

# Time-varying incidence of cancer after the onset of type 2 diabetes: evidence of potential detection bias

J. A. Johnson · S. L. Bowker · K. Richardson ·  
C. A. Marra

Received: 17 May 2011 / Accepted: 13 June 2011 / Published online: 12 July 2011  
© Springer-Verlag 2011

## Abstract

**Aims/hypothesis** Despite the vast body of epidemiological literature on the risk of cancer in people with diabetes, few studies have examined the pattern of cancer risk during different time windows following diabetes onset. The objective of the study was to examine the risks of site-specific cancer in people with incident type 2 diabetes during different time windows following diabetes onset.

**Methods** This was a population-based retrospective cohort study. The study period was 1 April 1994 to 31 March 2006; censoring occurred at 31 March 2006, at death or on departure from British Columbia, Canada. Using linked health databases, we identified incident cohorts with and without diabetes, who were matched by age, sex and index year. Following a minimum 2-year cancer washout period, first site-specific cancers were identified prospectively in both cohorts.

**Results** Within 3 months following diabetes onset, participants with diabetes had significantly increased risks of colorectal, lung, liver, cervical, endometrial, ovarian, pancreatic and prostate cancers. After the initial 3-month period, the risks for colorectal (HR 1.15, 95% CI 1.05, 1.25), liver (HR 2.53, 95% CI 1.93, 3.31) and endometrial (HR 1.58, 95% CI 1.28, 1.94) cancers remained significantly elevated compared with those without diabetes. The diabetes cohort remained at increased risk of pancreatic

cancer in later years, but followed a different pattern: HR 3.71 at 3 months–1 year, 2.94 at 1–2 years, 1.78 at 2–3 years and 1.65 at 3–10 years ( $p$  value for all  $<0.01$ ). After an initial period of elevated risk, men with type 2 diabetes subsequently had a decreased risk of prostate cancer (HR 0.82, 95% CI 0.76, 0.88).

**Conclusions/interpretation** People with type 2 diabetes are at increased risk of select cancers; this risk is particularly elevated at the time of diabetes onset, which is likely to be due to increased ascertainment.

**Keywords** Cancer incidence · Cohort study · Detection bias · Type 2 diabetes

## Abbreviations

ACS	American Cancer Society
BCCA	British Columbia Cancer Agency
BCLHD	British Columbia Linked Health Databases
ICD-9	International Classification of Diseases, Ninth Revision
ICD-O-3	International Classification of Diseases for Oncology–Third Revision
NDSS	National Diabetes Surveillance System
PH	Proportional Hazards
SES	Socioeconomic Status

## Introduction

Numerous epidemiological studies have identified associations between diabetes and several types of cancer or cancer mortality in various populations [1, 2]. The relationship between diabetes and cancer was highlighted by the Joint Consensus Report from the ADA and American Cancer

J. A. Johnson (✉) · S. L. Bowker  
Department of Public Health Sciences, University of Alberta,  
2-040 Li Ka Shing Center,  
Edmonton, Canada, AB T6G 2E1  
e-mail: jeff.johnson@ualberta.ca

K. Richardson · C. A. Marra  
Faculty of Pharmaceutical Sciences,  
University of British Columbia,  
Vancouver, BC, Canada

Society (ACS) [3]. Although the literature indicates a strong and consistent increased risk of cancer in people with type 2 diabetes, the strength of the association depends on the specific cancer site. The strongest relationships have been demonstrated for liver and pancreatic cancers, although these may reflect some degree of ‘reverse causality’, with the cancer itself leading to the onset of diabetes [1, 2]. The risk of endometrial cancer appears to be doubled in women with diabetes [1, 2]. Risks of breast, colorectal and bladder are about 20–40% higher in those with type 2 diabetes [1, 2]. Interestingly, prostate cancer is about 10–20% less likely to occur in men with type 2 diabetes, which is thought to be due, in part, to the reduced levels of circulating testosterone levels in these individuals [3, 4].

