Predicting values of lipids and white blood cell count for all-site cancer in type 2 diabetes

Xilin Yang1*, Wing-Yee So1*, Ronald C W Ma1, Gary T C Ko2, Alice P S Kong1,3, Qingsheng Wang4, Clive S Cockram1, Chun-Chung Chow1, Juliana C N Chan1,2,3 and Peter C Y Tong1

1Department of Medicine and Therapeutics, 2Hong Kong Institute of Diabetes and Obesity and 3Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, The Prince of Wales Hospital, Shatin, Hong Kong SAR, China
4Tianjin Cancer Hospital and Institute, Tianjin, China

(Correspondence should be addressed to J C N Chan; Email: jchan@cuhk.edu.hk)

* X Yang and W-Y So contributed equally to this work

Abstract

Type 2 diabetic patients have increased cancer risk. We developed and validated an all-site cancer risk score in a prospective cohort of 7374 Chinese type 2 diabetic patients free of known history of cancer at enrolment, using split-half validation. Spline Cox model was used to detect common risk factors of cancer and to guide linear transformation of non-linear risk factors. After a median follow-up period of 5.45 years, 365 patients (4.95%) developed cancer. Body mass index (BMI; <24.0 or ≥27.6 kg/m²), triglyceride (≥0.81 to <1.41 mmol/l), high-density lipoprotein cholesterol (<0.9 or ≥1.8 mmol/l), total cholesterol (<4.3 mmol/l) and white blood cell (WBC) count (<5.8×10⁹ count per litre) were associated with increased cancer risks and exhibited non-linear relationships. We further linear transformed these terms for selection using backward Cox regression (P<0.05 for stay) in the training dataset. In the test dataset, calibration was checked using Hosmer–Lemeshow test and discrimination checked using area under receiver operating characteristic curve. In addition to age and current smoking, only linear-transformed total cholesterol and WBC count were selected. The risk score was 0.0488×age (years)−0.5810×total cholesterol (mmol/l, coded to 4.3 if >4.3)−0.3596×WBC count (10⁹ counts/l, 5.8 if >5.8)+0.6390×current smoking status (1 if yes). The 5-year probability of cancer was 1−0.9590×EXP(0.9382×(RISK SCORE+1.5903)). The predicted cancer probability was not significantly different from the observed cancer probability during the 5-year follow-up. The adjusted area under receiver operating characteristic curve was 0.712. In conclusion, BMI, lipids and WBC count have predicting values for cancer.

Endocrine-Related Cancer (2008) 15 597–607

Introduction

There is a global epidemic of type 2 diabetes mellitus (T2DM; Wild et al. 2004). Besides increasing risks of coronary heart disease, stroke and chronic kidney disease (Manson et al. 1991, Almdal et al. 2004, Wu et al. 2005), T2DM also predisposes patients to increased risk of cancer including breast (Lipscombe et al. 2006) and endometrial cancers (Friberg et al. 2007) in women and prostate cancers in men (Rodriguez et al. 2005). Other cancers include colorectal (Seow et al. 2006), pancreatic (Huxley et al. 2005) and liver (Rousseau et al. 2006), suggesting that multiple-site cancers in T2DM may share common risk factors.

The Framingham Study developed a risk score for predicting coronary heart disease in general population (Anderson et al. 1991). Using the Hong Kong Diabetes Registry, our group has developed risk scores to predict end-stage renal disease (Yang et al. 2006b, 2007c), stroke (Yang et al. 2007b) and coronary heart disease (Yang et al. 2008a) in Chinese T2DM patients. More recently, our group has developed a hazard ratio (HR) curve technique to identify the full-range risk relationships...
between common risk factors and all-cause death (So et al. 2008) and coronary heart disease (Yang et al. 2007a). Based on the identified curving relationships between risk factors and all-cause death, we also developed a risk score to predict all-cause death in T2DM (Yang et al. 2008b). In this analysis, we applied similar methods to delineate the relationships between risk factors and all-site cancer followed by validation of a risk score to predict all-site cancer in T2DM.

