GMR
The Original

ASSOCIATION BETWEEN DIABETES MELLITUS AND SELF-REPORTED HISTORY OF CANCER AMONG ADULTS: A CROSS-SECTIONAL ANALYSIS OF NHANES DATA

Yamama Farouk Shakir[1], Aljuboori M. Nafea[2], Amash H. A. Alajmi[3], Amer Khlaif Faraj Alajaleen[4], Samer Ayman Mohammad Daradkah[5], Mohamed Yassin Ali Osman[6], Jakline Nelson Saddik[7]

[1]Ibn Sina University of Medical and Pharmaceutical Sciences, Baghdad, Iraq, Email: Yamama.alawssi@ibnsina.edu.iq, ORCID: https://orcid.org/0009-0005-3400-213X
[2]Department of Microbiology, College of Medicine, Ibn Sina University of Medical and Pharmaceutical Sciences, Baghdad, Iraq. Email: aljuboori.mohammed@ibnsina.edu.iq, ORCID: https://orcid.org/0009-0002-7381-6396
[3]School of Medicine, Jordan University of Science and Technology, Irbid, Jordan, Email: ahalajmi19@med.just.edu.jo, ORCID: https://orcid.org/0009-0007-0172-4827
[4]Faculty of Medicine, The University of Jordan, Amman, Jordan, Email: ameralajalen@yahoo.com, ORCID: https://orcid.org/0000-0002-0430-1586
[5]Department of General Surgery, Faculty of Medicine and Surgery, The University of Jordan, Amman, Jordan, Email: drsamer.daradkah@gmail.com, ORCID: https://orcid.org/0009-0003-4185-2648
[6]Lecturer of General and Plastic Surgery, Faculty of Medicine, Fayoum University, Fayoum, Egypt, Email: Yassinsurgeon@gmail.com, ORCID: https://orcid.org/0009-0004-9983-4220
[7]Department of Microbiology and Immunology, Faculty of Pharmacy, Nahda University, Beni Suef, Egypt, Email: jaqueline.nielsen@nub.edu.eg, ORCID: https://orcid.org/0009-0002-7189-709X

ABSTRACT

Background: Diabetes mellitus and cancer may coexist, but their observed relationship may reflect shared metabolic and behavioural risk factors. This study examined the association between diabetes mellitus and self-reported history of any cancer among U.S. adults.

Methods: This cross-sectional analysis used NHANES 2017–2018 data. Adults aged ≥20 years with complete primary-analysis data on self-reported diabetes, self-reported cancer history, body mass index, demographic variables, smoking, and mobile examination center weights were included (n = 4,347). The outcome was report that a health professional had ever diagnosed cancer or malignancy. Survey-weighted logistic regression estimated crude, demographic-adjusted, and fully adjusted odds ratios. A sensitivity analysis used diagnosed diabetes or glycated hemoglobin (HbA1c) ≥6.5% (n = 4,214).

Results: Weighted cancer-history prevalence was 20.1% (95% CI, 15.2%–26.1%) among adults with diabetes and 9.4% (95% CI, 7.6%–11.6%) among adults without diabetes. Diabetes was associated with higher unadjusted odds of cancer history (OR, 2.41; 95% CI, 1.57–3.70; $p < .001$). The estimate attenuated after adjustment for age, sex, and race/ethnicity (OR, 1.24; 95% CI, 0.76–2.02) and remained non-significant after full adjustment (OR, 1.31; 95% CI, 0.80–2.15; $p = .262$). The sensitivity model was also non-significant (OR, 1.34; 95% CI, 0.85–2.09; $p = .189$).

Conclusions: Diabetes was associated with greater unadjusted co-occurrence of self-reported cancer history, but no statistically significant independent cross-sectional association remained after measured confounding was addressed. Findings do not establish temporality, causation, cancer-site-specific associations, or incident cancer risk.

KEYWORDS: diabetes mellitus; cancer history; NHANES; cross-sectional study; survey-weighted logistic regression

1. INTRODUCTION

Diabetes mellitus and cancer are two important chronic conditions with significant public health impacts in the United States. Nationally representative estimates of the prevalence of total diabetes (diagnosed plus undiagnosed) in U.S. adults ranged from 15.8% in August 2021 to 15.8% in August 2023, and cancer statistics have continued to show a high cancer burden in the United States in terms of incident cancers and cancer-related deaths (Gwira et al., 2024; Siegel et al., 2024). Thus, the presence of diabetes and past cancer diagnosis is clinically relevant for prevention, screening, survivorship care and long-term management of chronic diseases (Giovannucci et al., 2010; Gallagher & LeRoith, 2015).

The association of type 2 diabetes with a number of site-specific cancers has been confirmed repeatedly in observational studies, but the quality, direction and causality of these associations is not uniform across cancer sites and study designs (Giovannucci et al., 2010; Tsilidis et al., 2015). Several mechanisms have been suggested to account for this, including hyperglycaemia, insulin resistance, compensatory hyperinsulinaemia, insulin-like growth factor signalling, obesity-related endocrine changes and chronic inflammation, but in addition there are important common risk factors, including age, adiposity, smoking, socioeconomic position, and healthcare contact (Gallagher & LeRoith, 2015; Giovannucci et al., 2010).

Beyond metabolic and endocrine pathways, immune-related mechanisms have also been proposed to contribute to the observed coexistence of diabetes and cancer. Chronic low-grade inflammation altered innate and adaptive immune responses, immune-cell dysfunction, and changes in cytokine signalling have been described in individuals with diabetes and may influence tumour initiation and progression indirectly. In particular, prolonged metabolic dysregulation may contribute to an immunological microenvironment characterised by persistent inflammatory activation and impaired immune surveillance. Nevertheless, these pathways remain complex and are difficult to distinguish from shared behavioural and metabolic determinants in population-based observational studies (Gallagher & LeRoith, 2015; Tsilidis et al., 2015; Mantovani et al., 2008).

The present study is purposefully more limited than an incidence or causal study. NHANES does not include an incident-cancer outcome for this cross-sectional analysis, but rather assesses whether respondents have been told by a professional in the past that they had cancer or a malignancy (National Center for Health Statistics [NCHS], 2020e). Therefore, the study estimates the association of diabetes mellitus and self-reported prevalence of any cancer in adults, not the risk of developing cancer nor the effect of diabetes on the risk of developing cancer at any or specific cancer site and not causal effects of diabetes (NCHS, 2020e; von Elm et al., 2007).

The objectives in this study are to describe the weighted characteristics of adults with and without self-reported cancer history; to estimate the weighted prevalence of self-reported cancer history by diabetes status; to quantify the crude and adjusted survey-weighted association; to determine whether the crude association remains after survey weighted adjustment; and to examine the robustness using a composite diabetes definition that includes glycated haemoglobin (HbA1c) (NCHS, 2020b, 2020c, 2020d, 2020e). It was hypothesized a priori that diabetes would be positively associated with the crude prevalence of self-reported cancer history and that such an association might differ after controlling for common demographic, socioeconomic, anthropometric, and smoking factors (Giovannucci et al., 2010; Tsilidis et al., 2015).

