

Clinical Characteristics and Development of a Risk-Identification Model for Incidentally Detected Cancer in Patients with Diabetes: A Case–Control Study

Nengi Cheang , Xueru Chen, Hongmei Zhang, Qing Su, Shichun Du

Department of Endocrinology, Xin Hua Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, 200092, People's Republic of China

Correspondence: Shichun Du; Qing Su, Email dushichun@xinhua.com.cn; suqing@xinhua.com.cn

Introduction: In hospitalised adults with diabetes, asymptomatic cancers may be overlooked during admission, delaying diagnosis and potentially contributing to poorer prognosis by missing opportunities for earlier treatment. Although cross-sectional imaging can detect lesions, it is not systematically performed in asymptomatic patients. A risk signal from routine measurements could prompt targeted imaging and enable diagnosis.

Methods: We conducted a single-centre retrospective case-control study at Xinhua Hospital. Cases were adults with diabetes with first, pathology-confirmed cancer (any site) detected incidentally during index admission without cancer-related symptoms. Controls were inpatients matched by age and diabetes duration. We developed two models. Model A predictors were age, sex, body mass index, HbA1c, mean glucose, mean amplitude of glycaemic excursions (MAGE) from the first 72 h of capillary glucose, 2-hour C-peptide, apolipoprotein A-I (ApoA-I), albumin, and sodium–glucose cotransporter 2 (SGLT2) inhibitor use. Model B added carcinoembryonic antigen (CEA), carbohydrate antigen 19–9 (CA19-9), carbohydrate antigen 125 (CA125), and cytokeratin 19 fragment (CYFRA21-1); albumin was not retained. Performance was assessed by area under the receiver operating characteristic curve (AUC), calibration, Brier score, decision curve analysis (DCA), and net reclassification improvement (NRI).

Results: Of 281 patients, 140 were cases and 141 were controls. Model A used 219 participants and Model B 156 participants. Higher 2-hour C-peptide and ApoA-I were associated with lower odds of cancer, whereas higher CYFRA21-1 was associated with higher odds. Model B outperformed Model A (AUC 0.740 to 0.852; Brier 0.199 to 0.147).

Conclusion: Using inpatient measurements, we developed a risk-identification model for incidentally detected, asymptomatic cancer among hospitalised adults with diabetes, which may help prioritise in-hospital diagnostic work-up. Adding tumour markers, particularly CYFRA21-1, improved discrimination. Prospective multicentre validation and assessment of clinical and economic impact are warranted before implementation.

Keywords: diabetes mellitus, incidentally detected cancer, inpatients, glycaemic variability

Introduction

Globally, an estimated 589 million adults are living with diabetes in 2024, and this number is projected to rise to 853 million by 2050.¹ In 2022, there were approximately 20 million new cancer cases worldwide and 9.7 million cancer deaths, and the global burden is expected to further increase with population ageing.² When diabetes mellitus (DM) and cancer co-occur, diagnosis is often delayed, treatment-related complications increase, and survival is worse, imposing substantial clinical and economic burdens on patients and health systems.^{3–9} As populations age and multimorbidity increases, early identification of cancer in people with DM is a clinical priority, including inpatient settings where incidental first detections occur.



Observational studies indicate that DM is associated with higher risks of several cancers. Large pooled analyses report approximately 10–30% higher incidence for cancers of the liver, pancreas, colorectum, breast, endometrium, and kidney.^{10,11} Some site-specific risks are larger; in women with diabetes, endometrial cancer risk is roughly twofold. An inverse association is often reported for prostate cancer, potentially reflecting hormonal and detection factors.¹² Beyond incidence, diabetes has been linked to higher cancer-specific mortality, likely influenced by biology, comorbidity, and treatment patterns.^{13,14} In inpatient practice, however, diagnostic work-up is commonly initiated by overt symptoms, whereas early-stage cancers may remain clinically silent. Consequently, cancer may be under-recognized and not specifically suspected during routine inpatient care, and may first be diagnosed incidentally during investigations undertaken for other indications. In a national English cancer audit, incidental diagnoses accounted for approximately 4% of cancers (about 1 in 25), and the odds increased with age¹⁵. This diagnostic gap underscores the need for pragmatic risk stratification approaches that leverage routinely collected inpatient measures without requiring specialized monitoring.

Emerging data suggest that short-term glycaemic variability (GV) relates to carcinogenesis through oxidative stress and inflammation.¹⁶ Mechanistically, rapid glucose excursions can trigger recurrent mitochondrial ROS surges and activation of redox-sensitive inflammatory signalling, leading to enhanced pro-inflammatory cytokine release. These oxidative and inflammatory perturbations can promote DNA damage, pro-proliferative and angiogenic programmes, and immune dysregulation, thereby facilitating tumour initiation and progression in the diabetic milieu.^{10,11,16} Prior studies often quantified variability with continuous glucose monitoring metrics (CGM), including time in range, coefficient of variation, and mean amplitude of glycaemic excursions.^{17,18} Because CGM is not used uniformly in inpatient care, we derived inpatient-compatible indices from 72-hour capillary glucose profiles to characterize both glycaemic exposure (mean glucose) and short-term variability, including the mean amplitude of glycaemic excursions (MAGE), using standard methods.¹⁹ Beyond glycaemic dysregulation, impaired β -cell reserve and adverse metabolic–nutritional status may further shape cancer susceptibility and clinical detectability; therefore, we also incorporated 2-hour C-peptide, apolipoprotein A-I (ApoA-I), and albumin as routinely available ward-based markers with biological relevance to tumour progression.^{20–22}

On this basis, we studied adults with diabetes admitted to a tertiary hospital and focused on incidentally detected asymptomatic cancer during the index admission. The primary objective was to determine whether routinely collected inpatient measures identify patients at higher likelihood of such detection. These measures included HbA1c; mean glucose and the mean amplitude of glycaemic excursions calculated from 72-hour capillary glucose profiles; 2-hour C-peptide; ApoA-I; albumin; and sodium–glucose cotransporter-2 (SGLT2) inhibitor use. We hypothesized that higher HbA1c and greater glycaemic variability, together with lower 2-hour C-peptide, ApoA-I, and albumin, would be independently associated with detection. We also tested whether adding common tumour markers (CEA, CA19-9, CA125, CYFRA21-1) improves model discrimination and assessed calibration and decision analysis. The aim was to support inpatient risk stratification and to aid clinicians in targeted evaluation for occult cancer using routine measures.

Methods

Study Design and Participants

This retrospective case-control study was conducted in accordance with the principles of the Declaration of Helsinki and was approved by the Ethics Committee of Xinhua Hospital, Shanghai Jiao Tong University School of Medicine (Ethics approval No. XHEC-C-2025-262-1). We screened consecutive adult inpatients with confirmed diabetes admitted to the Department of Endocrinology over a ten-year period. Diabetes was diagnosed and classified per the China Guideline for the Prevention and Treatment of Diabetes (2024 edition).²³ Diabetes type (type 1, type 2, or other specific types) was determined and recorded according to the same guideline. Only the first (index) admission per patient was included. Cases were patients in whom a cancer (malignant tumour) was first identified during the index admission while asymptomatic at presentation; diagnoses were histopathologically confirmed and coded with reference to the WHO Classification of Tumours (5th edition) and the International Classification of Diseases for Oncology, 3rd Edition (ICD-O-3).^{24,25} We excluded individuals with a prior cancer or hereditary cancer syndromes, severe organ dysfunction, long-term systemic glucocorticoid or other immunosuppressant exposure, or acute metabolic decompensation during the admission, including diabetic ketoacidosis or hyperosmolar hyperglycaemic state. Controls were contemporaneous

inpatients with diabetes and no clinical, imaging or laboratory evidence of cancer during the same admission; they were frequency-matched to cases by diabetes duration and age, with the sex distribution balanced within 5%, required to have negative inpatient tumour screening and complete metabolic laboratory data, and further excluded for prior cancer, acute metabolic decompensation or acute infection, immunosuppressed states (chronic biologic therapy, solid-organ transplantation or HIV infection) or missing key variables. After application of these criteria, 281 participants entered the analytic cohort (cases 140; controls 141). A step-wise summary of inclusions and exclusions, including the number of patients at each stage, is provided in [Supplementary Table S1](#).