Methods

Participants

The Prince of Wales Hospital is the teaching hospital of the Chinese University of Hong Kong and serves a population of over 1.2 million. The Hong Kong Diabetes Registry was established in 1995 as part of a quality improvement programme and enrols 30–50 ambulatory diabetic patients on a weekly basis. The referral sources included general practitioners, community clinics and other specialty clinics and patients discharged from the Prince of Wales Hospital or other regional hospitals. Enrolled patients with hospital admissions within 6–8 weeks prior to assessment accounted for <10% of all referrals. The 4-h assessment of complications and risk factors was performed on an outpatient basis, modified from the European DiabCare protocol (Piwernetz et al. 1993). Once a diabetic subject had undergone this comprehensive assessment, he/she was considered to have entered this study cohort and would be followed up till the time of death. Ethical approval was obtained from the Chinese University of Hong Kong Clinical Research Ethics Committee. The Declaration of Helsinki was adhered and informed consent obtained from every patient for data analysis and research purpose at the time of assessment.

For this analysis, the clinical endpoints including discharge diagnoses of hospital admissions and mortality on or before 30 July 2005 were recorded or otherwise censored on 30 July 2005. Details of all medical admissions of the cohort by that date were retrieved from the Hong Kong Hospital Authority Central Computer System that records admissions to all public hospitals in Hong Kong. The latter provide 95% of the total hospital bed-days in Hong Kong due to its heavily subsidized health care system. Mortality data from the Hong Kong Death Registry was also retrieved and all causes of death were further ascertained by review of case notes by an endocrinologist. These databases were matched by a unique identification number, the Hong Kong Identity Card number, which is compulsory for all residents in Hong Kong.

Hospital discharge principal diagnoses coded by the International Classification of Diseases, Ninth Revision (ICD-9) were used to identify first cancer events. The endpoint of this study was defined as having a first incident cancer (codes 140–208). Upper digestive tract cancer was defined as malignant neoplasm of esophagus and stomach (codes 150–151) and lower digestive tract cancer defined as malignant neoplasm of colon, rectum, rectosigmoid junction and anus (codes 153–154).

From 1995 to 2005, 7920 diabetic patients were enrolled in this Registry. Among them, 332 with type 1 diabetes defined as acute presentation with diabetic ketoacidosis, heavy ketonuria (>3+) or continuous requirement of insulin within 1 year of diagnosis (Laakso & Pyorala 1985), and 5 with uncertain type 1 diabetes status, were excluded from the analysis. Among them, 49 with non-Chinese or unknown nationality were excluded. Another 160 patients were excluded due to past history of cancer. A total of 7374 Chinese T2DM patients without cancer history were included in the analysis.

Clinical measurements and laboratory assays

Details of assessment methods and definitions have been described previously (Yang et al. 2006b). On the day of assessment, patients attended the centre after an 8-h fasting and underwent anthropometric measurements and laboratory investigations. This study used the abbreviated MDRD formula recalibrated for the Chinese (Ma et al. 2006) to estimate glomerular filtration rate expressed in ml/min 1.73/m²: eGFR = 186×(SCR×0.011)−1.154×(age)−0.203×(0.742 if female)×1.233, where SCR is serum creatinine expressed as μmol/l (original mg/dl converted to μmol/l) and 1.233 is the adjusting coefficient for the Chinese. Body weight and height were measured with subjects wearing light clothing and without shoes. Blood pressure (BP) was measured using a Dinamap PRO-100 device (Critikon, Milwaukee, WI, USA).

Complete blood picture was assayed using Beckman Coulter counter. Plasma glucose was measured by a hexokinase method (Hitachi 911 automated analyser, Boehringer Mannheim). Total cholesterol, triglyceride and high-density lipoprotein cholesterol (HDL-C) were measured by enzymatic methods on a Hitachi 911 automated analyser (Boehringer Mannheim) using the reagent kits supplied by the manufacturer of the analyser. The inter- and intra-assay coefficients of variation have been described previously (Yang et al. 2006b).
Statistical analyses

The Statistical Analysis System (Release 9.10) was used to perform the statistical analysis (SAS Institute Inc., Cary, NC, USA). Restricted cubic splines (RCS) are piecewise cubic polynomials connected across different intervals of a continuous variable (Harrell 2001). It can fit sharply curving shapes with an additional advantage that only \( k - 1 \) parameters must be estimated (\( k \) is the number of knots) (Harrell 2001). With the application of RCS in Cox models, we had developed a HR curve method to examine full-range associations between risk factors and endpoints such as death (So et al. 2008) and coronary heart disease (Yang et al. 2007). As before, we chose four knots at quantiles 0.05, 0.35, 0.65 and 0.95, which has been suggested to offer adequate fit of the model and is a good compromise between flexibility and loss of precision caused by overfitting a small sample (Harrell 2001). HR between two points of a continuous variable can be estimated by \( \text{EXP} (Y_2 - Y_1) \), where \( Y_1 \) and \( Y_2 \) are the corresponding spline function values of the two points. If we select a proper point \( Y_i \) as the referent, \( \text{EXP} (Y_2 - Y_i) \) stands for the HR of point 2 versus point 1. Thus, we obtained the HR curves by plotting the HRs of all other points versus the referent point.