Despite the vast body of epidemiological literature on the risk of cancer in people with diabetes, few studies have, to our knowledge, examined the pattern of cancer risk during different time windows following diabetes onset. In fact, the ADA/ACS Joint Consensus Report suggests that future epidemiological studies should address questions of the temporal relation between diabetes and cancer risk, and should consider the risk of various cancers separately [3]. Therefore, the objective of this study was to examine the risks of site-specific cancers in people with type 2 diabetes during different time windows following diabetes onset. We anticipated increased risks of site-specific cancers (and a decreased risk of prostate cancer), as suggested by previous epidemiology studies [1, 2]. Given the potential for reverse causality for liver and pancreatic cancers, and the possibility of detection bias playing a role in the increased risk of all types of cancer in people with type 2 diabetes, we examined the interaction between diabetes status and frequency of physician visits in the 2 years prior to the index date for the risk of site-specific cancers during different time windows.

## Methods

**Study population and data** We used the British Columbia Linked Health Databases (BCLHD) in this population-based retrospective cohort study [5]. Over 99% of the approximately 4.5 million people in British Columbia, Canada, are eligible for publicly funded health coverage, and their health utilisation data are captured in the BCLHD. The BCLHD includes information on physician services, hospital admissions, coverage period and vital statistics linked via a unique patient identifier. Physician visits and hospital admissions were classified according to the International Classification of Diseases, Ninth Revision (ICD-9; [www.icd9data.com/2007/V6lume1/240-279/250-259/250/default.htm](http://www.icd9data.com/2007/V6lume1/240-279/250-259/250/default.htm)). Family-level socioeconomic status (SES) was determined according to neighbourhood income

per person from the Canadian census and reported in quintiles.

In our data set, all traceable personal identifiers were removed to protect patient confidentiality. Ethics approval was obtained from the University of British Columbia Behavioral Research Ethics Board and the University of Alberta Health Research Ethics Board.

**Definition of cases of diabetes** Individuals likely to have a clinical diagnosis of diabetes were identified based on the established case definition for the Canadian National Diabetes Surveillance System (NDSS) [6]. The NDSS algorithm has been validated for use in Canadian administrative databases and has been found to have a sensitivity and specificity of 86% and 97%, respectively [7]. The index date for diabetes was defined as the first of: (1) a hospital admission for diabetes (ICD-9 code 250); or (2) the second of two medical fee-for-service claims coded ICD-9 250 within a 2-year period. Women with claims for gestational diabetes (ICD-9 648.8) were excluded. Individuals aged 30 years or older at the diabetes index date were considered to have type 2 diabetes.

**Cohort definitions** Our sample was drawn from two initial cohorts of subjects with and without diabetes identified from the BCLHD during the period 1 April 1994 to 31 March 2006. Subjects with diabetes were identified using the definition outlined in the previous section ( $n=328,994$ ); those who did not meet the criteria for diabetes during the study period, and were registered for health coverage in 1996 ( $n=327,572$ ), were eligible to be included in the non-diabetes cohort.

We identified individuals with incident diabetes as those meeting the definition of type 2 diabetes after two consecutive years of not meeting the criteria ( $n=185,100$ ). Non-diabetic participants registered for health coverage on the diabetes index date were then matched, with replacement, to each incident diabetic subject by sex and birth year. Both the incident diabetes and matched non-diabetes cohorts had a minimum 2-year cancer-free period prior to the index date. By having this minimum 2-year cancer-free period prior to the index date in both cohorts, we were more confident that we were capturing incident cancers subsequent to diabetes onset.

**Cancer outcomes** Cancer outcomes were provided via linkage to the British Columbia Cancer Agency (BCCA) database, which is estimated to cover at least 95% of all cancer cases in British Columbia [8]. The registry includes diagnosis date, cancer site (the International Classification of Diseases for Oncology–Third Revision [ICD-O-3]), and method of diagnosis of malignancy. Cancer sites (ICD-O-3 topography code) examined in this analysis include

colorectal (C18–C21), pancreatic (C25), lung (C34), liver (C22), breast (C50), cervical (C53), endometrial (C54), ovarian (C56), prostate (C61) and thyroid (C73).

The first cancer at each site was identified prospectively from the diabetes index date in both cohorts. Censoring occurred at 31 March 2006, at death or on departure from British Columbia, whichever was earliest. To avoid problems with competing risk, the development of a cancer in one site did not exclude subjects from analyses of a first cancer at another site.