Statistical Analysis

Continuous variables were assessed for normality using the Shapiro–Wilk test.²⁶ Normally distributed data are presented as mean \pm standard deviation (SD) and compared with the independent samples *t* test; skewed data are presented as median (IQR) and compared with the Mann–Whitney *U*-test.²⁷ Categorical variables are summarised as n (%) and compared using the χ^2 -test or Fisher’s exact test, as appropriate.²⁸ In addition to these comparisons, univariable associations with cancer were estimated by fitting separate logistic regression models for each candidate predictor, reporting odds ratios (ORs) with 95% confidence intervals. Given the frequency matching by age and diabetes duration, with sex distribution balanced, between-group balance was additionally summarised using standardised mean differences (SMDs), with absolute SMD values for age, sex, and diabetes duration all <0.1 ([Supplementary Table S2](#)). Descriptive summaries and univariable comparisons were performed in the full cohort (N=281). Multivariable Model A was fitted in participants with complete data for the clinical–metabolic predictors (n=219). Because tumour marker measurements were not available for all participants, Model B and head-to-head model-performance comparisons were conducted in the tumour-marker-complete subset (n=156; 64 cases and 92 controls). Two-sided tests with $\alpha=0.05$ were used throughout. Where multiple univariable comparisons were made, the false discovery rate was controlled using the Benjamini–Hochberg procedure. Analyses were performed in R (v4.3) and/or SPSS (v26.0).^{29,30}

Model Development and Performance Assessment

We developed multivariable logistic regression models to predict incident, histopathology-confirmed cancer during the index admission. Candidate predictors were prespecified on clinical grounds and routine availability: age, sex, BMI, glycaemic indices (HbA1c, mean glucose) and the mean amplitude of glycaemic excursions (MAGE), markers of β -cell reserve and lipid and nutritional status (2-hour C-peptide, ApoA-I, albumin), and SGLT2 inhibitor use; an extended specification additionally included tumour markers measured during the admission (CEA, CA19-9, CA125, CYFRA21-1). All laboratory variables were obtained from routine, standardized hospital-based testing performed as part of inpatient care and documented in the electronic medical records for the index admission. Mean glucose and 72-hour MAGE were computed from inpatient capillary glucose measurements. Clinically plausible extreme values were retained, and only clearly erroneous or non-physiologic readings were excluded prior to computation. Two multivariable models were fitted: Model A (clinical–metabolic) included age and sex together with the prespecified clinical and metabolic predictors described above; Model B (extended) additionally incorporated the tumour markers; albumin was not retained. Given prespecified candidate predictors and a limited effective sample size, backward stepwise selection was used to obtain a parsimonious model and reduce the risk of overfitting. Multicollinearity among predictors was assessed using variance inflation factors and tolerance statistics, with results provided in [Supplementary Table S3](#). Continuous predictors were modelled as linear terms. Results are presented as adjusted ORs with 95% confidence intervals. Discrimination was quantified by the area under the receiver operating characteristic (ROC) curve with DeLong confidence intervals; calibration was assessed using the calibration intercept and slope, the Brier score, and decile-based calibration plots.^{31–33} Clinical utility was evaluated using decision curve analysis (DCA) and clinical impact curves (CIC) across prespecified thresholds (0.05–0.40).^{34,35} Between-model comparisons used DeLong tests for AUC and reclassification metrics, including net reclassification improvement (NRI) and integrated discrimination improvement (IDI).^{36,37} Model A was fitted in the clinical–metabolic complete-case set (n=219). For head-to-head performance comparisons with Model B, discrimination and calibration metrics were computed in the tumour-marker-complete subset (n=156). Model B was fitted in this subset.

Results

During the study period, 281 adults with diabetes met inclusion criteria; 140 were diagnosed with incidentally detected, asymptomatic cancer during hospitalisation and 141 served as contemporaneous controls (Figure 1). Primary tumour origins among included cancer cases (n=140) are summarised in [Supplementary Table S4](#). Controls were frequency matched to cases by diabetes duration and age; sex distribution was balanced. Multivariable Model A was fitted in the clinical–metabolic complete-case set (n=219), whereas Model B was fitted in the tumour-marker-complete subset (n=156). Analysis of collected clinical and metabolic data showed that hyperlipidaemia was less common in cases (27.86% and 48.94% in the case and control groups, respectively; $P<0.001$). Median HbA1c was higher in cases, 8.70% (IQR, 7.67–10.20), than in controls, 8.00% (7.00–9.70) ($P=0.007$). Median mean glucose was 8.56 mmol/L in cases and 7.96 mmol/L in controls ($P=0.003$). Median 2-hour C-peptide was lower in cases (1.31 and 1.77 nmol/L in the case and control groups, respectively; $P=0.015$). MAGE did not differ significantly between groups (2.94 and 2.97 mmol/L, respectively; $P=0.198$) (Table 1). On univariable analysis, HbA1c was associated with cancer (OR=1.16, 95% CI 1.04–1.31; $P=0.010$), as was mean glucose (OR=1.28, 95% CI 1.11–1.49; $P<0.001$). ApoA-I (OR=0.21, 95% CI 0.09–0.50; $P<0.001$) and albumin (OR=0.92, 95% CI 0.87–0.98; $P=0.009$) were inversely associated. MAGE showed a non-significant increase in odds (OR=1.12, 95% CI 0.97–1.28; $P=0.118$) (Figure 2A–E).

Two multivariable models were then evaluated. Model A included routinely available clinical–metabolic variables. Model B extended Model A by adding common tumour markers (CEA, CA19-9, CA125 and CYFRA21-1); albumin was not retained.

In Model A, MAGE was independently associated with cancer (OR=1.34, 95% CI 1.04–1.73; $P=0.023$). 2-hour C-peptide and ApoA-I were inversely associated (OR=0.53, 0.35–0.80; $P=0.002$; and OR=0.23, 0.06–0.86; $P=0.029$). HbA1c and mean glucose were not significant (both $P=0.064$) (Figure 3A and Table 2). After adding tumour markers (Model B), CYFRA21-1 was positively associated with cancer (OR=1.41, 1.03–1.93; $P=0.034$). 2-hour C-peptide