Both prevalence comparison and crude and adjusted sequential ORs are provided allowing one to see both the difference in prevalence at the population level and the difference in prevalence after measured confounding is accounted for. This distinction is particularly relevant when the endpoints are general cancer history (self-reported) and when factors that are associated with both cancer and diabetes are not equally distributed between comparison groups (Giovannucci et al., 2010; Tsilidis et al., 2015; von Elm et al., 2007).

2. LITERATURE REVIEW

2.1 Biological plausibility and shared risk-factor structure

The biological rationale for a diabetes–cancer association is multifactorial. Insulin resistance may lead to compensatory hyperinsulinaemia and altered insulin-like growth factor (IGF) signalling, which can influence cellular proliferation and apoptosis. In parallel, chronic hyperglycaemia—reflected clinically by biomarkers such as glycated haemoglobin (HbA1c)—has been associated with oxidative stress and systemic inflammatory activation. These laboratory-based indicators strengthen the clinical interpretation of diabetes beyond self-reported status and support its role as a metabolically active condition with potential oncogenic relevance (Giovannucci et al., 2010; Gallagher & LeRoith, 2015).

Obesity may further exacerbate insulin resistance and systemic inflammation and is also associated with alterations in adipokines and sex-hormone pathways, thereby linking metabolic dysfunction with tumour-promoting biological environments. Nevertheless, observational evidence alone cannot establish causality, as shared determinants such as age, adiposity, smoking, socioeconomic status, and healthcare utilisation may contribute to the observed associations (Tsilidis et al., 2015). Beyond these biological pathways, reverse causation and surveillance bias remain possible because occult cancer, cancer treatment, or closer healthcare surveillance may influence diabetes recognition and documentation of cancer history (Giovannucci et al., 2010).

Observational data cannot prove causal effects just by the fact that there is a biological plausibility. Age, body size, smoking, socioeconomic status, available preventive services and health care utilisation are common factors which can generate or exacerbate crude associations between diabetes and cancer (Giovannucci et al., 2010; Tsilidis et al., 2015). Occult cancer, cancer treatment, or closer surveillance of other health issues may also be plausible as reverse causation, and surveillance bias may affect documentation of a cancer history and recognition of diabetes (Giovannucci et al., 2010).

2.2 Evidence from cohort studies and evidence syntheses

The evidence is prospective and indicates some general, small, but inconsistent, association between diabetes and cancer overall. Noto et al (2011) conducted a systematic review and meta-analysis of cohort studies and found a pooled adjusted risk ratio of 1.10 (95% CI, 1.04–1.17) for all cancer incidence in adult diabetes population. This finding would be indicative of an epidemiological association, but would not necessarily be interpreted as a uniform effect on all cancer types, nor as an individual patient's causal association (Noto et al., 2011; Giovannucci et al., 2010).

A meta-umbrella of observational meta-analyses revealed that the evidence for the association of type 2 diabetes with cancer was of varying credibility across different cancer sites, most notably liver, pancreas, endometrium, colorectum, breast and bladder, but also kidney (Tsilidis et al., 2015). A subsequent systematic review, which further triangulated observational meta-analysis evidence with Mendelian-randomisation evidence, also highlighted that whilst there was strong observational evidence for some associations with sites, there remained some uncertainties regarding causal interpretation (Pearson-Stuttard

et al., 2021). A meta-analysis of 121 cohorts also found evidence of sex and site of cancer heterogeneity, suggesting that an all-cancer estimate is a sum of different cancer types and is not a causal estimate for each type of cancer (Ohkuma et al., 2018). These evidence syntheses also help to understand the different interpretation that should be given to a broad ‘any cancer' outcome versus site-specific cancer incidence. It may be appropriate to present an all-cancer summary, but this can obscure differences in the aetiology, timing of diagnosis, survival, and confounding across individual cancers (Ohkuma et al., 2018; Pearson-Stuttard et al., 2021; Tsilidis et al., 2015).

2.3 Measurement and design considerations in population surveys

Cancer outcome in this study was based on NHANES item MCQ220 ("Ever told by a doctor or health professional that you had cancer or a malignancy of any kind") for adults 20 years and older (NCHS, 2020e). Whilst self-reported cancer history is useful for population surveillance it is not synonymous with registry confirmed diagnosis – validation studies have revealed a variation in the accuracy of self-reported cancer across populations and cancer type (Bergmann et al., 1998). Therefore, the cancer history in the manuscript was used as self-reported cancer history.

A cross-sectional design is also a source of an additional interpretation limitation in that diabetes and cancer history are measured in the same survey cycle. It cannot determine whether diabetes was a risk factor for cancer or whether cancer or its treatment affected glycaemic status or whether both diabetes and cancer were brought about by common risk factors (von Elm et al., 2007; Giovannucci et al., 2010). This is the right question to ask, rather than whether diabetes causes cancer or increases its risk (NCHS, 2020e; von Elm et al., 2007).

2.4 Study rationale and analytical contribution

NHANES is an ideal population-level association study due to the standardized interview, examination, and laboratory testing procedures, and a complex probability sample of the noninstitutionalized civilian U.S. population (Chen et al., 2020; NCHS, 2020f). The components used here are from the 2017-2018 cycle, which allows for diabetes classification from the interview, sensitivity definition based on HbA1c, measured BMI, and for smoking and key demographic and socioeconomic factors to be included in the analysis (NCHS, 2020a, 2020b, 2020c, 2020d, 2020g).

The study provides a new association analysis, where the exposure is diabetes and self-reported history of any cancer is the outcome. The main sequence of evidence is thus a direct comparison of the weighted prevalence of the conditions and crude, demographic weighted, and fully survey weighted logistic models. This sequence focuses on demonstrating that the association between two factors is not just simple co-occurrence of the factors but the association that remains after the adjustment for the other factors to prevent overstating the evidence for crosssectional data as causal or for cancer site-specific (Giovannucci et al., 2010; Tsilidis et al., 2015; von Elm et al., 2007).