In this analysis, a forward stepwise algorithm (\( P < 0.05 \) for entry and stay) was used to identify the risk factors of all-site cancer. The detail on how to use stepwise algorithm in the spline Cox model has been described (So et al. 2008). Based on the shapes of these curved relationships between risk factors and all-site cancer, these continuous variables were categorized into two or three levels and then traditional Cox analysis was performed to confirm the findings in the risk curve analysis. We also attempted to devise simple formulae to mimic the curved relationships between identified risk factors and all-site cancer to develop a cancer risk score.

Split-half validation was used in developing and validating the risk score (Yang et al. 2008). In the training subsample, Cox proportional hazard regression with the backward algorithm (\( P < 0.05 \) for stay) was used to obtain a subset of baseline predictors for all-site cancer, which were identified in the above step. Proportional hazards assumption and functional form were checked using the Supremum test (Lin et al. 1993). \( P < 0.05 \) for proportional hazards was considered to violate the assumption while \( P < 0.05 \) for functional form suggests that further improvement in the linear transformation remains possible. The construction of a risk score and a t-year probability equation have been described in detail elsewhere (Yang et al. 2008a). Let LR denote the likelihood ratio \( \chi^2 \) and \( P \) the number of predictors in the final model, then the estimated shrinkage is \( (\text{LR-p})/\text{LR} \) and a shrinkage below 0.85 raises concern of overfitting (Harrell 2001).

Validation of the risk score was performed using the test dataset. Calibration was checked using the Hosmer–Lemeshow test (Yang et al. 2008b). \( P < 0.05 \) indicates significant difference between the predicted and observed rates, suggesting poor calibration. We used Pencina & D’Agostino (2004) method to calculate C index (or c statistics) for overall discrimination and Chambless & Diao (2006) method to calculate the aROC, sensitivity and specificity for the 5-year specific discrimination and selection of cut-off points. Kaplan–Meier estimator was also used to calculate the cumulative incidence of all-site cancer in patients with a risk score at or above and below the cut-off point.

In order to identify the contributions of cancer subtypes to the overall predicting ability of the risk score, a series of Cox regression analyses using the whole cohort were performed, with the risk score as the sole predictor and major subtypes of cancer in the cohort as endpoints. \( P < 0.05 \) is considered to be statistically significant.

Results

Characteristics of study population

The median age of the cohort was 57 (interquartile range (IQR) 47–67) years. At the entry of the study, they had a median duration of 6 years of diabetes (IQR 1–11 years). 4.95% (365 patients) of the cohort developed cancer during a median follow-up period of 5.45 (IQR 2.92–7.78) years and the annual incidence rate was 9.20 (95% CI 8.26–10.14) per 1000 person-years. There were 732 deaths, including 161 deaths with cancer as the principal diagnosis. Comparisons of baseline variables between the cancer cases and their non-cancer counterparts are listed in Table 1. The training dataset and the test dataset contained 3701 (total cancer 187) and 3673 (total cancer 178) patients, respectively. Baseline variables of the two groups are listed in Table 1.

Non-linear relations between risk factors and cancer

Age, body mass index (BMI), current smoker status, total cholesterol, HDL-C, triglyceride and white blood cell (WBC) were selected as predictors of incident cancer by the stepwise spline Cox regression model.
Except for age that was linearly associated with cancer, other continuous risk factors were associated with all-site cancer in a non-linear manner (Fig. 1). BMI was associated with cancer in a linear manner in univariate analysis but in a U-shaped relationship in multivariate analysis. HDL-C was associated with cancer in a U-shaped manner with the nadir at 1.22 mmol/l. Total cholesterol was negatively associated with rapid increase in cancer risk at levels >4.3 mmol/l. From 0.2 to 2.5 mmol/l, triglyceride was associated with cancer in an A-shaped manner with the zenith at 1.17 mmol/l. WBC count exhibited a negative risk association with all-site cancer as the count fell below 5.8 \times 10^9 \text{ counts per litre.}