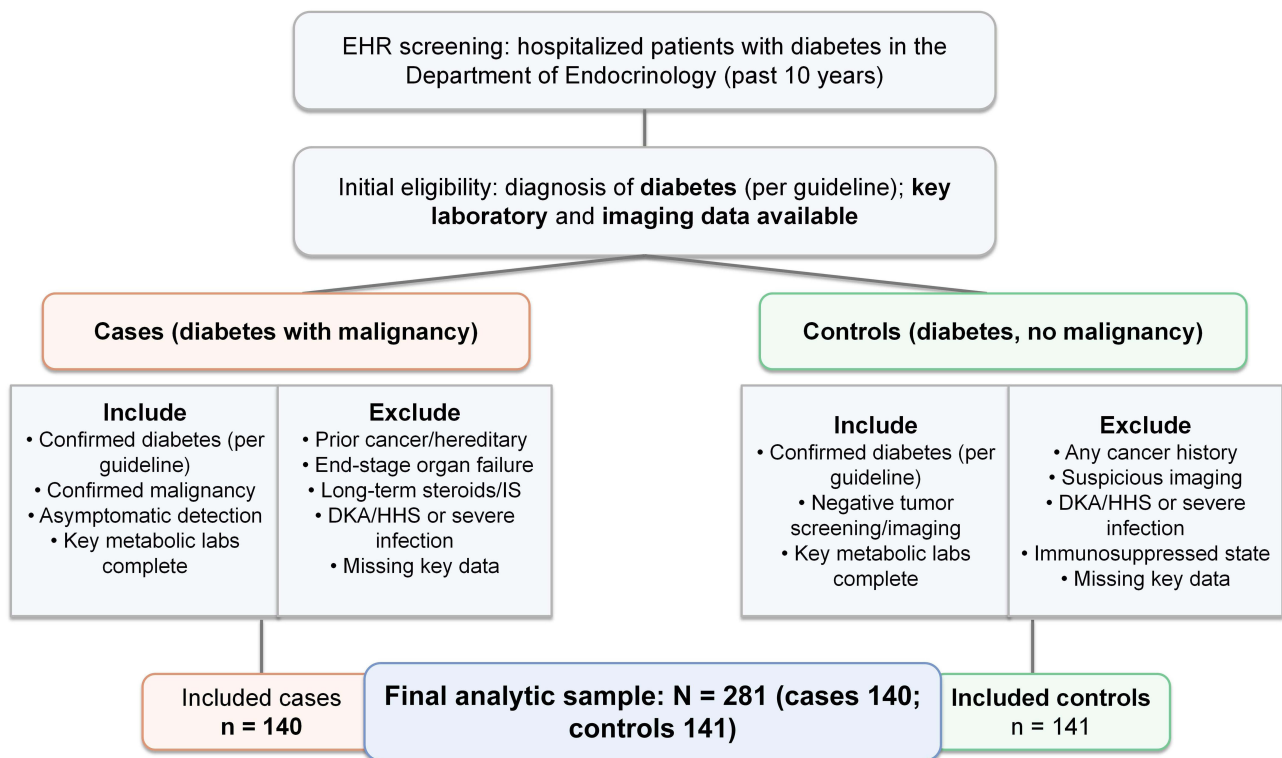


Figure 1 Flow diagram of participant inclusion and exclusion. We screened hospitalised adults with diabetes with key laboratory and imaging data. Cases were incident, asymptomatic cancer confirmed during the index admission. Controls were contemporaneous inpatients without cancer and with negative imaging. Final analytic sample 281 patients with 140 cases and 141 controls.

Table 1 Baseline Clinical and Metabolic Characteristics of Participants with Diabetes by Cancer Status

Variables	Cases (n=140)	Controls (n=141)	P value
Age (years)	67.63 ± 8.53	67.26 ± 8.81	0.742
Sex			0.865
Male (n)	99 (70.71)	101 (71.63)	
Female (n)	41 (29.29)	40 (28.37)	
BMI (kg/m²)	23.98 ± 3.62	24.53 ± 3.38	0.406
Waist circumference (cm)	92.22 ± 9.52	91.59 ± 9.97	0.278
Hip circumference (cm)	96.41 ± 7.37	94.81 ± 10.92	0.309
Waist-to-hip ratio	0.96 (0.91–1.00)	0.95 (0.92–0.99)	0.544
Height (m)	1.68 ± 0.08	1.67 ± 0.09	0.300
Weight (kg)	69.52 ± 12.16	66.96 ± 12.37	0.051
Smoking	45 (32.14)	43 (30.50)	0.766
Drinking	23 (16.43)	19 (13.48)	0.488
Hypertension	88 (62.86)	85 (60.28)	0.657
Hyperlipidemia	39 (27.86)	69 (48.94)	<0.001
Carotid plaque	22 (15.71)	35 (24.82)	0.086
HbA1c (%)	8.70 (7.67–10.20)	8.00 (7.00–9.70)	0.007
Mean glucose (mmol/L)	8.56 (7.50–10.09)	7.96 (7.16–8.98)	0.003
Insulin (pmol/L)	60.53 (38.87–100.26)	55.20 (40.49–85.58)	0.557
C-peptide (nmol/L)	0.59 (0.34–0.95)	0.59 (0.38–0.87)	0.941
2-h insulin (pmol/L)	197.91 (129.67–293.09)	219.12 (137.62–335.65)	0.315
2-h C-peptide (nmol/L)	1.31 (0.80–2.12)	1.77 (1.10–2.55)	0.015
MAGE (mmol/L)	2.94 (1.75–4.02)	2.97 (1.12–3.85)	0.198

Notes: Baseline clinical and metabolic characteristics of the study participants. Data are presented as mean ± SD for approximately normally distributed continuous variables and as median (IQR: 25th–75th percentile) for skewed continuous variables; categorical variables are summarized as n (%). P values were obtained using Student's *t* test or Mann–Whitney *U*-test for continuous variables, as appropriate, and χ^2 -test (or Fisher's exact test when expected counts < 5) for categorical variables. Two-sided tests; $\alpha = 0.05$.

remained protective (OR=0.49, 0.28–0.86; $P=0.012$), and HbA1c became inversely associated (OR=0.68, 0.48–0.95; $P=0.024$). Notably, higher ApoA-I was independently associated with lower odds of cancer (OR=0.07, 0.01–0.53; $P=0.010$). MAGE was not significant; other variables were not significant (Figure 3B). Discrimination improved from an AUC of 0.740 (95% CI 0.659–0.821) for Model A to 0.852 (0.792–0.913) for Model B (DeLong $P=0.002$) (Figure 3C). Calibration was also stronger with Model B (Brier 0.147 vs 0.199; intercept 0.000 and slope 1.000; Hosmer–Lemeshow $P=0.921$ vs 0.227), with closer agreement across deciles of predicted risk (Figure 3D). Full regression coefficients are provided in Supplementary Tables S5–S6. Given incomplete availability of tumour marker measurements, analyses involving Model B were restricted to the tumour-marker-complete subset. To evaluate potential selection bias arising from this restriction, we refitted Model A within the same subset; effect directions for key predictors were broadly consistent (Supplementary Table S7). In additional sensitivity analyses, we further adjusted the tumour marker-extended model (Model B) for erythrocyte sedimentation rate (ESR) and estimated glomerular filtration rate (eGFR) to account for potential confounding by systemic inflammation and renal function. Discrimination remained similar after adding ESR alone or ESR plus eGFR (Supplementary Figure S1).

Having established discrimination and calibration, we next assessed clinical utility and reclassification in the tumour-marker-complete subset ($n=156$). Across threshold probabilities of 5–40%, Model B provided greater net benefit than Model A and either default strategy (treat-all or treat-none) (Figure 4A). Clinical impact curves showed that, at matched thresholds, Model B flagged fewer individuals as high risk while maintaining a similar number of true positives (Figure 4B). Risk reclassification from Model A to Model B was favourable (categorical NRI=0.418, 95% CI 0.239–0.606), with controls more often reclassified downward and cases upward (Figure 4C). Predicted risks were higher in cases than controls for both models (both $P<0.001$) (Figure 4D).

