated high school</td>
<td>2.362</td>
<td>0.629</td>
<td>9.013</td>
<td>1.634</td>
<td>0.201</td>
<td></td>
</tr>
<tr>
<td>Cancer diagnosis (Wave 1) * Some college</td>
<td>3.555</td>
<td>0.938</td>
<td>13.477</td>
<td>3.481</td>
<td>0.062</td>
<td></td>
</tr>
<tr>
<td>Cancer diagnosis (Wave 1) * Graduated college and higher</td>
<td>2.353</td>
<td>0.566</td>
<td>9.779</td>
<td>1.385</td>
<td>0.239</td>
<td></td>
</tr>
<tr>
<td>Cancer diagnosis (Wave 2) * Less than high school</td>
<td>Reference</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cancer diagnosis (Wave 2) * GED to graduated high school</td>
<td>0.344</td>
<td>0.102</td>
<td>1.159</td>
<td>2.965</td>
<td>0.085</td>
<td></td>
</tr>
<tr>
<td>Cancer diagnosis (Wave 2) * Some college</td>
<td>0.202</td>
<td>0.058</td>
<td>0.697</td>
<td>6.402</td>
<td>&lt;0.05*</td>
<td></td>
</tr>
<tr>
<td>Cancer diagnosis (Wave 2) * Graduated college and higher</td>
<td>0.271</td>
<td>0.075</td>
<td>0.980</td>
<td>3.963</td>
<td>&lt;0.05*</td>
<td></td>
</tr>
<tr>
<td>Cancer diagnosis (Wave 3) * Less than high school</td>
<td>Reference</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cancer diagnosis (Wave 3) * GED to graduated high school</td>
<td>1.500</td>
<td>0.278</td>
<td>8.091</td>
<td>0.223</td>
<td>0.637</td>
<td></td>
</tr>
<tr>
<td>Cancer diagnosis (Wave 3) * Some college</td>
<td>2.256</td>
<td>0.450</td>
<td>11.307</td>
<td>0.980</td>
<td>0.322</td>
<td></td>
</tr>
<tr>
<td>Cancer diagnosis (Wave 3) * Graduated college and higher</td>
<td>1.459</td>
<td>0.263</td>
<td>8.998</td>
<td>0.186</td>
<td>0.666</td>
<td></td>
</tr>
</tbody>
</table>
MDD ($\beta = 0.944$, 95 % CI: 0.096–1.793, $p < 0.05$), in contrast to cancer diagnoses at Wave 1 (95 % CI: 1.062–0.182, $p = 0.166$) or Wave 3 (95 % CI: 1.381–0.546, $p = 0.396$). Comprehensive coefficient parameters are included in Supplementary Table 2. (17)

The interaction term’s significance also suggests effect modification on a multiplicative scale. Stratified analysis outcomes are shown in Fig. 3. For individuals with education level ranging from partial grade school to incomplete high school, cancer diagnosis at Wave 2 was significantly correlated with a 0.92-point increase in MDD diagnostic scores in Wave 3, exhibiting a 95 % CI of 0.07–1.77 and $p < 0.05$. Conversely, the association between cancer diagnosis at Wave 1 and MDD diagnostic scores in individuals with some college education lost its significance (95 % CI: 0.044–0.149, $p = 0.122$). Additional information was included in Supplementary Table 3.

### 3.2. Secondary outcome: sleep problems

In the comparable MSM that utilized sleep issues as the binary outcome, there was no significant correlation between the years of cancer diagnosis and the occurrence of MDD. The odds ratios for Wave 1, Wave 2, and Wave 3 were 0.657 (95 % CI: 0.248–1.738; $p = 0.397$), 1.494 (95 % CI: 0.500–0.560; $p = 0.423$), and 1.233 (95 % CI: 0.404–3.766; $p = 0.713$), respectively. Furthermore, no significant interactions were observed between the history of cancer diagnosis and education levels. Supplementary Table 4 provides more detailed information.

### 4. Discussion

Our primary finding underscores that education level serves as a significant effect modifier in the relationship between a history of cancer diagnosis and subsequent MDD. Specifically, individuals with the lowest educational attainment who received a cancer diagnosis in Wave 2 exhibited increased odds of experiencing MDD in Wave 3.