3. METHODOLOGY

3.1 Study design and reporting framework

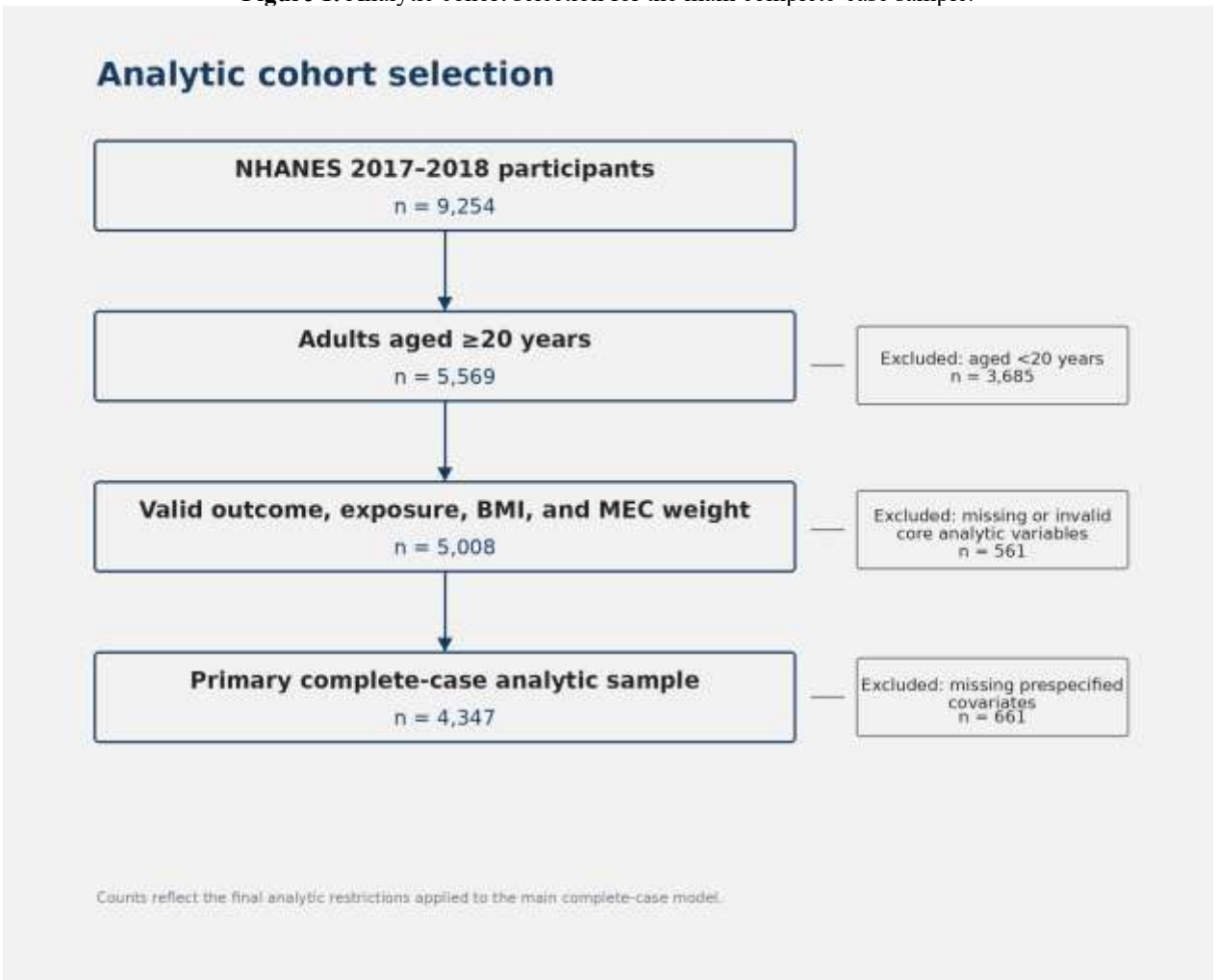
The current study was designed as a cross-sectional study to investigate the relationship between diabetes mellitus and history of cancer (self-report) in adults in the NHANES 2017–2018. NHANES is a multistage probability-based design that is used to obtain estimates for the civilian, non institutionalizedized, U.S. population (National Center for Health Statistics [NCHS], 2020f; Chen et al., 2020). Reporting was designed around the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines for cross sectional studies (von Elm et al., 2007).

3.2 Data source and study population

The NHANES 2017–2018 component files were linked to each other using the unique respondent identifier (SEQN) and made public. Analytic files included in the analytic dataset included the Diabetes (DIQ_J), Body Measures (BMX_J), Glycohemoglobin (GHB_J), Medical Conditions (MCQ_J), Smoking-Cigarette Use (SMQ_J), and Demographic Variables and Sample Weights (DEMO_J) files (NCHS, 2020a, 2020b, 2020c, 2020d, 2020e, 2020g). To be eligible, participants had to be 20 years or older and have complete information on self-reported diabetes, self-reported cancer history, BMI, and survey-design variables as well as the weight from the mobile examination center (MEC) examination. Complete-case approach was used for prespecified covariates in the primary model. The sensitivity analysis using HbA1c information used the composite diabetes classification as outlined below and only included 4214 adults with complete covariate and definitive composite classification.

The main analytic sample, as summarised in Figure 1, was constructed. Of 9,254 participants in the 2017–2018 cycle, 5,569 were aged 20 years or older. Using the eligibility and complete-case criteria, 4,347 adults were analyzed. After applying the sensitivity analysis criteria of the HbA1c, it yielded 4,214 adults.

Figure 1. Analytic cohort selection for the main complete-case sample.



Note. MEC, mobile examination center; BMI, body mass index. The flow diagram reports the analytic restrictions used for the main complete-case model.

3.3 Outcome definition

The outcome was self-reported history of cancer or malignancy. NHANES respondents age 20 and older were asked if anyone ever told them they had cancer or a malignancy (MCQ220) which was coded as a yes or no response (NCHS, 2020e). The outcome included history of any cancer, and was not categorized for cancer site, stage, time of diagnosis, treatment or survival.

3.4 Exposure definition

Exposures most commonly reported were diabetes mellitus. Diabetes was determined as having answered "yes" to the question: Has a doctor or health professional ever told you that you had diabetes or sugar diabetes (DIQ010)? Answers of "no" were considered to be a lack of diabetes. For the primary comparison, those who reported borderline diabetes, or responded with missing, refused, or do not know were not included (NCHS, 2020c).

A composite Diabetes definition was used in the prespecified sensitivity analysis. People who self-identified as having diabetes and those with glycated hemoglobin (HbA1c) ≥6.5% were considered to have diabetes. Participants who did not report having diabetes and had an HbA1c less than 6.5% (American Diabetes Association Professional Practice Committee, 2026; NCHS, 2020d) were considered to be without diabetes. Participants who indicated that they had diabetes were considered as having diabetes, participants who indicated they did not have diabetes needed to have the HbA1c data to be considered as no diabetes. For this sensitivity analysis, participants who did not have a definite composite classification were excluded from the analyses. To enhance clinical validity, a laboratory-based glycaemic biomarker (HbA1c) was incorporated in sensitivity analyses. Participants with HbA1c ≥6.5% were classified as having diabetes, consistent with established diagnostic criteria (American Diabetes Association Professional Practice Committee, 2026; NCHS, 2020d). This approach integrates clinical diagnosis with objective biochemical measurement, improving phenotype precision in a population-based setting.

From a pathophysiological perspective, HbA1c also reflects chronic glycaemic exposure, which has been linked to oxidative stress and systemic inflammatory activation. These processes represent overlapping metabolic–inflammatory pathways that may be relevant to chronic disease co-occurrence, including cancer-related outcomes.

3.5 Covariates

These prespecified covariates included sex, race/ethnicity, educational attainment, family poverty-income ratio (PIR), BMI category and cigarette-smoking status (continuous, modelled per 10-year increase). In the Body Measures component, BMI was calculated using actual height and weight measurements. Lifetime cigarette use and current-use responses (NCHS, 2020a, 2020g) were used to construct smoking status. Education level was grouped into four categories: < high school, high school/GED, some college/associate degree and college graduate; PIR was also grouped into three categories: <1.30, 1.30 to <3.50, and ≥3.50; BMI was grouped into three categories: <25, 25 to <30 and ≥30 kg/m²; and smoking status was grouped into three categories: never smoker, former smoker or current smoker.

3.6 Statistical analysis

All analyses accounted for the complex NHANES survey design using appropriate sampling weights (WTMEC2YR), strata (SDMVSTRA), and primary sampling units (SDMVPSU), with variance estimation performed using Taylor-series linearization (Chen et al., 2020; NCHS, 2020b–2020d). Weighted descriptive statistics and survey-adjusted logistic regression models were applied.