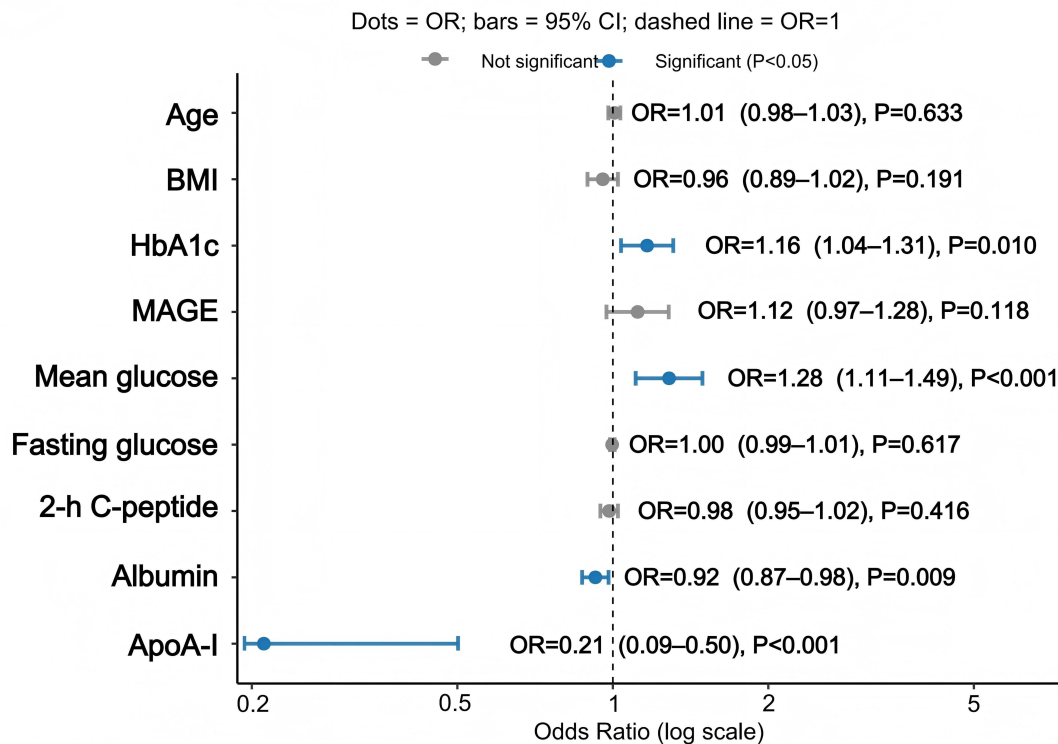
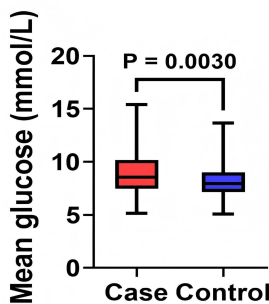
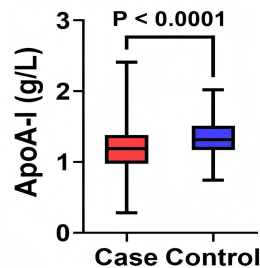
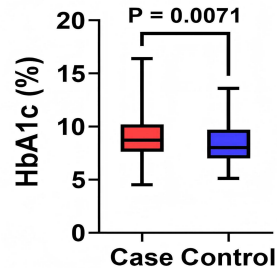
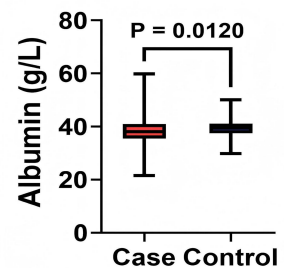
A**Univariable Logistic Regression****B****C****D****E**

Figure 2 Univariable associations and distributions of key clinical variables. **(A)** Univariable logistic regression for case status. Points indicate odds ratios with 95% confidence intervals on a logarithmic scale. The dashed vertical line marks an odds ratio of 1 and blue points indicate predictors with $P < 0.05$. **(B–E)** Mean glucose, ApoA-I, HbA1c and albumin in cases (red) and controls (blue). Box and whisker plots show medians and interquartile ranges with whiskers extending to the data range. Two-sided P values for between-group comparisons are shown above each panel.

Discussion

Diabetes is associated with higher site-specific cancer incidence and poorer outcomes. In hospitalised adults with diabetes, the first recognition of cancer is often incidental rather than informed by structured risk assessment. This creates a need for a practical risk signal based on routine ward data. We conducted a single-centre retrospective case-control study that included consecutive adult inpatients with diabetes. Cases were incident, asymptotically detected cancers confirmed histopathologically during the index admission. Controls were contemporaneous inpatients frequency matched for age and duration of diabetes. From routinely collected measurements we prespecified biologically plausible predictors and fitted a clinical and metabolic model using 72-hour capillary glucose profiles to derive mean glucose and the mean amplitude of glycaemic excursions, 2-hour C-peptide as an index of stimulated β -cell reserve, and ApoA-I. An extended model included four low-cost tumour markers CEA, CA19-9, CA125 and CYFRA21-1. The clinical and metabolic model showed moderate discrimination and clinically coherent coefficients. The extended model improved discrimination and calibration, increased net clinical