Using categorized continuous variables, all these predictors remained significant on multivariate analysis (Table 2). The significance of low and high BMI, high HDL-C, low WBC count and low total cholesterol persisted after excluding patients with a follow-up period of \(<2.5\) years. The significance of male gender in univariate analysis disappeared \(P > 0.10\) once smoking status or WBC count was included in the model. Subgroup analysis did not reveal marked difference in effect size by sex (see Supplementary Table 1, which can
be viewed online at http://erc.endocrinology-journals.org/supplemental/ for details). Other variables including HbA1c, albuminuria, renal function and BP and drug use were not selected into the model.

The risk score and the probability equation

In the training database, age, current smoking and formula linear-transformed total cholesterol and WBC count were selected to predict incident cancer while ex-smoking and the formula linear-transformed BMI, HDL-C and triglyceride were not selected into the predicting model by the backward algorithms (Table 3). The overall C index of the predicting model in the training dataset was 0.716 (95% CI 0.685–0.747). The shrinkage of the predicting model was, if any, small (shrinkage = 0.9382) and the formulae linear-transformed TC and WBC were adequate ($P > 0.05$ for functional form). The constructed risk score was $0.0488 \times \text{age (years)} - 0.5810 \times \text{total cholesterol (mmol/l, coded to 4.3 if } > 4.3) - 0.3596 \times \text{WBC}$

Figure 1 Hazard ratio curves for age, body mass index, white blood cell count and lipids for all-site cancer in Chinese patients with type 2 diabetes. Dot (black) curves derived from univariate analyses; star (grey) curves derived from multivariate models adjusted for smoking status (current/ex) and spline functions of (a) age, (b) body mass index (BMI), (c) white blood cell (WBC) count, (d) high-density lipoprotein cholesterol (HDL-C), (e) total cholesterol and (f) triglyceride; $P < 0.05$ for all curves. Unselected variables by the stepwise spline Cox model included duration of diabetes, HbA1c, spot urine albumin:creatinine ratio, glomerular filtration rate, systolic/diastolic blood pressure, peripheral arterial disease, sensory neuropathy, retinopathy and baseline drug use (lipid-lowering drugs, ACE inhibitors or angiotensin II receptor blockers and insulin).
The 5-year probability of cancer was \(0.9590 \exp(0.9382 \times \text{RISK SCORE} + 1.5903)\).

**Validation of the cancer risk score in the test data**

In the test database, the predicted cancer probability was not significantly different from the observed probability of cancer over 5 years of follow-up (Fig. 2). The overall C index was 0.695 (95% CI 0.667–0.724). With follow-up time and censoring status considered, the adjusted aROC was 0.712 for a 5-year observation period. Using a cut-off point of 1.2133, the sensitivity was 56.6% and the specificity was 71.4%. Sensitivity and specificity at other selected cut-off points are listed in Table 4.

**Cumulative incidences of all-site cancer in patients with type 2 diabetes**

<table>
<thead>
<tr>
<th>Parameter estimates and hazard ratio of baseline variables for cancer</th>
<th>Univariate</th>
<th>Multivariate*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hazard ratio (95% CI)</strong></td>
<td><strong>P value</strong></td>
<td><strong>Hazard ratio (95% CI)</strong></td>
</tr>
<tr>
<td>Age (years)</td>
<td>1.05 (1.05–1.06)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Female sex</td>
<td>0.74 (0.61–0.91)</td>
<td>0.0047</td>
</tr>
<tr>
<td>Body mass index (kg/m²)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;24.0</td>
<td>1.33 (1.06–1.68)</td>
<td>0.0156</td>
</tr>
<tr>
<td>≥24.0 to &lt;27.6</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>≥27.6</td>
<td>0.95 (0.70–1.28)</td>
<td>0.7264</td>
</tr>
<tr>
<td>Total cholesterol (mmol/l)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;0.90</td>
<td>1.58 (1.18–2.10)</td>
<td>0.0021</td>
</tr>
<tr>
<td>≥0.90 to &lt;1.80</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>≥1.80</td>
<td>1.61 (1.16–2.22)</td>
<td>0.0039</td>
</tr>
<tr>
<td>White blood cell count (10⁹ count/l)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;5.8</td>
<td>1.56 (1.20–2.04)</td>
<td>0.0011</td>
</tr>
<tr>
<td>≥5.8 to &lt;9.0</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>≥9.0</td>
<td>1.16 (0.84–1.61)</td>
<td>0.3756</td>
</tr>
<tr>
<td>Smoking status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Ex-smoking</td>
<td>1.73 (1.30–2.29)</td>
<td>0.0001</td>
</tr>
<tr>
<td>Current smoking</td>
<td>1.44 (1.14–1.84)</td>
<td>0.0027</td>
</tr>
</tbody>
</table>