Survey-weighted logistic regression was used to estimate odds ratios (ORs) and 95% confidence intervals (CIs) for the association between diabetes and self-reported cancer history.

The modelling strategy included:

- Crude model: diabetes only
- Demographic-adjusted model: diabetes + age + sex + race/ethnicity
- Fully adjusted model: diabetes + demographic + socioeconomic + lifestyle covariates (BMI, smoking, education, PIR)

The survey design degrees of freedom (30 PSUs – 15 strata = 15) were used to calculate the confidence intervals and conduct the hypothesis testing. All tests were performed two-sided and a p value of < .05 was considered statistically significant. All the analyses were repeated using the composite diabetes definition in the sensitivity model, which included 4,214 participants. To strengthen the clinical–laboratory interpretation, HbA1c-based diabetes classification was used in sensitivity analyses. In addition, HbA1c was considered not only as a categorical diagnostic marker but also as a continuous biomarker proxy of long-term glycaemic exposure.

Given emerging evidence that chronic metabolic dysregulation may interact with systemic inflammatory and immune-related pathways, HbA1c was interpreted as an indirect marker of metabolic–inflammatory burden rather than a direct immune biomarker. However, no direct immunological assays (e.g., cytokines or immune cell profiling) were available within NHANES 2017–2018; therefore, immune-related mechanisms were considered only at the interpretative level.

In this cross-sectional study, the odds ratios reported are associative measures of the odds of self-reported prevalent cancer. The models that were staged were designed to illustrate the shift in the association between diabetes and the estimate when plausible shared determinants were added; this approach does not help identify mediation pathways, eliminate residual confounding, or establish temporality (Giovannucci et al., 2010; von Elm et al., 2007).

3.7 Ethics, data availability, and interpretive scope

The NHANES public-use data are de-identified and available for public. This is a secondary analysis where there were no direct contacts with the participants. The study is a cross-sectional design with self-reported cancer history, which does not allow for temporality or causality to be determined (NCHS, 2020e, 2020f).

4. RESULTS

4.1 Characteristics of the analytic sample

The characteristics of the 4,347 adults in the main analytic sample, including 453 adults who reported a history of cancer, are summarised by weights in Table 1. The weighted prevalence of having ever been diagnosed with cancer was 10.7% and the weighted prevalence of having ever been diagnosed with diabetes was 11.7%. The weighted mean age was higher among the participants with cancer history compared with the participants without cancer history (63.6 vs. 46.1 years). Significant differences were also noted in terms of higher weighted adult (≥60 years old) and adult with diabetes (22.0% vs. 10.5%).

Table 1. Weighted characteristics of adults in the main analytic sample, overall and by self-reported cancer history.

Characteristic	Overall N = 4,347	No cancer history n = 3,894	Cancer history n = 453
Age, years, weighted mean	48.0	46.1	63.6
BMI, kg/m², weighted mean	29.8	29.8	29.6
Sex			
Male	2,102 (48.2)	1,888 (48.9)	214 (42.4)
Female	2,245 (51.8)	2,006 (51.1)	239 (57.6)

Characteristic	Overall N = 4,347	No cancer history n = 3,894	Cancer history n = 453
Age group, years			
20–39	1,356 (37.1)	1,325 (40.3)	31 (10.0)
40–59	1,381 (34.5)	1,296 (36.3)	85 (19.4)
≥60	1,610 (28.4)	1,273 (23.4)	337 (70.6)
Race/ethnicity			
Mexican American	551 (8.3)	515 (8.9)	36 (3.5)
Other Hispanic	381 (6.2)	354 (6.6)	27 (3.3)
Non-Hispanic White	1,600 (64.4)	1,331 (62.3)	269 (82.1)
Non-Hispanic Black	977 (10.7)	907 (11.4)	70 (5.4)
Non-Hispanic Asian	608 (5.4)	581 (5.9)	27 (2.0)
Other/multiracial	230 (4.9)	206 (5.1)	24 (3.6)
Educational attainment			
<High school	811 (10.5)	736 (10.7)	75 (8.5)
High school/GED	1,050 (27.2)	942 (27.6)	108 (23.8)
Some college/AA	1,437 (31.3)	1,278 (31.4)	159 (30.9)
College graduate	1,049 (31.0)	938 (30.3)	111 (36.8)
Family poverty-income ratio			
<1.30	1,253 (20.2)	1,145 (21.2)	108 (12.5)
1.30–<3.50	1,791 (35.9)	1,599 (36.1)	192 (33.9)
≥3.50	1,303 (43.9)	1,150 (42.7)	153 (53.6)
BMI category			
<25 kg/m²	1,141 (25.9)	1,034 (26.4)	107 (21.8)
25–<30 kg/m²	1,385 (31.2)	1,227 (30.7)	158 (35.3)
≥30 kg/m²	1,821 (42.9)	1,633 (42.9)	188 (42.9)
Smoking status			
Never	2,486 (57.2)	2,268 (58.0)	218 (50.6)
Former	1,063 (25.2)	894 (23.9)	169 (35.5)
Current	798 (17.6)	732 (18.0)	66 (14.0)
Self-reported diabetes			
No	3,642 (88.3)	3,310 (89.5)	332 (78.0)
Yes	705 (11.7)	584 (10.5)	121 (22.0)

Note. Values are unweighted n (weighted %) unless otherwise indicated. Age and BMI are weighted means. Percentages are column percentages estimated using WTMEC2YR. BMI, body mass index; GED, General Educational Development; AA, associate degree; PIR, poverty-income ratio.

4.2 Direct unadjusted association between diabetes and self-reported cancer history

The direct association or unadjusted association per the study title is shown in Table 2. Among adults without self-reported diabetes, the weighted prevalence of self-reported cancer history was 9.4% (95% CI, 7.6%–11.6%) and among adults with self-reported diabetes, the weighted prevalence was 20.1% (95% CI, 15.2%–26.1%). Higher odds of self-reported cancer history (crude OR, 2.41; 95% CI, 1.57–3.70; p<.001) were found for those who self-reported diabetes in the diabetes-only survey-weighted model.

The difference in prevalence was 10.7 absolute weighted prevalence points. This unadjusted contrast indicates the relationship between diabetes and a previous self-reported cancer diagnosis that were observed in a population at one survey time point

and is not an estimate of the incidence of cancer, relative risk, or a causal effect of diabetes (NCHS, 2020e; von Elm et al., 2007).