**Statistical analysis** Incident rates and 95% CI were first determined for each cancer site using Poisson regression analysis. Cox regression was then used to estimate the HR and 95% CI for the association between incident diabetes and cancer. Analyses were adjusted for characteristics at the diabetes index date including age, sex, SES, frequency of physician visits in the 2 years prior to the index date (categorised into quintiles) and index year. The proportional hazards (PH) assumption was assessed by visual inspection of the log-minus-log plot. If the PH assumption was violated for diabetes, the follow-up time was split and the HR estimated in each interval using a multivariable piecewise PH model [9].

To account for time-varying HR, we split the follow-up time into early and later intervals based on observed patterns of risk. Splitting the follow-up time into two intervals at 3 months was deemed appropriate for all cancer sites, except for pancreatic and liver cancer. A further four intervals (3 months–1 year, 1–2 years, 2–3 years and >3 years) were deemed appropriate for pancreatic cancer because of a slower decline in risk over time, and a constant HR over time was valid for liver cancer. To further assess the potential for detection bias, we tested an interaction term for diabetes status and frequency of physician visits in the 2 years prior to the index date, within the early and later time windows. Cancer incidence HR and 95% CI were plotted against mean cancer/censoring time within the intervals for each cancer site. SAS version 9.2 (Cary, NC, USA) was used for analyses and R version 2.8 (Vienna, Austria) for plotting.

## Results

There were 185,100 participants in each of the incident type 2 diabetes and matched non-diabetes cohorts (Table 1). The mean±SD age of the two cohorts was 60.7±13.5 years, and 54% of the participants were men. Nearly 45% of the diabetes cohort fell into the lowest two SES quintiles, compared with 37% of the participants without diabetes. The diabetes cohort also had more physician visits than the non-diabetes cohort in the 2 years prior to the index date

(Table 1). The mean cancer-free period before the index date was 6.9 years and 7.1 years for cohorts with and without diabetes, respectively. The mean±SD follow-up was 4.3±2.8 and 4.4±2.8 years for the diabetes and non-diabetes cohorts, respectively.

Throughout the entire follow-up, the overall incidence rate of cancer (per 1,000 person-years) was 16.04 (95% CI 15.76, 16.32) in the diabetes cohort, compared with 14.25 (95% CI 13.95, 14.56) in the control cohort (Table 2). The unadjusted incidence rates for colorectal, pancreatic, lung, liver and endometrial cancers were all higher in the diabetes cohort, compared to those without diabetes. The incidence rates for breast, cervical, ovarian and thyroid cancer were similar in the two cohorts, and, as expected, the incidence rate of prostate cancer was lower in the diabetes cohort.

In the first 3 months following the diabetes index date, the risks of colorectal, lung, cervical, endometrial, ovarian and prostate cancer in participants with type 2 diabetes were significantly elevated compared with those without diabetes (Table 3). There was a trend towards an increased risk of thyroid and breast cancer in the diabetes cohort during the first 3 months after the diabetes index date, but these findings were not statistically significant. Figure 1 illustrates the clear peaks in risk for each cancer subtype in the early time period subsequent to the diabetes index date.

In the later time period of 3 months to 10 years after the diabetes index date, the diabetes cohort remained at a significantly elevated risk of colorectal and endometrial cancer compared with those without diabetes (Table 3). A non-significant increased risk of thyroid cancer remained in those with diabetes. The risk of prostate cancer during this later time period was significantly lower in the diabetes cohort compared with those without diabetes. The risks of lung, cervical, ovarian and breast cancers in the diabetes cohort compared with those without diabetes were close to the null in this later time period. Importantly, risk estimates for several cancers (colorectal, lung, cervical, endometrial, ovarian and prostate) for the full follow-up period appeared to be overestimated compared with those risks excluding the first 3 months after diabetes onset (Table 3).

The risk of pancreatic cancer in the first 3 months after diabetes onset was significantly higher in the diabetes cohort (adjusted HR 13.84, 95% CI 7.49, 25.58;  $p<0.0001$ ). The adjusted HRs for incident pancreatic cancer in the diabetes cohort, for the time periods 3 months–1 year, 1–2 years, 2–3 years and 3–10 years after the diabetes index date were 3.71 (95% CI 2.55, 5.39;  $p<0.0001$ ), 2.94 (95% CI 2.00, 4.33;  $p<0.0001$ ), 1.78 (95% CI 1.14, 2.77;  $p=0.01$ ) and 1.65 (95% CI 1.28, 2.13;  $p=0.0001$ ), respectively. The risk of liver cancer for those with diabetes remained constantly elevated throughout the entire follow-up period (HR 2.53; 95% CI 1.93–3.31;  $p<0.0001$ ).