*Multivariate model adjusted for age, sex, body mass index, high-density lipoprotein cholesterol, total cholesterol, triglyceride, white blood cell count and smoking status. Other unselected variables by the stepwise Cox model included duration of diabetes, glycated haemoglobin, spot urine albumin:creatinine ratio, glomerular filtration rate, systolic/diastolic blood pressure, peripheral arterial disease, sensory neuropathy, retinopathy and baseline drug use (lipid-lowering drugs, ACE inhibitors or angiotensin II receptor blockers and insulin).

**Table 3 Parameter estimates and hazard ratios of baseline variables for cancer prediction in the training dataset**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Estimates of β values</th>
<th>Hazard ratio</th>
<th>Assumptions checking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter estimates</td>
<td>S.E.M.</td>
<td>Hazard ratio</td>
<td>95% CI</td>
</tr>
<tr>
<td>Age (years)</td>
<td>0.0488</td>
<td>0.0071</td>
<td>1.05</td>
</tr>
<tr>
<td>Total cholesterol (mmol/l, coded to 4.3 if &gt;4.3)</td>
<td>-0.5810</td>
<td>0.2608</td>
<td>0.56</td>
</tr>
<tr>
<td>White blood cell count (10⁹ count/l, recoded to 5.8 if &gt;5.8)</td>
<td>-0.3596</td>
<td>0.1634</td>
<td>0.70</td>
</tr>
<tr>
<td>Current smoking (1 if yes, 0 otherwise)</td>
<td>0.6390</td>
<td>0.2031</td>
<td>1.90</td>
</tr>
</tbody>
</table>

PH, proportional hazards; FF, functional forms.

*Chi² test for likelihood ratio test = 64.75, DF = 4, P<0.0001 and the shrinkage of the model is (64.75–4)/64.75 = 0.9382.

**Validation of the cancer risk score in the test data**

In the test database, the predicted cancer probability was not significantly different from the observed probability of cancer over 5 years of follow-up (Fig. 2). The overall C index was 0.695 (95% CI 0.667–0.724). With follow-up time and censoring status considered, the adjusted aROC was 0.712 for a 5-year observation period. Using a cut-off point of 1.2133, the sensitivity was 56.6% and the specificity was 71.4%. Sensitivity and specificity at other selected cut-off points are listed in Table 4.

Cumulative incidences of all-site cancer in patients with type 2 diabetes.
a risk score at or above and below the cut-off point are shown in Fig. 3.

**Predictive values for cancer subtypes**

In the combined database, the HR of patients with the risk scores at or above the suggested cut-off point of $-1.2133$ compared with those with risk score less than the cut-off point were as follows:

- 26.19 (95% CI 3.31–207.26, number of events = 12) for malignant neoplasm of lymphatic and hematopoietic tissue.
- 9.55 (95% CI 4.32–21.15, number of events = 44) for malignant neoplasm of respiratory and intrathoracic organs.
- 5.02 (95% CI 2.29–11.00, number of events = 29) for malignant neoplasm of upper digestive tract cancer.
- 4.09 (95% CI 2.37–7.04, number of events = 78) for malignant neoplasm of lower digestive tract cancer.
- 2.72 (95% CI 1.40–5.28, number of events = 47) for malignant neoplasm of other and unspecified sites.
- 2.40 (95% CI 1.29–4.47, number of events = 49) for malignant neoplasm of liver and intrahepatic bile ducts.
- 1.95 (95% CI 1.07–3.56, number of events = 53) for malignant neoplasm of bone, connective tissue, skin and breast.
- 1.91 (95% CI 1.03–3.55, number of events = 62) for malignant neoplasm of genitourinary organs.
- 1.56 (95% CI 0.46–5.32, number of events = 12) for malignant neoplasm of lip, oral cavity and pharynx.