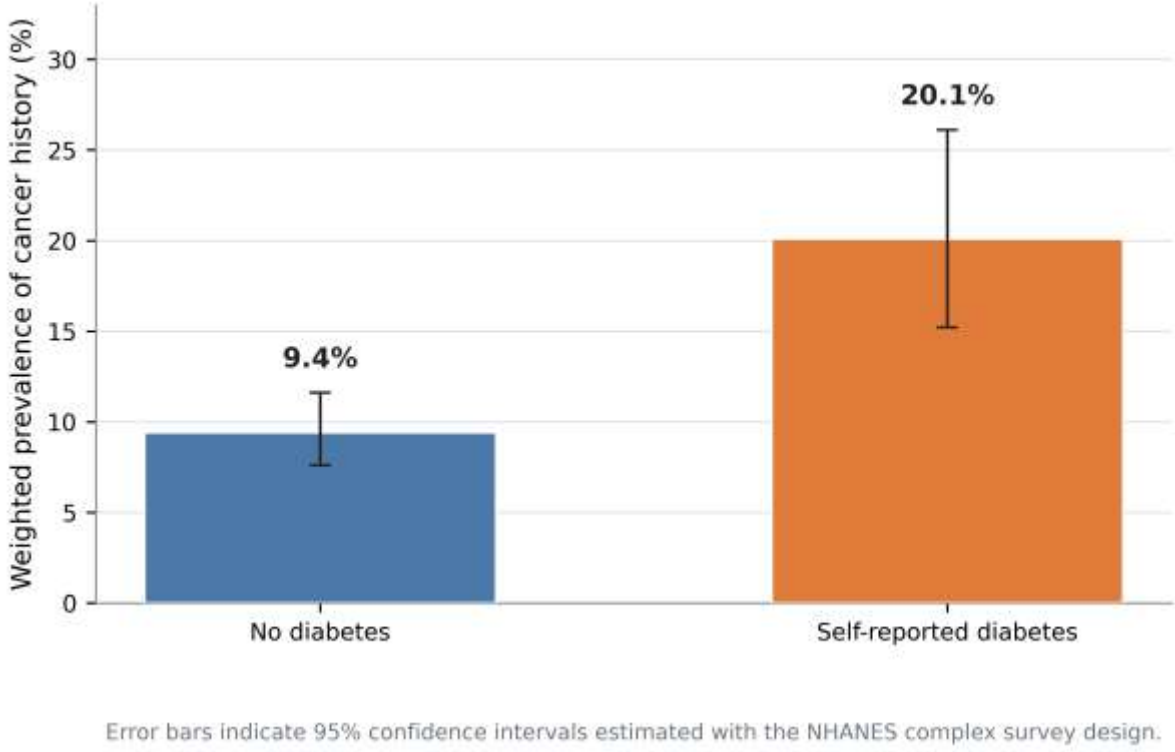
Table 2. Direct unadjusted association between self-reported diabetes and self-reported cancer history.

Diabetes status	Unweighted sample, n	Cancer history, n	Estimated population, million	Weighted prevalence of cancer history, % (95% CI)	Crude OR (95% CI)
No diabetes	3,642	332	180.9	9.4 (7.6–11.6)	Reference
Self-reported diabetes	705	121	24.0	20.1 (15.2–26.1)	2.41 (1.57–3.70)

Note. Prevalence estimates, 95% CIs, population totals, and crude ORs account for WTMEC2YR, SDMVSTRA, and SDMVPSU. Crude ORs were estimated from a diabetes-only survey-weighted logistic-regression model; no diabetes is the reference group. Cancer history is self-report of ever being told by a health professional that the participant had cancer or a malignancy.

Figure 2 visualises the direct weighted prevalence contrast in Table 2. This comparison is intentionally unadjusted and should be interpreted as co-occurrence rather than as an independent or causal association.

Figure 2. Survey-weighted prevalence of self-reported cancer history by self-reported diabetes status.



Note. Error bars indicate 95% confidence intervals estimated using the NHANES complex survey design.

4.3 Adjusted association between diabetes and self-reported cancer history

The sequential survey-weighted logistic-regression analysis is shown in Table 3. The unadjusted odds of self-reported history of cancer were higher in adults reporting diabetes than in adults without diabetes (OR 2.41, 95% CI 1.57 to 3.70; p< .001). After adjustment for age, sex, and race/ethnicity, the estimate was attenuated (OR, 1.24; 95% CI, 0.76–2.02; p=.371). In the fully adjusted model (including education, poverty-income ratio, BMI category, and smoking status), diabetes was no longer associated with self-reported cancer history (OR, 1.31; 95% CI, 0.80-2.15; p=.262).

This high level of attenuation after demographic adjustment was also consistent with the strong age pattern exhibited in Table 1, but this was not the intent of the sequential models to assign the adjustment in the diabetes estimate to a specific covariate. The fully adjusted 95% CI (0.80–2.15) contained the null and values consistent with either a small inverse or positive

relationship. So the non-significant p value was not interpreted as an absence of the association of diabetes with cancer outcome, but rather as evidence that this association was not independent in this analysis.

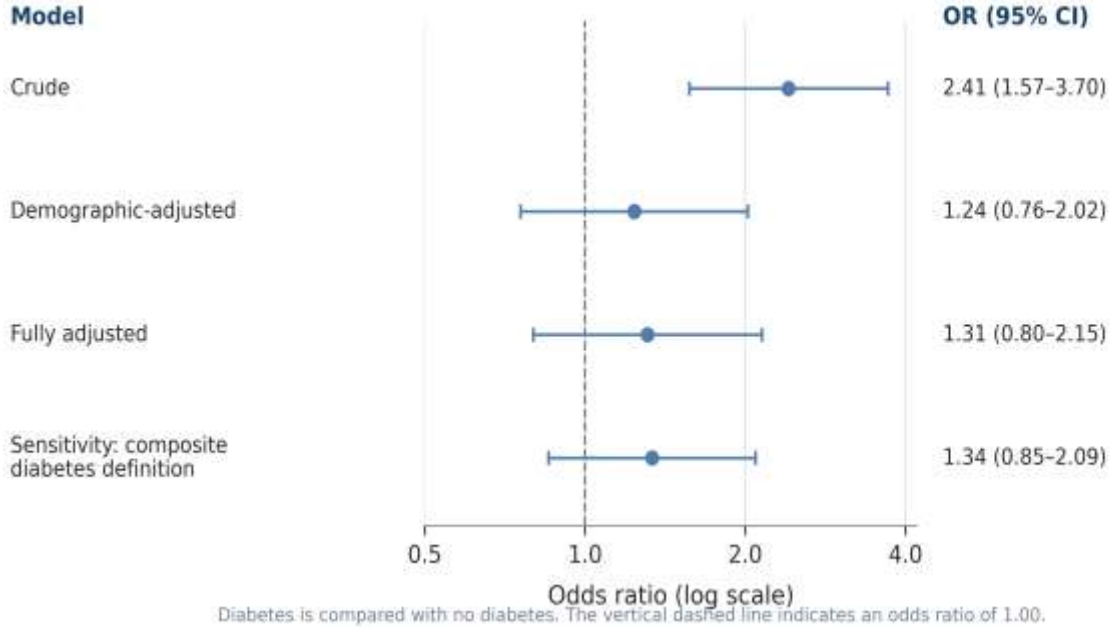
Table 3. Association between diabetes and self-reported cancer history across survey-weighted logistic regression models.

Model	Analytic n	Adjustment set	OR	95% CI	p value
Crude model	4,347	Unadjusted	2.41	1.57–3.70	<.001
Demographic-adjusted model	4,347	Age, sex, race/ethnicity	1.24	0.76–2.02	.371
Fully adjusted model	4,347	Age, sex, race/ethnicity, education, PIR, BMI, smoking	1.31	0.80–2.15	.262
Sensitivity: composite diabetes definition	4,214	Fully adjusted; diabetes = self-reported diagnosis or HbA1c ≥6.5%	1.34	0.85–2.09	.189

Note. ORs compare diabetes with no diabetes. The demographic-adjusted model includes age, sex, and race/ethnicity. The fully adjusted model additionally includes education, PIR, BMI category, and smoking status. CIs and p values use the NHANES survey-design degrees of freedom (15). The composite sensitivity analysis included 4,214 participants with complete covariates and a definitive composite diabetes classification. HbA1c, glycated hemoglobin; PIR, poverty-income ratio; BMI, body mass index.