**Table 1** Baseline demographics for incident type 2 diabetes and matched non-diabetes control cohorts

Characteristics at index date	Incident diabetes ( <i>N</i> =185,100)		No diabetes ( <i>N</i> =185,100) <sup>a</sup>	
	<i>n</i>	%	<i>n</i>	%
Sex				
Women	84,506	45.7	84,506	45.7
Men	100,594	54.3	100,594	54.3
Index date				
Apr 96–Mar 97	17,500	9.5	17,500	9.5
Apr 97–Mar 98	17,399	9.4	17,399	9.4
Apr 98–Mar 99	16,622	9.0	16,622	9.0
Apr 99–Mar 00	17,567	9.5	17,567	9.5
Apr 00–Mar 01	17,787	9.6	17,787	9.6
Apr 01–Mar 02	18,435	10.0	18,435	10.0
Apr 02–Mar 03	19,317	10.4	19,317	10.4
Apr 03–Mar 04	19,544	10.6	19,544	10.6
Apr 04–Mar 05	20,866	11.3	20,866	11.3
Apr 05–Mar 06	20,063	10.8	20,063	10.8
Age (years)				
30–40	12,262	6.6	12,311	6.7
40–50	30,410	16.4	30,357	16.4
50–60	47,539	25.7	47,481	25.7
60–70	45,325	24.5	45,444	24.6
70–80	34,372	18.6	34,259	18.5
>80	15,192	8.2	15,248	8.2
Mean±SD	60.7±13.5		60.7±13.5	
SES				
First quintile	43,131	23.3	34,964	18.9
Second quintile	39,154	21.2	34,114	18.4
Third quintile	35,172	19.0	34,796	18.8
Fourth quintile	32,569	17.6	35,901	19.4
Fifth quintile	29,149	15.7	38,052	20.6
Missing	5,925	3.2	7,273	3.9
Frequency of physician visits (quintiles) <sup>b</sup>				
0–5 (reference)	13,465	7.3	39,587	21.4
6–12	32,466	17.5	38,318	20.7
13–20	37,621	20.3	35,020	18.9
21–34	45,881	24.8	35,945	19.4
>34	55,667	30.1	36,230	19.6
Minimum cancer-free period before index date (years)				
Mean±SD	6.9±2.9		7.1±2.9	
Median (IQR)	6.9 (4.3–9.4)		7.2 (4.6–9.6)	
Follow-up time at risk of cancer (years)				
Mean±SD	4.2±2.8		4.3±2.8	
Median (IQR)	3.8 (1.8–6.5)		3.9 (1.8–6.6)	
Follow-up time, years				
Mean±SD	4.3±2.8		4.4±2.8	
Median (IQR)	4.0 (1.8–6.7)		4.1 (1.9–6.8)	

<sup>a</sup> Matched by sex and birth year (with replacement)<sup>b</sup> In the 2 years prior to the index date

IQR interquartile range

Interactions between diabetes status and frequency of physician visits in the 2 years prior to the index date were statistically significant for lung, breast, cervical and prostate

cancer in the first 3 months (Table 4). All cancer sites, except liver and ovarian, followed a similar pattern of substantially elevated risk for those subjects with diabetes

**Table 2** Incident rates for cancers in the type 2 diabetes and control cohorts, in British Columbia, for the whole follow-up period 1996–2006