**Discussion**

Our study demonstrates that low total cholesterol, low or high HDL-C and a triglyceride level around 1.17 mmol/l and low WBC count are associated with increased cancer risk in type 2 diabetic patients in addition to age and smoking status.

The nature of associations between cancer and lipids remains controversial. Large epidemiological studies of general population suggest an inverse total cholesterol–cancer relationship (Schatzkin et al. 1987, Schuit et al. 1993, Steenland et al. 1995). In support of these findings, we further identified the threshold value of total cholesterol $<4.30$ mmol/l for the inverse total cholesterol–cancer relationship in T2DM. Schuit et al. (1993) proposed several hypotheses for this inverse relationship:

Table 4 Sensitivity and specificity of selected cut-off points of the risk score and their corresponding 5-year probability of cancer in the test dataset

<table>
<thead>
<tr>
<th>Risk score</th>
<th>5-year cancer probability$^a$</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Positive predictive value</th>
<th>Negative predictive value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-1.8824$</td>
<td>0.0313</td>
<td>0.901</td>
<td>0.362</td>
<td>0.064</td>
<td>0.987</td>
</tr>
<tr>
<td>$-1.6131$</td>
<td>0.0402</td>
<td>0.820</td>
<td>0.484</td>
<td>0.071</td>
<td>0.982</td>
</tr>
<tr>
<td>$-1.6084$</td>
<td>0.0403</td>
<td>0.797</td>
<td>0.499</td>
<td>0.071</td>
<td>0.981</td>
</tr>
<tr>
<td>$-1.5503$</td>
<td>0.0425</td>
<td>0.766</td>
<td>0.523</td>
<td>0.072</td>
<td>0.979</td>
</tr>
<tr>
<td>$-1.5108$</td>
<td>0.0441</td>
<td>0.721</td>
<td>0.545</td>
<td>0.071</td>
<td>0.976</td>
</tr>
<tr>
<td>$-1.4620$</td>
<td>0.0461</td>
<td>0.702</td>
<td>0.574</td>
<td>0.074</td>
<td>0.976</td>
</tr>
<tr>
<td>$-1.4340$</td>
<td>0.0473</td>
<td>0.671</td>
<td>0.581</td>
<td>0.072</td>
<td>0.974</td>
</tr>
<tr>
<td>$-1.4084$</td>
<td>0.0484</td>
<td>0.659</td>
<td>0.604</td>
<td>0.074</td>
<td>0.973</td>
</tr>
<tr>
<td>$-1.3596$</td>
<td>0.0507</td>
<td>0.633</td>
<td>0.632</td>
<td>0.077</td>
<td>0.973</td>
</tr>
<tr>
<td>$-1.3104$</td>
<td>0.0530</td>
<td>0.581</td>
<td>0.661</td>
<td>0.076</td>
<td>0.970</td>
</tr>
<tr>
<td>$-1.2261$</td>
<td>0.0572</td>
<td>0.573</td>
<td>0.692</td>
<td>0.082</td>
<td>0.971</td>
</tr>
<tr>
<td>$-1.2133$</td>
<td>0.0579</td>
<td>0.566</td>
<td>0.714</td>
<td>0.087</td>
<td>0.972</td>
</tr>
<tr>
<td>$-1.1694$</td>
<td>0.0602</td>
<td>0.532</td>
<td>0.742</td>
<td>0.090</td>
<td>0.971</td>
</tr>
<tr>
<td>$-1.1666$</td>
<td>0.0604</td>
<td>0.524</td>
<td>0.742</td>
<td>0.089</td>
<td>0.970</td>
</tr>
<tr>
<td>$-1.1458$</td>
<td>0.0615</td>
<td>0.487</td>
<td>0.751</td>
<td>0.086</td>
<td>0.968</td>
</tr>
<tr>
<td>$-1.1064$</td>
<td>0.0638</td>
<td>0.461</td>
<td>0.770</td>
<td>0.088</td>
<td>0.967</td>
</tr>
</tbody>
</table>

$^a$Calculated using the equation: the 5-year probability of cancer $= 1 - 0.9590^{\exp(0.9382 \times (RISK \ SPACE + 1.5903))}$. 