Figure 3 shows the diabetes estimates from the sequential models and from the sensitivity analysis informed by the HbA1c data. An unadjusted strong positive association was observed between presence of self-reported history of cancer and diabetes in adults, but this association was reduced and was not statistically significant when the models were fully adjusted for demographic factors. The composite-diabetes sensitivity estimate was also not significant. This pattern is consistent with the role of common risk-factor structures, including the large age gap between adults with and without cancer history within this sample, but the direction or mechanism is not determined (Giovannucci et al., 2010; Tsilidis et al., 2015).

Figure 3. Survey-weighted odds ratios for self-reported cancer history associated with diabetes across sequential models.



Note. Points indicate odds ratios and horizontal bars indicate 95% confidence intervals. The dashed line marks an odds ratio of 1.00.

4.4 Sensitivity analysis and primary interpretation

Among the 4,214 adults eligible for the composite diabetes definition, the fully adjusted association was not significant when diabetes was defined by self-reported diagnosis (OR, 1.34; 95% CI, 0.85 to 2.09; p = .189) or HbA1c ≥6.5% (OR, 1.35; 95% CI, 0.83 to 2.19; p = .220). So there was no change in the main conclusion when the wider definition of diabetes was employed. Overall, the findings suggest that the increased co-occurrence of diabetes and the self-reported history of cancer observed may be attributable to more general demographic, cardiometabolic and behavioural characteristics than diabetes alone. As the data are cross-sectional and the outcome is self-reported, the results should not be interpreted as evidence for a causal relationship

between diabetes and cancer, or as evidence of diabetes predicting the risk of new cancer incidence (NCHS, 2020e; von Elm et al., 2007).

5. DISCUSSION

5.1 Principal findings and interpretation

The results of this nationally representative cross-sectional analysis revealed a clear unadjusted co-occurrence of diabetes and self-reported history of cancer. The weighted prevalence contrast was 20.1% vs 9.4% whereas the OR was 2.41 (95% CI 1.57 to 3.70) for adults reporting diabetes and reporting a history of cancer compared with those without diabetes. But this odds ratio decreased to 1.24 after adjustment for the demographic variables and was 1.31 (95% CI, 0.80–2.15) in the fully adjusted model. Similar overall inference (OR: 1.34; 95% CI: 0.85–2.09) was obtained with the sensitivity analysis that was performed using the HbA1c. Therefore, the primary finding is increased crude co-occurrence; and the cross-sectional association between these two conditions is not statistically significant when considering the measured covariates.

The large demographic adjustment reduction indicates that different population composition plays a key role in the difference. The adults with a self-reported history of cancer in this sample were significantly older than adults without a self-reported history of cancer, and there is a strong association between age and diabetes, as well as the cumulative incidence to age of a cancer diagnosis. This relationship is similar to the one observed with crude diabetes–cancer relationships, where common determinants such as age, adiposity, smoking, socioeconomic circumstances and healthcare contact (Giovannucci et al., 2010; Tsilidis et al., 2015) may affect the results. The present estimates cannot tell if diabetes occurred before cancer, after cancer or cancer treatment, or if it simply coincided with cancer history, by means of common causes (NCHS, 2020e; von Elm et al., 2007).

It should also be noted that the fully adjusted estimate is very accurate. The confidence interval of this was sufficiently broad that it included the null and possible meaningful relationships in either direction. Thus, it can be concluded that this cross-sectional sample failed to show an independent association, not that a diabetes–cancer association has been disproved. Longitudinal data with validated site-specific cancer incidence and subsequent, repeated diabetes measurements are needed to determine the relationships more precisely (NCHS, 2020e; von Elm et al., 2007).

5.2 Comparison with previous evidence

The crude association observed broadly agrees with the prospective evidence that diabetes is related to a slightly increased average incidence of all cancer types, but which does not measure the same effect as the current study. Noto et al. (2011) calculated the pooled adjusted risk ratio (ARR) for all-cancer incidence from cohort studies to be 1.10, while in a larger meta-analysis, ARRs varied by sex and cancer site (Ohkuma et al., 2018). Similarly, umbrella reviews have found evidence for a number of site-specific associations, and highlighted the heterogeneity and uncertainty about causality (Pearson-Stuttard et al., 2021; Tsilidis et al., 2015).

Comparing numbers directly should be taken with caution, however. All previous prospective studies focused on incidence of cancers or cancer mortality while this study relied on an adult respondent's report of ever being told by a health professional that they had any cancer or malignancy. The NHANES outcome, therefore, combines various cancer sites, time of diagnosis, and cancer survivorship experiences, and is not able to determine if diabetes is associated with specific cancers (NCHS, 2020e). Thus, the attenuation here does not suggest there would not be an independent association between PSB and the more general outcome of prevalent self-reported cancer history at any of the sites, but rather that such an association was not observed in this cross-sectional sample.

5.3 Biological interpretation and study strengths

Biological mechanisms suggested in the literature include insulin resistance, compensatory hyperinsulinaemia, insulin like growth factor (IGF) signalling, chronic hyperglycaemia, obesity-related hormonal change and inflammation (Gallagher & LeRoith, 2015; Giovannucci et al., 2010). These pathways are not amenable to the present analysis, and should not be treated as a mechanistic analysis. Rather, the reduction in the estimate of the diabetes after covariate adjustment highlights the need to distinguish between biological hypotheses and the impact of shared structures of risk factors in observational studies (Pearson-Stuttard et al., 2021; Tsilidis et al., 2015).

There are a number of strengths to the study. NHANES involves a standardised interview, physical examination, laboratory testing and probability sample that represent the noninstitutionalized civilian U.S. population (Chen et al., 2020; NCHS, 2020f). The use of measured BMI, prespecified covariates, MEC weights, strata and Psu's led to a better alignment of the analysis with the survey design. The fully-adjusted non-significant estimate (OR, 1.34; 95% CI, 0.85-2.09) was similar when the sensitivity analysis was performed using the HbA1c-informed diabetes definition and 4,214 adults with a definitive composite classification, confirming the principal finding of this study was not dependent on the narrower self-reported diabetes definition (American Diabetes Association Professional Practice Committee, 2026, NCHS, 2020d).

5.4 Public-health interpretation

This discovery of increased co-occurrence of crude is still pertinent from a public health perspective. Shared age-related, metabolic and behavioural risk factors within the same adult population might contribute to a clustering of diabetes and self-reported history of cancer. Diabetes status should thus not be used as a stand-alone measure of cancer burden in routine clinical

and preventive care. A comprehensive evaluation, which takes adiposity, smoking habits, metabolic health, the demographic context and other pertinent risk factors into account, may be more enlightening in identifying adults that would benefit from preventive measures and the proper health monitoring. This fits with the shared-risk-factor theory outlined in the diabetes–cancer literature, and the current analysis is not able to calculate the impact of any one strategy on cancer incidence or inform decision-making on cancer screening based on diabetes status alone (Gallagher & LeRoith, 2015; Giovannucci et al., 2010; Tsilidis et al., 2015).