Cancer site	Incident diabetes (N=185,100)						No diabetes (N=185,100) <sup>a</sup>					
	0–3 months since index date			>3 months since index date			0–3 months since index date			>3 months since index date		
	Incident cancer (n)	Incidence rate (/1,000 person-years)	95% CI	Incident cancer (n)	Incidence rate (/1,000 person-years)	95% CI	Incident cancer (n)	Incidence rate (/1,000 person-years)	95% CI	Incident cancer (n)	Incidence rate (/1,000 person-years)	95% CI
Colorectal	219	4.90	4.29, 5.60	1,452	1.93	1.83, 2.03	78	1.73	1.38, 2.17	1,368	1.75	1.64, 1.86
Pancreas	150	3.36	2.86, 3.94	488	0.65	0.59, 0.71	11	0.24	0.14, 0.44	234	0.29	0.25, 0.34
Lung	227	5.08	4.46, 5.79	1,463	1.94	1.84, 2.04	83	1.84	1.49, 2.29	1,438	1.80	1.70, 1.92
Liver	25	0.56	0.38, 0.83	246	0.33	0.29, 0.37	8	0.18	0.09, 0.36	100	0.13	0.10, 0.16
Breast (female)	89	4.36	3.54, 5.37	1,089	3.17	2.98, 3.36	72	3.50	2.78, 4.41	1,131	3.18	2.97, 3.40
Cervical	13	0.64	0.37, 1.10	95	0.27	0.22, 0.33	4	0.19	0.07, 0.52	92	0.26	0.21, 0.33
Endometrial	29	1.42	0.99, 2.04	297	0.86	0.76, 0.96	13	0.63	0.37, 1.09	212	0.59	0.51, 0.69
Ovarian	25	1.22	0.83, 1.81	128	0.37	0.31, 0.44	13	0.63	0.37, 1.09	129	0.37	0.30, 0.44
Prostate	220	9.07	7.94, 10.35	1,557	3.86	3.67, 4.06	111	4.54	3.77, 5.47	1,955	4.68	4.45, 4.93
Thyroid	8	0.18	0.09, 0.36	66	0.09	0.07, 0.11	4	0.09	0.03, 0.24	48	0.06	0.05, 0.08
Any neoplasm	1,573	35.35	33.64, 37.14	10,865	14.86	14.59, 15.15	635	14.13	13.07, 15.28	10,838	14.20	13.89, 14.52

<sup>a</sup> Matched on sex, birth year and index year (with replacement)

**Table 3** HRs for first cancer at each site in the incidence diabetes cohort compared to the matched cohort without diabetes for different time periods following the diabetes index date

Cancer site	Analysis	Diabetes HR by time since index date								
		0–3 months			3 months–10 years			0 months–10 years		
		HR	95% CI	<i>p</i> value	HR	95% CI	<i>p</i> value	HR	95% CI	<i>p</i> value
Colorectal <i>n</i> =3,117	Unadjusted	2.82	2.17, 3.65	<0.0001	1.09	1.01, 1.18	0.04	1.19	1.10, 1.27	<0.0001
	Adjusted <sup>a</sup>	2.89	2.22, 3.75	<0.0001	1.15	1.05, 1.25	0.002	1.24	1.14, 1.35	<0.0001
Lung <i>n</i> =3,211	Unadjusted	2.74	2.13, 3.53	<0.0001	1.05	0.96, 1.13	0.28	1.14	1.05, 1.23	0.001
	Adjusted <sup>a</sup>	2.70	2.09, 3.47	<0.0001	1.05	0.97, 1.14	0.23	1.14	1.06, 1.24	0.001
Breast (female) <i>n</i> =2,381	Unadjusted	1.24	0.91, 1.70	0.17	0.99	0.91, 1.09	0.86	1.01	0.92, 1.10	0.88
	Adjusted <sup>a</sup>	1.23	0.90, 1.68	0.19	0.99	0.90, 1.09	0.82	1.00	0.92, 1.10	0.93
Cervical <i>n</i> =204	Unadjusted	3.27	1.07, 10.02	0.04	1.06	0.78, 1.44	0.72	1.39	1.18, 1.63	<0.0001
	Adjusted <sup>a</sup>	3.46	1.10, 10.86	0.03	1.10	0.79, 1.53	0.57	1.50	1.26, 1.77	<0.0001
Endometrial <i>n</i> =551	Unadjusted	2.24	1.17, 4.31	0.02	1.44	1.19, 1.76	0.0003	1.49	1.23, 1.80	<0.0001
	Adjusted <sup>a</sup>	2.43	1.26, 4.68	0.008	1.58	1.28, 1.94	<0.0001	1.63	1.33, 1.99	<0.0001
Ovarian <i>n</i> =295	Unadjusted	1.93	0.99, 3.78	0.05	1.02	0.78, 1.33	0.87	1.11	0.86, 1.42	0.42
	Adjusted <sup>a</sup>	2.06	1.05, 4.05	0.04	1.10	0.84, 1.44	0.48	1.19	0.93, 1.53	0.17
Prostate <i>n</i> =3843	Unadjusted	1.99	1.59, 2.51	<0.0001	0.81	0.76, 0.87	<0.0001	0.88	0.82, 0.94	0.0002
	Adjusted <sup>a</sup>	1.98	1.57, 2.48	<0.0001	0.82	0.76, 0.88	<0.0001	0.88	0.82, 0.95	0.0007
Thyroid <i>n</i> =126	Unadjusted	2.01	0.61, 6.68	0.25	1.41	0.94, 2.11	0.1	1.46	0.99, 2.14	0.06
	Adjusted <sup>a</sup>	1.83	0.55, 6.12	0.32	1.29	0.85, 1.95	0.23	1.29	0.87, 1.91	0.21