$\chi^2 = 11.12, P > 0.10$
1) low total cholesterol may enhance carcinogenesis by affecting the integrity of cellular membranes; 2) both low total cholesterol and predisposition to cancers are expressions of an underlying, presumably genetic factor; 3) total cholesterol may exert a conditional influence (acting indirectly via hormonal or immunological mechanisms) or promote the action of an already initiated carcinogenic process. Other authors speculated that low total cholesterol might be attributable to cholesterol-lowering action of preclinical cancer (International Collaborative Group 1982, Sherwin et al. 1987). In our cohort, the significance of total cholesterol < 4.3 mmol/l for cancer persisted after excluding patients with a follow-up period of < 2.5 years. Large cohort studies with long follow-up (≥ 10 years) showed that the inverse total cholesterol–cancer relationship remained significant for measurements taken 5 or more years before diagnosis of cancer (Schatzkin et al. 1987, Schuit et al. 1993). Taken together, our findings suggest that the inverse total cholesterol–cancer relationship in general population is also applicable to T2DM with a threshold value around 4.3 mmol/l.

Findings relating to associations of HDL-C and triglyceride with cancer have been inconsistent. In a study of 38 823 Norwegian women aged 17–54 years with a median follow-up period of 17.2 years, women with HDL-C above 1.64 mmol/l were at decreased risk of breast cancer than their counterparts with a HDL level of 1.20 mmol/l and below (Furberg et al. 2004). In another study of 5207 Danish women followed for 26 years, inverse association between HDL-C and breast cancer had also been reported (Hoyer & Engholm 1992). Other investigators reported that breast cancer incidence covaried with triglyceride levels in ex-smokers (Manjer et al. 2001), although others failed to detect a significant overall negative relationship between serum triglycerides and breast cancer risk (Vatten & Foss 1990) or cancer mortality (Cowan et al. 1990, Bahl et al. 2005). Our study suggests that both high and low HDL-C were associated with cancer, especially high HDL-C for longer term cancer risk. In all the above studies, the HR curves of lipids for cancer were not examined. The non-linear nature of lipid association with cancer risk may explain previous negative reports where lipids were analysed either as continuous variables or categorical ones. Despite the significant associations on risk curve analysis, HDL-C and triglyceride were not selected into the cancer risk score possibly due to unsatisfactory data transformation, weak effects or small sample size.

In the Second National Health and Nutrition Examination Survey (Erlinger et al. 2004), increased WBC was significantly associated with high total cancer mortality. On the other hand, in a Chinese population, low lymphocyte count was associated with total cancer mortality (Huang et al. 2003). In this study, there was a tendency for WBC count to exhibit a U-shaped risk association with rapid increase in cancer risk as WBC count fell below < 5.8 × 10^9 counts per litre. There was a trend for high WBC above 10 × 10^9 counts per litre to be associated with increased risk of cancer, albeit short of significance.

Obesity was considered to be a risk factor for cancers including colon, liver, breast, pancreatic and kidney (Calle & Kaaks 2004, Fryzek et al. 2005, Frezza et al. 2006). Conversely, low BMI was associated with increased risk of esophageal cancer (Engeland et al. 2004), oral cavity and oropharyngeal cancer (Nieto et al. 2003). In this study, we found that BMI < 24 kg/m^2 and ≥ 27.6 kg/m^2 were both associated with increased risk of cancer compared with patients with a BMI level ≥ 24 to < 27.6 kg/m^2.

The nature of the close links between T2DM and increased cancer risk remains unclear, although it has long been recognized that lipoproteins exhibit regulatory effects on immune responses, immune mediation pathways, cellular metabolism of the immune system and antigen-nonspecific host resistance mediated by the lymphoid and reticuloendothelial systems. Other supporting evidence includes differences in lipoproteins in cancer patients (Edginton & Curtiss 1981). In the early 1980s, Vitale & Broitman (1981) hypothesized that, ‘If tumour cells or transformed cells (premalignant) are no longer regulated in terms of cholesterol, then changes in the concentration of lipids may enhance either the growth or transformation of aberrant cells to malignant cells by further deregulating cholesterol metabolism’. Against this background, the predictive values of low total cholesterol and WBC count for cancer in our analysis support the possible causal role of abnormal lipid metabolism and immune responses, frequently observed...
in T2DM, in the development of cancer. While experimental studies are needed to test these hypotheses, further investigations are also needed to examine whether these lipid–WBC–cancer relationships can also be generalized to non-diabetic populations.