5.5 Limitations and future research

There are some caveats that need to be taken with a grain of salt. First, self-reported cancer history is not the same as cancer history reported on cancer registers, and may differ according to population and the type of cancer (Bergmann et al., 1998). Second, MCQ220 captures any previous cancer or malignancy but does not give any information about the cancer site, stage of cancer, date of diagnosis, treatment, survival or recurrence (NCHS, 2020e). Hence, the evaluation of pancreatic cancer or any other specific cancer site is not done in this manuscript.

Third, the cross-sectional design does not allow determining temporality and the issue of selection due to survival remains unsolved (Giovannucci et al., 2010; von Elm et al., 2007). Fourth, it is important to note that there remains the possibility of residual confounding, since information on medication use, detailed diet information, physical activity, family cancer history, cancer treatment and utilization of healthcare was not taken into account in the primary model (Giovannucci et al., 2010; Tsilidis et al., 2015). Lastly, when missingness is related to exposure, missingness is related to outcome, or missingness is related to any of the other factors, precision and generalizability of the complete-case design might be affected (Chen et al., 2020). To better understand the timing and site-specific associations, prospective cohorts or linked EHR and cancer-registry datasets with clinically confirmed incident cancer and site-specific outcomes, along with repeated diabetes assessments are required.

6. CONCLUSION

Among the adults in this survey-weighted analysis, diabetes was linked to increased unadjusted prevalence of self-reported cancer history, and increased crude odds of cancer history. This association was greatly diminished after taking age, sex, race/ethnicity, socioeconomic characteristics, BMI and smoking into account, and was not statistically significant at the fully adjusted level (OR, 1.31; 95% CI, 0.80–2.15; p = .262). The HbA1c-informed sensitivity analysis yielded a consistent result (OR, 1.34; 95% CI, 0.85–2.09; p = .189). These results indicate a trend towards higher rates of unadjusted co-occurrence of diabetes and self-reported cancer history but did not show a statistically significant independent association between the two at cross-section, after considering measured factors. Thus, the increased cancer-history burden seen in adults with diabetes might be because they share some similar demographic and cardiometabolic features. From a clinical perspective, a thorough evaluation of metabolic, lifestyle, and demographic factors may be more useful than to look at diabetes as a stand-alone parameter of cancer burden. The results do not provide information on temporality, causation, associations with cancer sites or on cancer incidence risk. The relationship between diabetes and incident cancer is yet to be understood from the temporal perspective and prospective studies with clinically established incident cancer outcome and repeated diabetes measures are warranted.

REFERENCES

1. American Diabetes Association Professional Practice Committee. (2026). 2. Diagnosis and classification of diabetes: Standards of Care in Diabetes—2026. Diabetes Care, 49(Supplement 1), S27–S49. https://doi.org/10.2337/dc26-S002
2. Bergmann, M. M., Calle, E. E., Mervis, C. A., Miracle-McMahill, H. L., Thun, M. J., & Heath, C. W., Jr. (1998). Validity of self-reported cancers in a prospective cohort study in comparison with data from state cancer registries. American Journal of Epidemiology, 147(6), 556–562. https://doi.org/10.1093/oxfordjournals.aje.a009487
3. Chen, T.-C., Clark, J., Riddles, M. K., Mohadjer, L. K., & Fakhouri, T. H. I. (2020). National Health and Nutrition Examination Survey, 2015–2018: Sample design and estimation procedures (Vital and Health Statistics Series 2, No. 184). National Center for Health Statistics. https://www.cdc.gov/nchs/data/series/sr_02/sr02-184-508.pdf
4. Gallagher, E. J., & LeRoith, D. (2015). Obesity and diabetes: The increased risk of cancer and cancer-related mortality. Physiological Reviews, 95(3), 727–748. https://doi.org/10.1152/physrev.00030.2014
5. Giovannucci, E., Harlan, D. M., Archer, M. C., Bergenstal, R. M., Gapstur, S. M., Habel, L. A., Pollak, M., Regensteiner, J. G., & Yee, D. (2010). Diabetes and cancer: A consensus report. Diabetes Care, 33(7), 1674–1685. https://doi.org/10.2337/dc10-0666
6. Gwira, J. A., Fryar, C. D., & Gu, Q. (2024). Prevalence of total, diagnosed, and undiagnosed diabetes in adults: United States, August 2021–August 2023 (NCHS Data Brief No. 516). National Center for Health Statistics. https://www.cdc.gov/nchs/products/databriefs/db516.htm
7. Mantovani, A., Allavena, P., Sica, A., & Balkwill, F. (2008). Cancer-related inflammation. Nature, 454(7203), 436–444. https://doi.org/10.1038/nature07205
8. National Center for Health Statistics. (2020a). 2017–2018 data documentation, codebook, and frequencies: Body measures (BMX_J). Centers for Disease Control and Prevention. https://wwwn.cdc.gov/Nchs/Data/Nhanes/Public/2017/DataFiles/BMX_J.htm

9. National Center for Health Statistics. (2020b). 2017–2018 data documentation, codebook, and frequencies: Demographic variables and sample weights (DEMO_J). Centers for Disease Control and Prevention. https://wwwn.cdc.gov/Nchs/Data/Nhanes/Public/2017/DataFiles/DEMO_J.htm
10. National Center for Health Statistics. (2020c). 2017–2018 data documentation, codebook, and frequencies: Diabetes (DIQ_J). Centers for Disease Control and Prevention. https://wwwn.cdc.gov/Nchs/Data/Nhanes/Public/2017/DataFiles/DIQ_J.htm
11. National Center for Health Statistics. (2020d). 2017–2018 data documentation, codebook, and frequencies: Glycohemoglobin (GHB_J). Centers for Disease Control and Prevention. https://wwwn.cdc.gov/Nchs/Data/Nhanes/Public/2017/DataFiles/GHB_J.htm
12. National Center for Health Statistics. (2020e). 2017–2018 data documentation, codebook, and frequencies: Medical conditions (MCQ_J). Centers for Disease Control and Prevention. https://wwwn.cdc.gov/Nchs/Data/Nhanes/Public/2017/DataFiles/MCQ_J.htm
13. National Center for Health Statistics. (2020f). NHANES 2017–2018 overview. Centers for Disease Control and Prevention. https://wwwn.cdc.gov/nchs/nhanes/continuousnhanes/overview.aspx?BeginYear=2017
14. National Center for Health Statistics. (2020g). 2017–2018 data documentation, codebook, and frequencies: Smoking—cigarette use (SMQ_J). Centers for Disease Control and Prevention. https://wwwn.cdc.gov/Nchs/Data/Nhanes/Public/2017/DataFiles/SMQ_J.htm
15. Noto, H., Tsujimoto, T., Sasazuki, T., & Noda, M. (2011). Significantly increased risk of cancer in patients with diabetes mellitus: A systematic review and meta-analysis. Endocrine Practice, 17(4), 616–628. https://doi.org/10.4158/EP10357.RA
16. Ohkuma, T., Peters, S. A. E., & Woodward, M. (2018). Sex differences in the association between diabetes and cancer: A systematic review and meta-analysis of 121 cohorts including 20 million individuals and one million events. Diabetologia, 61(10), 2140–2154. https://doi.org/10.1007/s00125-018-4664-5
17. Pearson-Stuttard, J., Papadimitriou, N., Markozannes, G., Cividini, S., Kakourou, A., Gill, D., Rizos, E. C., Monori, G., Ward, H. A., Kyrgiou, M., Gunter, M. J., & Tsilidis, K. K. (2021). Type 2 diabetes and cancer: An umbrella review of observational and Mendelian randomization studies. Cancer Epidemiology, Biomarkers & Prevention, 30(6), 1218–1228. https://doi.org/10.1158/1055-9965.EPI-20-1245
18. Siegel, R. L., Giaquinto, A. N., & Jemal, A. (2024). Cancer statistics, 2024. CA: A Cancer Journal for Clinicians, 74(1), 12–49. https://doi.org/10.3322/caac.21820
19. Tsilidis, K. K., Kasimis, J. C., Lopez, D. S., Ntzani, E. E., & Ioannidis, J. P. A. (2015). Type 2 diabetes and cancer: Umbrella review of meta-analyses of observational studies. BMJ, 350, g7607. https://doi.org/10.1136/bmj.g7607
20. von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., & Vandenbroucke, J. P. (2007). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. PLoS Medicine, 4(10), e296. https://doi.org/10.1371/journal.pmed.0040296