<sup>a</sup> Adjusted for age, sex, SES, number of physician visits and year of diagnosis

who had fewer physician visits compared with their counterparts without diabetes, although this did not reach statistical significance (data not shown).

## Discussion

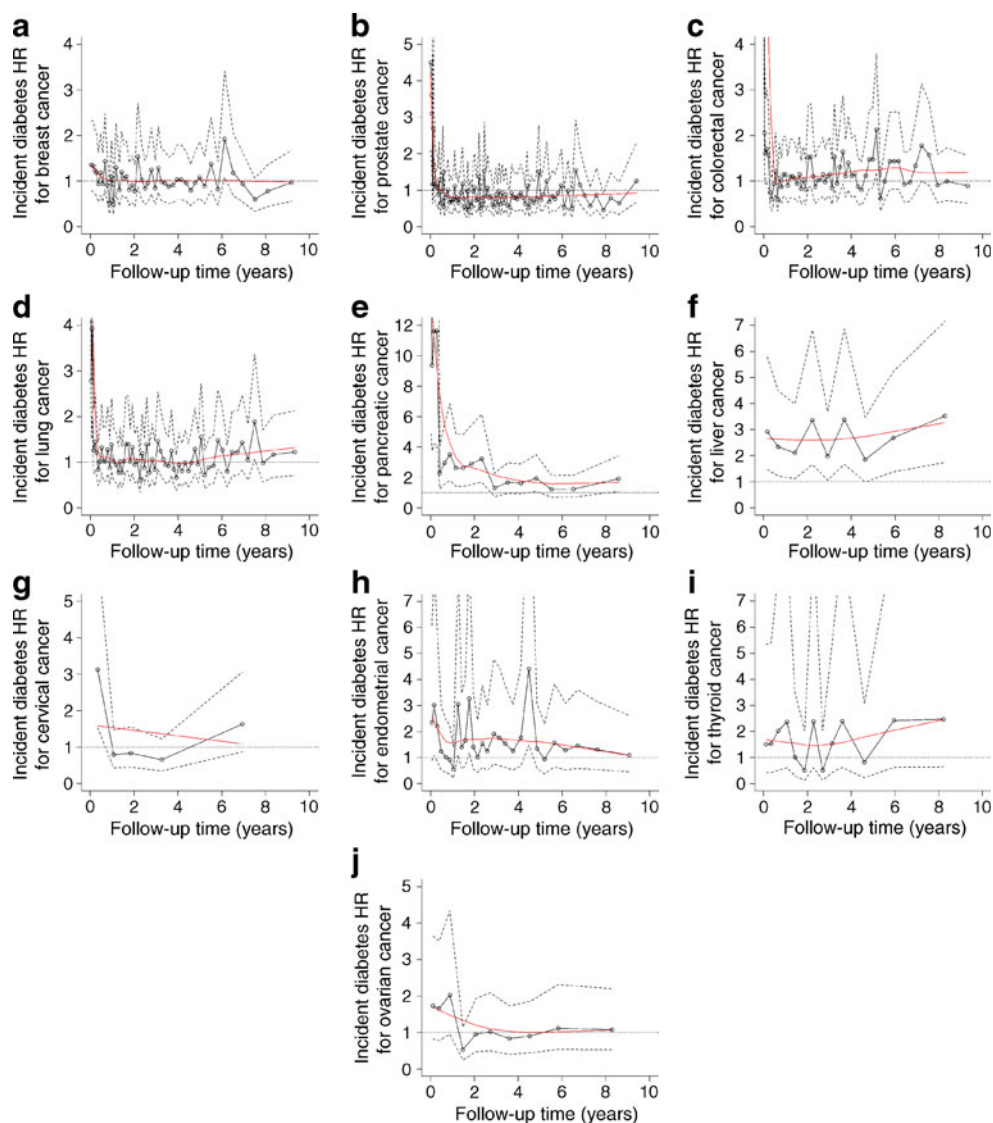
In a large population-based cohort of people from British Columbia, Canada, we examined the pattern of cancer incidence during different time windows following diabetes onset. The most striking observation was an initial spike in risk for most cancers immediately following the onset of diabetes, and a subsequent levelling off of risk throughout the remainder of follow-up. Interestingly, the peak in cancer risk within 0–3 months of diabetes onset was evident for all cancers, with the exception of liver cancer. For some cancers (colorectal, liver and endometrial cancer), the risk levelled off but remained elevated in those with diabetes, while for other cancers (lung, breast, cervical and ovarian cancer), the subsequent risk was the same as that observed in participants without diabetes, and for prostate cancer, the risk was subsequently lower in men with diabetes.