In this analysis, we further confirmed the utility of these risk relationships by developing and validating an all-site cancer risk score with moderate discrimination and calibration. The predicted probability was similar to the actual probability in the test dataset, suggesting an adequate calibration of the risk score. The Framingham coronary heart disease risk score achieved a c statistics of 0.79 for men and 0.83 for women in their original population, 0.63–0.75 for men and 0.66–0.83 for women in separate USA populations (D'Agostino et al. 2001) and 0.71 for men and 0.74 for women in Chinese populations (Liu et al. 2004). Using the Hong Kong Diabetes Registry, we have developed various risk scores to predict stroke (aROC = 0.77) (Yang et al. 2007b), end-stage renal disease risk (aROC > 0.94) (Yang et al. 2006b, 2007c), coronary heart disease (aROC = 0.74) (Yang et al. 2008a), heart failure (aROC = 0.88) (Yang et al. 2006a) and all-cause death (aROC = 0.85) (Yang et al. 2008b). Our current all-cancer risk score has an aROC of 0.71, which is higher than that of 0.66 for a breast cancer risk scores using the Gail model (Barlow et al. 2006, Chen et al. 2006). However, compared with other risk scores, cancer remains the least predictable endpoint in T2DM.

Our subgroup analysis suggests that cancer risk score may have different predicting abilities for the subtypes of cancer. For some of these cancers, the risk scores may be used in combination with other validated screening tests to help clinicians make informed decisions. Besides, there were large differences in the HR of the risk score for cancer subtypes, suggesting that population differences in cancer patterns and aetiologies may further influence the performance of the risk score. Thus, the risk score may not be readily extrapolated from one population to another, and recalibration of the probability and/or adjustment of coefficients might be needed.

This study has several limitations. First, the current risk score only aims to predict all types of cancer. Site-specific cancer risk scores may have better performance, although due to sample size limitations, we did not develop such risk scores. Second, BMI, HDL-C and triglyceride which exhibited significant curved relationships were not selected into the predicting model. Their non-selection was likely due to our inability to devise adequate formulae to mimic their sharply curving shapes. Third, the overall C index or aROC is not optimal and that additional biomarkers need to be discovered for inclusion to improve its discrimination.

In conclusion, based on a relatively large prospective cohort of Hong Kong Chinese patients with T2DM, we detected that low total cholesterol (< 4.3 mmol/l), high and low HDL-C (< 0.9 or ≥ 1.8 mmol/l), triglyceride range of 0.81–1.41 mmol/l and low WBC (< 5.8 × 10^9 count per litre) were associated with increased risk of all-site cancer. Using total cholesterol and WBC, we were able to develop an all-site cancer risk score with moderate performance in the test dataset. In the light of the increased risk for multiple-site cancer in diabetic patients and continuing progress in the screening strategies and treatment of cancers, especially if detected early, the external validation of the risk score in other diabetic populations will be of public health importance.

**Acknowledgements**

We thank all medical and nursing staff of the Prince of Wales Hospital Diabetes Centre for recruiting and managing these patients. This study is supported by the Hong Kong Foundation for Research and Development in Diabetes, established under the auspices of the Chinese University of Hong Kong. The authors declare that there is no conflict of interest that would prejudice the impartiality of this scientific work.

**References**

Almdal T, Scharling H, Jensen JS & Vestergaard H 2004 The independent effect of type 2 diabetes mellitus on ischemic heart disease, stroke, and death: a population-based study of 13,000 men and women with 20 years of follow-up. Archives of Internal Medicine 164 1422–1426.


Yang XL, Kong PS, So WY, Ma RC, Cockram CS, Chow CC, Tong PC & Chan JC 2006a Risk scores for predicting hard coronary heart disease and nonfatal heart failure in Hong Kong Chinese type 2 diabetic patients – The Hong Kong Diabetes Registry The 19th World Diabetes Congress. South Africa: Cape Town.


Yang X, So WY, Tong PC, Ma RC, Kong AP, Lam CW, Ho CS, Cockram CS, Ko GT, Chow CC et al. 2008b Development and validation of an all-cause mortality risk score in type 2 diabetes – The Hong Kong Diabetes Registry. *Archives of Internal Medicine* **168**.