This is consistent with existing literature which has shown that lower education levels can amplify the risk of MDD following a cancer diagnosis (Lopes et al., 2022; Mols et al., 2018). The disparity is possibly due to factors such as limited access to correct medical information, challenges in treatment acceptance, and barriers to accessing and completing treatments among those with lower education. As supported by Bettencourt et al., 2007, individuals with less education often lack the knowledge needed to address their mental health concerns adequately (Bettencourt et al., 2007).

Another essential observation was the strong temporal association between cancer diagnosis and the risk of MDD. Specifically, a cancer diagnosis at Wave 2 substantially elevated the likelihood of MDD at Wave 3. This might be attributed to the heightened emotional turmoil immediately after the diagnosis, aggressive initial treatments, and their associated physical and psychological side-effects (Fann et al., 2008; Pitman et al., 2018). However, as patients progressed in their cancer journey, they possibly developed coping strategies, thereby reducing the MDD risk in subsequent years (Halstead and Fernsler, 1994; Kyngäs et al., 2001; Mols et al., 2018).

To improve the mental health outcomes for cancer patients with lower education, several strategies can be adopted. These include normalizing mental health discussions, enhancing patient education on mental health, and facilitating connections to support networks and primary care settings.

One of the major strengths of this study lies in the prospective assessment of the relationship between cancer and MDD, as well as the risk factors, within a large population-based study. Furthermore, our study utilizes long-term longitudinal analyses to observe individual developmental trends or changes during recovery or relapse of the disease, avoiding the cohort effect of cross-sectional design.

The study faces several limitations. Firstly, the exposure variable relied on self-reported questionnaires, potentially leading to recall bias. Future research may benefit from employing objective measures for this variable. Secondly, the use of more recent data could have offered a
deeper understanding of the association between a history of cancer diagnosis and MDD, especially given the moderating role of education. It’s noteworthy that the upcoming Wave 4 of MIDUS, which will potentially offer more insights, won’t start until 2025. As such, this study was constrained by the available data. Future research is encouraged to utilize the latest data to understand the evolving consequences of cancer diagnosis. Lastly, factors like genetic predispositions, childhood trauma, and family history of cancer, while potentially significant, were not captured in the dataset used for this study. We recommend future research to delve into more extensive datasets that encompass these variables to better address potential confounding effects (Clark et al., 2011; Comijs et al., 2013; Hu et al., 2021; Nguyen and Massagué, 2007; Sullivan et al., 2000). (19)

5. Conclusion

A recent cancer diagnosis was associated with an increased likelihood of experiencing MDD, but not for more distant cancer diagnoses. Cancer survivors with lower education levels were more susceptible to MDD following a recent cancer diagnosis compared to those with higher education levels. Both education level and the timing of cancer diagnosis should be considered when evaluating the mental health of cancer survivors.

**Abbreviations**

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIDUS</td>
<td>Midlife Development in the United States</td>
</tr>
<tr>
<td>MDD</td>
<td>major depressive disorder</td>
</tr>
<tr>
<td>GLOBOCAN</td>
<td>Global Cancer Observatory</td>
</tr>
<tr>
<td>MIDMAC</td>
<td>John D. and Catherine T. MacArthur Foundation Research Network on Successful Midlife Development</td>
</tr>
<tr>
<td>RDD</td>
<td>Random Digit Dialing</td>
</tr>
<tr>
<td>SAQ</td>
<td>Self-Administered Questionnaire</td>
</tr>
<tr>
<td>APA</td>
<td>American Psychiatric Association’s</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organization’s</td>
</tr>
<tr>
<td>CIDI</td>
<td>Composite International Diagnostic Interview</td>
</tr>
<tr>
<td>GED</td>
<td>General Education Diploma</td>
</tr>
<tr>
<td>AST</td>
<td>Alcohol Screening Test</td>
</tr>
<tr>
<td>BMI</td>
<td>Body Mass Index</td>
</tr>
<tr>
<td>DAG</td>
<td>Directed Acyclic Graph</td>
</tr>
<tr>
<td>MSM</td>
<td>Marginal Structural Model</td>
</tr>
<tr>
<td>PMM</td>
<td>Predictive Mean Matching</td>
</tr>
<tr>
<td>EM</td>
<td>Expectation Maximization</td>
</tr>
<tr>
<td>SE</td>
<td>Standard Error</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence Interval</td>
</tr>
<tr>
<td>LL</td>
<td>Lower Limit</td>
</tr>
<tr>
<td>UL</td>
<td>Upper Limit</td>
</tr>
<tr>
<td>CBT</td>
<td>Cognitive Behavioral Therapy</td>
</tr>
</tbody>
</table>