Supplementary Materials

The supplementary materials provide detailed covariate estimates and a clustered presentation of selected Table 1 characteristics. They are provided for transparency and should be interpreted as supporting context; the principal evidence for the study title is contained in Tables 2–3 and Figures 2–3.

Supplementary Table S1

Fully adjusted survey-weighted logistic-regression estimates for self-reported cancer history.

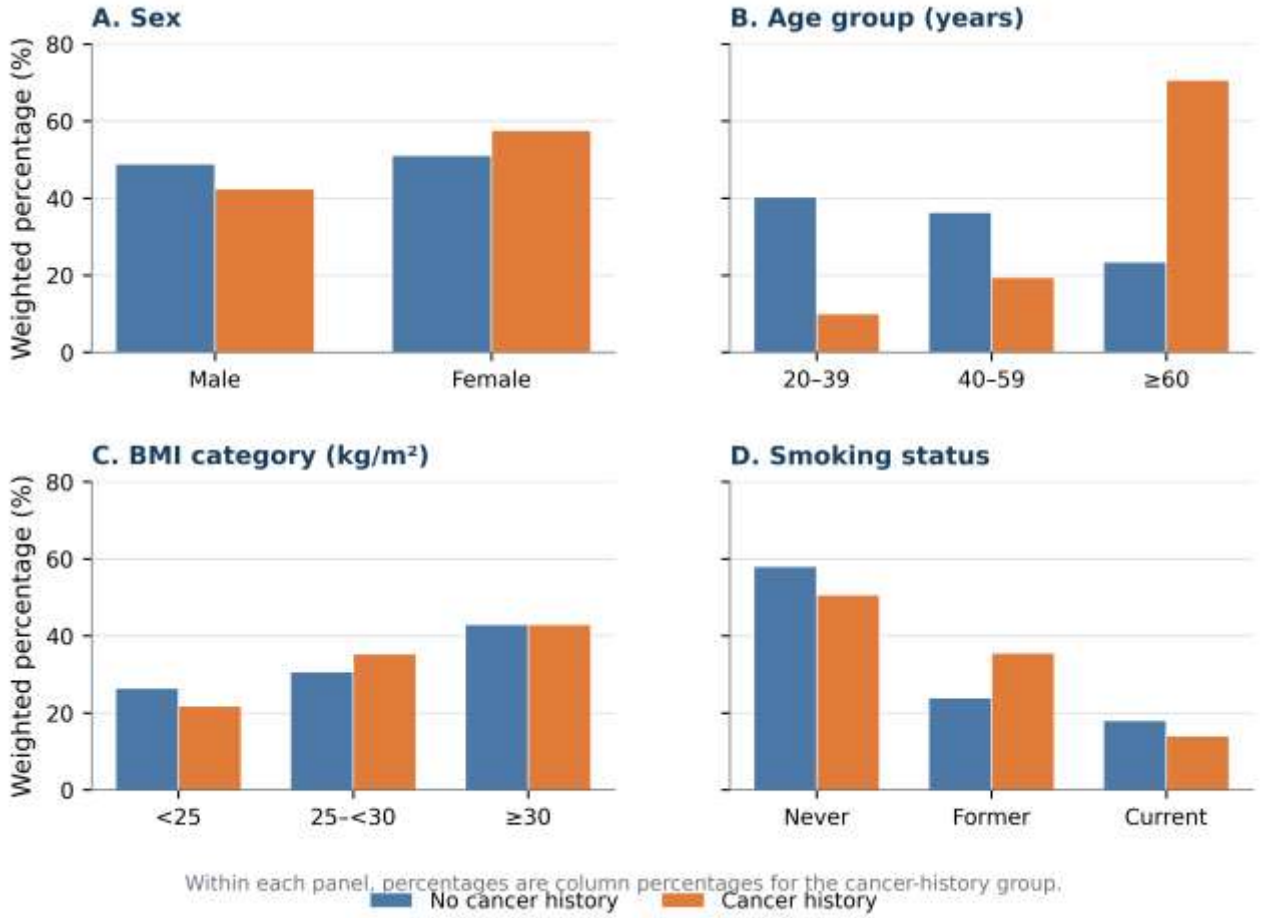
Variable	Adjusted OR	95% CI	p value
Self-reported diabetes (yes vs. no)	1.31	0.80–2.15	.262
Age (per 10-year increase)	1.95	1.69–2.26	<.001
Female (vs. male)	1.37	0.98–1.91	.067
Mexican American (vs. non-Hispanic White)	0.63	0.39–1.01	.054
Non-Hispanic Asian (vs. non-Hispanic White)	0.34	0.20–0.58	<.001
Non-Hispanic Black (vs. non-Hispanic White)	0.49	0.38–0.62	<.001
Other Hispanic (vs. non-Hispanic White)	0.61	0.24–1.52	.267
Other/multiracial (vs. non-Hispanic White)	0.79	0.26–2.38	.660
<High school (vs. college graduate)	0.63	0.39–1.02	.059
High school/GED (vs. college graduate)	0.66	0.36–1.21	.166
Some college/AA (vs. college graduate)	0.86	0.54–1.37	.510
PIR 1.30–<3.50 (vs. ≥3.50)	0.72	0.48–1.08	.104

Variable	Adjusted OR	95% CI	p value
PIR <1.30 (vs. ≥3.50)	0.76	0.48–1.19	.207
BMI 25–<30 kg/m² (vs. <25 kg/m²)	1.08	0.71–1.65	.707
BMI ≥30 kg/m² (vs. <25 kg/m²)	1.03	0.67–1.58	.890
Current smoker (vs. never)	1.57	1.00–2.45	.048
Former smoker (vs. never)	1.34	0.94–1.91	.104

Note. Reference categories are no diabetes, male, non-Hispanic White, college graduate, poverty-income ratio ≥3.50, BMI <25 kg/m², and never smoker. The model includes all variables shown simultaneously. OR, odds ratio; CI, confidence interval; PIR, poverty-income ratio; BMI, body mass index; GED, General Educational Development; AA, associate degree.

Supplementary Figure S1

Selected Table 1 characteristics by self-reported cancer-history status.



Note. Each panel displays weighted column percentages within the relevant cancer-history group. BMI, body mass index.