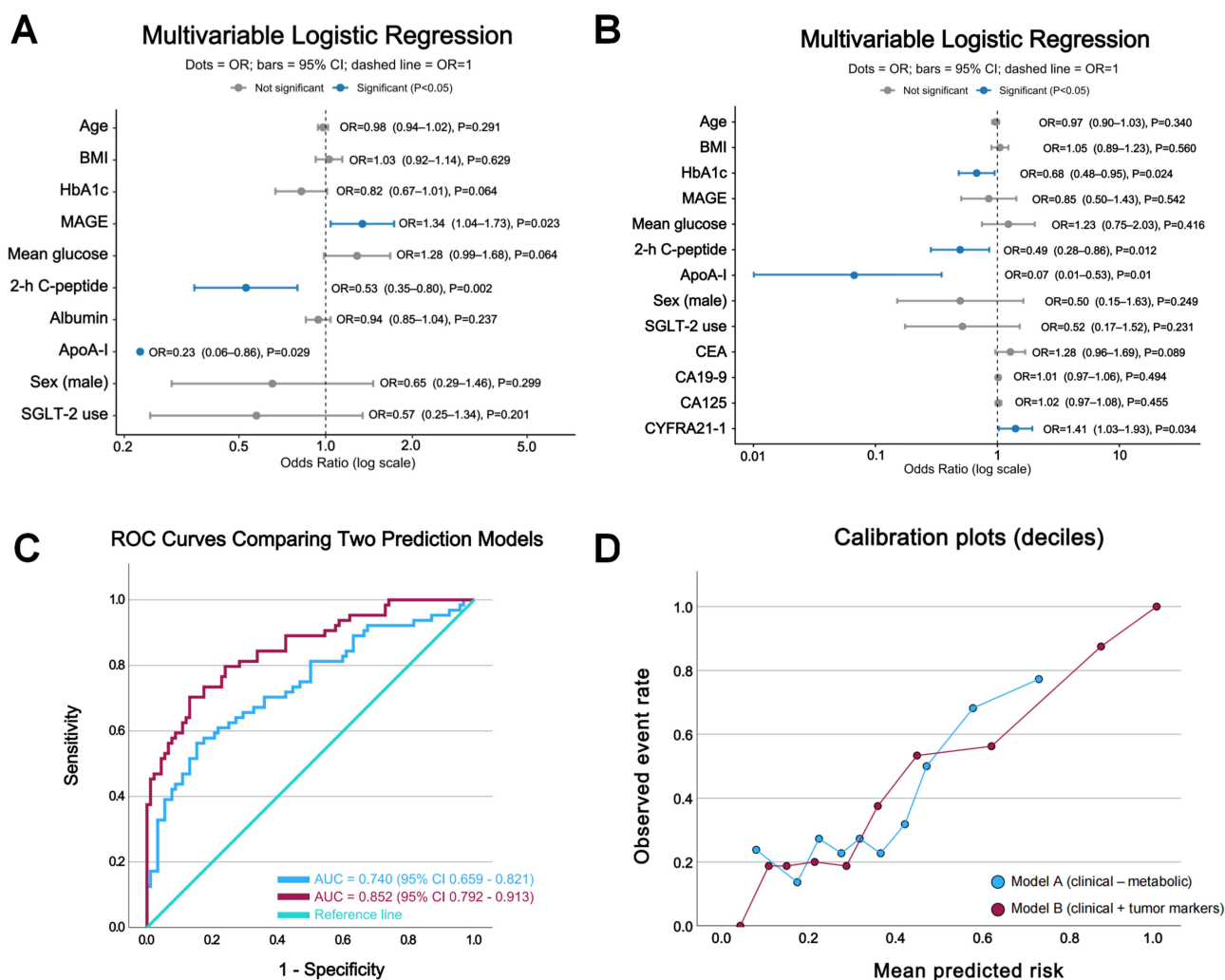


Figure 3 Performance of the clinical metabolic model and the extended model. **(A and B)** Multivariable logistic regression for case status in Model A (clinical plus metabolic predictors) and Model B (Model A plus four tumour markers, with albumin not retained). Points are odds ratios with 95% confidence intervals on a log scale, the vertical line marks an odds ratio of 1, and blue points indicate predictors with P less than 0.05. **(C)** Receiver operating characteristic curves for Model A and Model B with the corresponding areas under the curve. **(D)** Calibration of both models by deciles of predicted risk, with observed event rates plotted against mean predicted risk.

benefit on decision curve analysis, and enhanced categorical net reclassification. Among candidate markers, CYFRA21-1 emerged as an independent predictor that added incremental predictive value beyond clinical and metabolic variables. Greater short-term glycaemic variability, lower stimulated β -cell reserve and lower ApoA-I were associated with a higher likelihood of incidental cancer detection during hospitalisation.

Most prior studies assessed cancer risk in diabetes in community or outpatient cohorts and relied on long-horizon indices such as HbA1c. Inpatient cohorts and incident asymptotically detected cancer during the same admission were rarely examined. Short-term glycaemic variability has mechanistic links to oxidative and inflammatory stress but has seldom been operationalised in risk models.^{16–18} We quantified variability using MAGE from 72-hour capillary glucose profiles with the standard method.¹⁹ In our multivariable clinical and metabolic model, higher MAGE, lower 2-hour C-peptide and lower ApoA-I were associated with a higher likelihood of detection, which is consistent with literature on variability,^{16–18} β -cell related neoplasia risk²⁰ and the role of ApoA-I in lipid transport and systemic inflammation.²¹ Albumin did not remain significant after adjustment. In an inpatient setting it is influenced by hydration, acute-phase responses, hepatic synthetic function and renal loss, which can dilute any cancer-specific signal. Correlation with ApoA-I and other metabolic indices can further reduce its independent contribution, and the short diagnostic window during hospitalisation limits the informativeness of this relatively slow-moving marker.²²

Table 2 Multivariable Logistic Regression for Cancer in Participants with Diabetes

Variable	Model A OR (95% CI)	P value	Model B OR (95% CI)	P value
Demographics				
Age (per 1 year)	0.98 (0.94–1.02)	0.291	0.97 (0.91–1.03)	0.340
Male sex (ref: female)	0.65 (0.29–1.46)	0.299	0.50 (0.15–1.63)	0.249
BMI (per 1 kg/m ²)	1.03 (0.92–1.14)	0.629	1.05 (0.89–1.23)	0.560
Metabolic				
HbA1c (per 1%)	0.82 (0.67–1.01)	0.064	0.68 (0.48–0.95)	0.024
MAGE (per 1 mmol/L)	1.34 (1.04–1.73)	0.023	0.85 (0.50–1.43)	0.542
Mean glucose (per 1 mmol/L)	1.28 (0.99–1.68)	0.064	1.23 (0.75–2.03)	0.416
2-h C-peptide (per 1 nmol/L)	0.53 (0.35–0.80)	0.002	0.49 (0.28–0.86)	0.012
Albumin (per 1 g/L)	0.94 (0.85–1.04)	0.237	—	—
ApoA-I (per 1 g/L)	0.23 (0.06–0.86)	0.029	0.07 (0.01–0.53)	0.010
Medication				
SGLT2 inhibitor use (yes; ref: no)	0.57 (0.25–1.34)	0.201	0.52 (0.17–1.52)	0.231
Tumour markers (Model B only)				
CEA (per 1 ng/mL)	—	—	1.28 (0.96–1.69)	0.089
CA19-9 (per 1 U/mL)	—	—	1.01 (0.97–1.06)	0.494
CA125 (per 1 U/mL)	—	—	1.02 (0.97–1.08)	0.455
CYFRA21-1 (per 1 ng/mL)	—	—	1.41 (1.03–1.93)	0.034
Model performance				
AUC (95% CI)	0.740 (0.659–0.821)		0.852 (0.792–0.913)	
DeLong P (A versus B)		0.002		
Hosmer–Lemeshow P		0.227		0.921
Brier score		0.199		0.147
Calibration intercept and slope		0.244, 1.116		0.000, 1.000

Notes: Values are odds ratios (ORs) with 95% confidence intervals (CI); P values are from Wald tests. Model A included age, sex (ref: female), BMI, SGLT2 inhibitor use (ref: no), HbA1c, MAGE, mean glucose, 2-h C-peptide, albumin, and ApoA-I; Model B additionally included tumour markers (CEA, CA19-9, CA125, CYFRA21-1); albumin was not retained. Age, sex, BMI, and SGLT2 inhibitor use were forced; other variables were selected by backward logistic regression (PIN = 0.05, POUT = 0.10). Regression coefficients were estimated in the model-specific analytical samples (Model A: n = 219; Model B: n = 156). Model performance (AUC with 95% CI, DeLong test for AUC comparison), Brier score, and calibration (Hosmer–Lemeshow goodness-of-fit P value; calibration intercept and slope from logistic recalibration using logit (p) as the predictor) were computed on the same complete-case sample for all Model B predictors (N = 156; cases = 64; controls = 92), which was used for head-to-head comparisons between Model A and Model B.

In the extended model that added tumour markers, CYFRA21-1 remained an independent predictor and improved discrimination and calibration beyond clinical and metabolic variables. Unlike CEA, CA19-9 and CA125, CYFRA21-1 reflects circulating cytokeratin 19 fragments released during epithelial cell turnover and cytolysis.³⁸ Because cytokeratin 19 is widely expressed in epithelial tissues and across many epithelial cancers, CYFRA21-1 may provide a less organ-restricted tumour-related biomarker signal in cohorts with heterogeneous primary sites, thereby complementing organ-oriented markers.³⁹ Clinically, CYFRA21-1 has been most extensively applied in non-small cell lung cancer for disease monitoring and prognostic stratification, and elevations have also been reported in several other epithelial tumours.^{38–40} Notably, CYFRA21-1 is not a screening test and can be influenced by non-malignant conditions (including impaired renal function); therefore, in our setting it should be interpreted as an adjunctive triage marker for risk stratification rather than definitive evidence of cancer.⁴¹ Consistent with this interpretation, routine tumour markers showed no consistent between-group differences among asymptomatic hospitalised adults with diabetes, and CEA, CA19-9 and CA125 did not remain independently associated on multivariable analysis. This pattern likely reflects three factors. First, these assays are organ specific. In a cohort that includes different tumour sites, many cancers have no corresponding marker in the panel.^{42,43} In addition, none of these assays is recommended for screening asymptomatic individuals. Second, admission sampling captures a short diagnostic window and many incidental cancers are at an early stage, so circulating concentrations are often within the reference range. For example, the sensitivity of CEA in early colorectal cancer is low at usual diagnostic thresholds.^{44,45} Third,

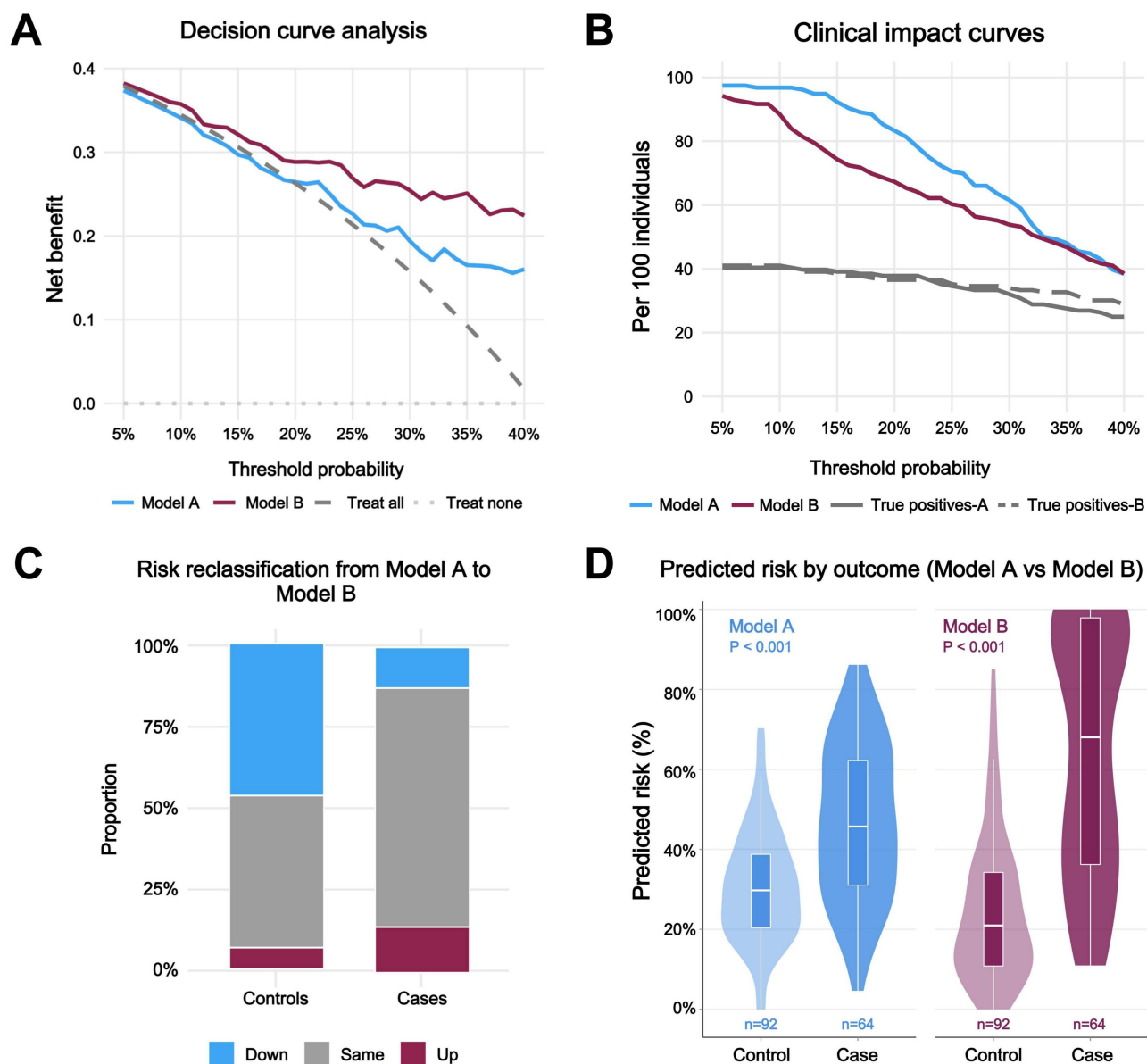


Figure 4 Decision analysis and risk reclassification for the two prediction models. **(A)** Decision curve analysis of net benefit across threshold probabilities for the clinical metabolic model (Model A) and the extended model with tumour markers (Model B), with treat all and treat none strategies shown for reference. **(B)** Clinical impact curves displaying, for each threshold, the number of individuals per 100 classified as high risk and the corresponding number of expected cases for each model. **(C)** Risk reclassification from Model A to Model B among controls and cases, with bars showing the proportions whose predicted risk decreased, stayed the same, or increased. **(D)** Distributions of predicted risk by outcome for Model A and Model B. Violin plots with embedded box plots illustrate separation between controls and cases.

common non-malignant conditions in diabetes wards can influence these markers and reduce specificity and positive predictive value. For example, benign hepatobiliary disease and inflammation can raise CA19-9.⁴⁶

Against this background, ApoA-I remained independently associated with lower odds of cancer after inclusion of tumour markers, and 2-hour C-peptide also remained protective, whereas MAGE did not retain significance in Model B despite being positively associated in the clinical and metabolic specification. ApoA-I is the principal protein component of HDL and a key determinant of HDL functionality, including reverse cholesterol transport.^{47,48} Beyond lipid transport, ApoA-I exerts anti-inflammatory and anti-oxidative effects through modulation of immune cells, cytokine signalling and lipid peroxidation.⁴⁹ Lower ApoA-I indicates an inflammatory, catabolic state and impaired lipid flux that is permissive to tumour initiation and progression in diabetes. Consistent with this interpretation, ApoA-I provides a non-glycaemic host signal that complements HbA1c, mean glucose, and variability metrics and is routinely available on wards. In this context, the lower prevalence of

diagnosed hyperlipidaemia among cases should be interpreted cautiously, as hyperlipidaemia status reflects prior detection and lipid-lowering treatment and may also be influenced by inflammatory or catabolic states during hospitalisation. Therefore, ApoA-I may offer a more objective lipid-related host signal than diagnosed hyperlipidaemia alone.

Regarding the glycaemic indices, MAGE did not remain significant once tumour markers were added, whereas the association for 2-hour C-peptide persisted. HbA1c was not significant in Model A and became an inverse predictor only after tumour markers were included. This pattern is clinically plausible and was not attributable to problematic multicollinearity among glycaemic indices ([Supplementary Table S3](#)). Consistent with the backward stepwise model selection, MAGE was retained and statistically significant in Model A, whereas its association was attenuated and no longer statistically significant after tumour markers were added in Model B ([Tables S5–S6](#)). MAGE tends to move with mean glucose and with early changes in inpatient therapy such as insulin adjustment, diet and fluids, which may reduce its independent signal.^{50,51} 2-hour C-peptide reflects stimulated beta cell reserve and the capacity to sustain an anabolic state under stress, and greater reserve aligned with lower odds of cancer across specifications.^{52,53} HbA1c integrates longer term exposure and is often lowered by treatment changes during admission while short term variability persists.^{54,55}

Our analyses indicate that routine ward measurements can generate an interpretable risk signal for incident, asymptotically detected cancer within a single admission. A clinical and metabolic model based on MAGE, 2-hour C-peptide and ApoA-I achieved moderate discrimination, and an extended model that added four low-cost tumour markers identified CYFRA21-1 as an independent predictor with gains in discrimination, calibration and clinical utility. However, this single-centre retrospective analysis has limitations, including heterogeneity in data sources and testing workflows, limited temporal alignment and completeness of some variables, and incomplete control of confounding, which may affect causal interpretation. In this context, our analyses are intended for risk identification and prediction rather than causal effect estimation, and the observed associations should therefore not be interpreted as causal relationships. Accordingly, while retrospective analyses provide valuable preliminary insight, the findings should be supplemented by prospective studies and, where appropriate, experimental or mechanistic validation.

In sum, this case–control study characterises the clinical profile of incidentally detected cancer in hospitalised adults with diabetes and develops a risk-identification model from routine ward measurements. It shifts inpatient risk assessment from long-horizon exposure metrics to short-horizon physiology. MAGE, 2-hour C-peptide and ApoA-I capture risk dimensions that HbA1c does not, and CYFRA21-1 is identified as a novel independent predictor in the inpatient setting. Because the clinical and metabolic model relies on tests routinely obtained during admission, risk can be computed without additional investigations to identify patients who may warrant focused evaluation for cancer, with further refinement where CYFRA21-1 is available. These findings provide an actionable pathway for bedside triage and set clear hypotheses for mechanistic and site-specific studies.

Conclusion

In this retrospective, single-centre case–control study of hospitalised adults with diabetes, we developed a practical risk-identification model for incidentally detected, asymptomatic cancer during the index admission using routine ward measurements. The clinical–metabolic model achieved moderate discrimination; MAGE, 2-hour C-peptide and ApoA-I were independently associated with detection, and after adding tumour markers MAGE was no longer significant. Adding four low-cost tumour markers identified CYFRA21-1 as an independent predictor and improved discrimination, calibration and clinical utility, with improved reclassification metrics. As the required variables are routinely obtained, risk can be computed without additional testing to prioritise focused evaluation as a pragmatic inpatient triage tool rather than a screening test. However, given the retrospective single-centre design and the lack of external validation, prospective multicentre validation with predefined thresholds and assessment of clinical and economic impact are warranted before implementation. Future studies should also evaluate workflow integration and downstream outcomes, including diagnostic yield, potential harms and cost-effectiveness, under protocolised testing and follow-up.

Use of Generative AI

Generative artificial intelligence tools (ChatGPT, OpenAI, GPT-5.1 Thinking) were used to assist in language editing and polishing of this manuscript. All content was reviewed and verified by the authors, who take full responsibility for the integrity and accuracy of the work.

Data Sharing Statement

The datasets generated and analyzed during the current study are not publicly available due to institutional data protection policies but are available from the corresponding author upon reasonable request.

Ethics Approval and Consent to Participate

This retrospective case–control study was conducted in accordance with the principles of the Declaration of Helsinki. The study protocol was reviewed and approved by the Ethics Committee of Xinhua Hospital, Shanghai Jiao Tong University School of Medicine (Ethics approval No. XHEC-C-2025-262-1). The Ethics Committee granted a waiver of informed consent because this was a retrospective study using de-identified routine clinical data and posed minimal risk to participants. Data were anonymized prior to analysis and handled confidentially in accordance with institutional policies, with access restricted to authorized study personnel.

Acknowledgments

We thank all the patients for allowing us to use their data.

Author Contributions

Nengi Cheang and Xueru Chen are co-first authors.

Nengi Cheang: Conceptualization, Methodology, Investigation, Data Curation, Formal Analysis, Writing – Original Draft.

Xueru Chen: Conceptualization, Methodology, Investigation, Data Curation, Formal Analysis, Writing – Review & Editing.

Hongmei Zhang: Investigation, Writing – Review & Editing.

Qing Su: Conceptualization, Methodology, Supervision, Writing – Review & Editing.

Shichun Du: Conceptualization, Methodology, Supervision, Writing – Review & Editing.

All authors gave final approval of the version to be published; agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

This study was supported by the Shanghai Natural Science Foundation (No. 24ZR1449000). The funder had no role in study design, data collection, data analysis, data interpretation, manuscript preparation, or the decision to publish.

Disclosure

The authors declare that they have no competing interests.

References

1. Federation ID. Facts & figures – IDF diabetes atlas 2025. 2025; Available from: <https://www.idf.org/aboutdiabetes/facts-figures.html>. Accessed February 02, 2026.
2. Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Ca A Cancer J Clinicians*. 2024;74(3):229–263. doi:10.3322/caac.21834
3. Bjornsdottir HH, Rawshani A, Rawshani A, et al. A national observation study of cancer incidence and mortality risks in type 2 diabetes compared to the background population over time. *Sci Rep*. 2020;10(1):17376. doi:10.1038/s41598-020-73668-y
4. Saydah SH, Loria CM, Eberhardt MS, Brancati FL, et al. Abnormal glucose tolerance and the risk of cancer death in the United States. *Am J Epidemiol*. 2003;157(12):1092–1100. doi:10.1093/aje/kwg100
5. Coughlin SS, Calle EE, Teras LR, et al. Diabetes mellitus as a predictor of cancer mortality in a large cohort of US adults. *Am J Epidemiol*. 2004;159(12):1160–1167. doi:10.1093/aje/kwh161

6. Van de Poll-Franse L, Houterman S, Janssen-Heijnen ML, et al. Less aggressive treatment and worse overall survival in cancer patients with diabetes: a large population based analysis. *Int J Cancer*. 2007;120(9):1986–1992. doi:10.1002/ijc.22532
7. Bradley CJ. Cancer, financial burden, and medicare beneficiaries. *J Clin Oncol*. 2017;35(22):2461–2462. doi:10.1200/JCO.2017.73.1877
8. Carrera PM, Kantarjian HM, Blinder VS. The financial burden and distress of patients with cancer: understanding and stepping-up action on the financial toxicity of cancer treatment. *CA Cancer J Clin*. 2018;68(2):153–165.
9. Davis-Ajami ML, Lu ZK, Wu J. Multiple chronic conditions and associated health care expenses in US adults with cancer: a 2010–2015 medical expenditure panel survey study. *BMC Health Serv Res*. 2019;19(1):981. doi:10.1186/s12913-019-4827-1
10. Zhu B, Qu S. The relationship between diabetes mellitus and cancers and its underlying mechanisms. *Front Endocrinol*. 2022;13:800995. doi:10.3389/fendo.2022.800995
11. Pliszka M, Szablewski L. Associations between diabetes mellitus and selected cancers. *Int J Mol Sci*. 2024;25(13):7476. doi:10.3390/ijms25137476
12. Carstensen B, Read SH, Friis S, et al. Cancer incidence in persons with type 1 diabetes: a five-country study of 9,000 cancers in type 1 diabetic individuals. *Diabetologia*. 2016;59(5):980–988. doi:10.1007/s00125-016-3884-9
13. Tsilidis KK, Kasimis JC, Lopez DS, et al. Type 2 diabetes and cancer: umbrella review of meta-analyses of observational studies. *Bmj*. 2015;350:g7607
14. Chen Y, Wu F, Saito E, et al. Association between type 2 diabetes and risk of cancer mortality: a pooled analysis of over 771,000 individuals in the Asia cohort consortium. *Diabetologia*. 2017;60(6):1022–1032. doi:10.1007/s00125-017-4229-z
15. Koo MM, Rubin G, McPhail S, et al. Incidentally diagnosed cancer and commonly preceding clinical scenarios: a cross-sectional descriptive analysis of english audit data. *BMJ Open*. 2019;9(9):e028362. doi:10.1136/bmjopen-2018-028362
16. Saisho Y. Glycemic variability and oxidative stress: a link between diabetes and cardiovascular disease? *Int J Mol Sci*. 2014;15(10):18381–18406. doi:10.3390/ijms151018381
17. Mao X, Cheung KS, Tan J-T, et al. Optimal glycaemic control and the reduced risk of colorectal adenoma and cancer in patients with diabetes: a population-based cohort study. *Gut*. 2024;73(8):1313–1320. doi:10.1136/gutjnl-2023-331701
18. Jospe MR, Liao Y, Giles ED, et al. A low-glucose eating pattern is associated with improvements in glycemic variability among women at risk for postmenopausal breast cancer: an exploratory analysis. *Front Nutr*. 2024;11:1301427. doi:10.3389/fnut.2024.1301427
19. Service FJ, Molnar GD, Rosevear JW, et al. Mean amplitude of glycemic excursions, a measure of diabetic instability. *Diabetes*. 1970;19(9):644–655. doi:10.2337/diab.19.9.644
20. Chen L, Li L, Wang Y, et al. Circulating C-peptide level is a predictive factor for colorectal neoplasia: evidence from the meta-analysis of prospective studies. *Cancer Causes Control*. 2013;24(10):1837–1847. doi:10.1007/s10552-013-0261-6
21. Zamanian-Daryoush M, DiDonato JA. Apolipoprotein A-I and cancer. *Front Pharmacol*. 2015;6:265. doi:10.3389/fphar.2015.00265
22. Yang Z, Zheng Y, Wu Z, et al. Association between pre-diagnostic serum albumin and cancer risk: results from a prospective population-based study. *Cancer Med*. 2021;10(12):4054–4065. doi:10.1002/cam4.3937
23. Chinese Diabetes S. China guideline for the prevention and treatment of diabetes (2024 edition). *Chin J Diabetes Mellitus*. 2025;17(1).
24. International Agency for Research on Cancer. *WHO Classification of Tumours Online (5th Edition)*. Lyon: IARC.
25. World Health Organization. *International Classification of Diseases for Oncology, 3rd Edition (ICD-O-3)*. World Health Organization.
26. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). *Biometrika*. 1965;52(3–4):591–611. doi:10.1093/biomet/52.3-4.591
27. Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. *Ann Math Statist*. 1947;18(1):50–60. doi:10.1214/aoms/1177730491
28. Agresti A. *An Introduction to Categorical Data Analysis*. 3 ed. Hoboken, NJ: Wiley; 2019.
29. Corp I.B.M. *IBM SPSS Statistics for Windows, Version 26.0*. Armonk, NY: IBM Corp.; 2019.
30. Team RC. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2024.
31. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44(3):837–845. doi:10.2307/2531595
32. Van Calster B, McLernon DJ, van Smeden M, et al. Calibration: the achilles heel of predictive analytics. *BMC Med*. 2019;17(1):230. doi:10.1186/s12916-019-1466-7
33. Huang Y, Pepe MS, Feng Z, Gabriel RA, Ohno-Machado L. A tutorial on calibration measurements and calibration models for clinical prediction models. *J Am Med Inf Assoc*. 2020;27(4):621–633. doi:10.1093/jamia/ocz228
34. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Mak*. 2006;26(6):565–574. doi:10.1177/0272989X06295361
35. Vickers AJ, Van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. *Diagnost Prognost Res*. 2019;3(1):18. doi:10.1186/s41512-019-0064-7
36. Pencina MJ, D'Agostino RB, D'Agostino RB, et al. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med*. 2008;27(2):157–172. doi:10.1002/sim.2929
37. Pencina MJ, D'Agostino RB, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med*. 2011;30(1):11–21. doi:10.1002/sim.4085
38. Pujol JL, Grenier J, Daurès JP, et al. Serum fragment of cytokeratin subunit 19 measured by CYFRA 21-1 immunoradiometric assay as a marker of lung cancer. *Cancer Res*. 1993;53(1):61–66.
39. Sheard MA, Vojtesek B, Simickova M, et al. Release of cytokeratin-18 and -19 fragments (TPS and CYFRA 21-1) into the extracellular space during apoptosis. *J Cell Biochem*. 2002;85(4):670–677. doi:10.1002/jcb.10173
40. Pujol JL, Molinier O, Ebert W, et al. CYFRA 21-1 is a prognostic determinant in non-small-cell lung cancer: results of a meta-analysis in 2063 patients. *Br J Cancer*. 2004;90(11):2097–2105. doi:10.1038/sj.bjc.6601851
41. Mikkelsen G, Åsberg A, Hultström ME, et al. Reference limits for chromogranin A, CYFRA 21-1, CA 125, CA 19-9 and carcinoembryonic antigen in patients with chronic kidney disease. *Int J Biol Markers*. 2017;32(4):e461–e466. doi:10.5301/ijbm.5000278
42. Force USPST, Curry SJ, Owens DK. Screening for ovarian cancer: US preventive services task force recommendation statement. *JAMA*. 2018;319(6):588–594. doi:10.1001/jama.2017.21926
43. Locker GY, Hamilton S, Harris J, et al. ASCO 2006 update of recommendations for the use of tumour markers in gastrointestinal cancer. *J Clin Oncol*. 2006;24(33):5313–5327. doi:10.1200/JCO.2006.08.2644

44. Wang R, Wang Q, Li P. Significance of carcinoembryonic antigen detection in the early diagnosis of colorectal cancer: a systematic review and meta-analysis. *World J Gastrointest Surg.* 2023;15(12):2907–2918. doi:10.4240/wjgs.v15.i12.2907
45. Gulhati P, Yin J, Pederson L, et al. Threshold change in CEA as a predictor of non-progression to first-line systemic therapy in metastatic colorectal cancer patients with elevated CEA. *J Natl Cancer Inst.* 2020;112(11):1127–1136. doi:10.1093/jnci/djaa020
46. Lin MS, Huang JX, Yu H. Elevated serum level of carbohydrate antigen 19-9 in benign biliary stricture diseases can reduce its value as a tumour marker. *Int J Clin Exp Med.* 2014;7(3):744–750.
47. Phillips MC. New insights into the determination of HDL structure by apolipoproteins: thematic review series: high density lipoprotein structure, function, and metabolism. *J Lipid Res.* 2013;54(8):2034–2048. doi:10.1194/jlr.R034025
48. Huang R, Silva RAGD, Jerome WG, et al. Apolipoprotein A-I structural organization in high-density lipoproteins isolated from human plasma. *Nat Struct Mol Biol.* 2011;18(4):416–422. doi:10.1038/nsmb.2028
49. Rosenson RS, Brewer HB, Ansell BJ, et al. Dysfunctional HDL and atherosclerotic cardiovascular disease. *Nat Rev Cardiol.* 2016;13(1):48–60. doi:10.1038/nrcardio.2015.124
50. Saisho Y, Tanaka C, Tanaka K, et al. Relationships among different glycemic variability indices obtained by continuous glucose monitoring. *Prim Care Diabetes.* 2015;9(4):290–296. doi:10.1016/j.pcd.2014.10.001
51. Dungan K, Binkley P, Osei K. Glycemic variability during algorithmic titration of insulin among hospitalized patients with type 2 diabetes and heart failure. *J Diabetes Complications.* 2016;30(1):150–154. doi:10.1016/j.jdiacomp.2015.09.009
52. Leighton E, Sainsbury CAR, Jones GCR. A practical review of C-peptide testing in diabetes. *Diabetes Therapy.* 2017;8(3):475–487. doi:10.1007/s13300-017-0265-4
53. Maddaloni E, Bolli GB, Frier BM, et al. C-peptide determination in the diagnosis of type of diabetes and its management: a clinical perspective. *Diabetes Obesity Metabol.* 2022;24(10):1912–1926. doi:10.1111/dom.14785
54. Spencer DH, Grossman BJ, Scott MG, Hirano T. Red cell transfusion decreases hemoglobin A1c in patients with diabetes. *Clin Chem.* 2011;57(1):57. doi:10.1373/clinchem.2010.149559
55. Vatcheva KP, Lee M, McCormick JB, Rahbar MH, et al. Multicollinearity in regression analyses conducted in epidemiologic studies. *North Am J Med Sci.* 2016;8(7):361–371.

Diabetes, Metabolic Syndrome and Obesity

Publish your work in this journal

Diabetes, Metabolic Syndrome and Obesity is an international, peer-reviewed open-access journal committed to the rapid publication of the latest laboratory and clinical findings in the fields of diabetes, metabolic syndrome and obesity research. Original research, review, case reports, hypothesis formation, expert opinion and commentaries are all considered for publication. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/diabetes-metabolic-syndrome-and-obesity-journal>

Dovepress
Taylor & Francis Group