While the mechanisms underlying the epidemiological associations between diabetes and cancer are not entirely

clear, common risk behaviours, such as unhealthy diets, physical inactivity, inflammatory markers and hyperinsulinaemia, are important considerations. Of particular concern, however, are methodological issues related to reverse causality and detection bias [3]. Our data support the concern that the elevated risk of some cancers in people with diabetes may be partly due to increased detection around the time of diabetes diagnosis. However, reverse causality does not appear to account for the elevated risk of cancer in people with type 2 diabetes.

The increased risks of pancreatic and liver cancers have been of particular concern with regards to reverse causality, as dysfunction of these metabolically active organs would result in impairments of glycaemic control. We were able to address this concern by having, on average, a 7-year cancer washout period (with a minimum of 2 years), for both the diabetes and control cohorts prior to the diabetes index date. In addition, regardless of any reverse causality that might have been present, if at all, we continued to observe a strong and statistically significant increased risk for both liver (HR 2.53) and pancreatic cancers (HR 1.65) in the latter time windows. As pancreatic cancer is rapidly progressive and generally fatal [10], it is highly unlikely that such a diagnosis preceded the diabetes index date in later time periods.

**Fig. 1** Adjusted HRs for site-specific cancers in incident diabetes cohorts over time since onset of diabetes: (a) breast cancer; (b) prostate cancer; (c) colorectal cancer; (d) lung cancer; (e) pancreatic cancer; (f) liver cancer; (g) cervical cancer; (h) endometrial cancer; (i) thyroid cancer; (j) ovarian cancer



On the other hand, there does appear to be evidence of a substantial detection bias in the assessment of cancer risk in the diabetes population. Detection, or ascertainment, bias arises due to the increase in the number of clinical investigations associated with a new diagnosis of diabetes, which may then lead to the discovery of a cancer [3]. The general pattern of an initial elevated peak in cancer risk at the time of diabetes onset, which is substantially higher than subsequent risk, suggests that many cancers are being diagnosed at or around the time of recognition of diabetes, or vice versa. Moreover, the pattern of risks associated with the interaction between diabetes status and frequency of physician visits in the 2 years prior to diabetes onset further suggests a substantial degree of detection bias. The likelihood of having a cancer diagnosis within 0–3 months of diabetes onset was much greater in those who had fewer physician visits in the previous years. Given that we excluded individuals with cancer diagnoses prior to their

index date, this suggests that the likelihood of diagnosing cancer was increased by the fact that diabetes had recently been identified.

This raises one of the potential limitations of our methods, in particular the index date assigned to diabetes onset. The case definition for diabetes is based on one hospitalisation or two physician visits over a 2-year period, with the index date being defined as the later of those dates. Thus, is it possible for the clinical diagnosis of diabetes to have preceded the case date in the BCLHD; however, previous validation of this diabetes case definition suggests that the median duration between qualifying physician visits is 39 days (interquartile range 12–150 days) (J. Johnson, unpublished data). We also ensured that participants were free of a cancer diagnosis in the 2 years prior to the index date, coinciding with the same window as the case definition for diabetes. We recognise, however, when diabetes and cancer are diagnosed within

**Table 4** HRs and 95% CI for risk of cancer by site among the incident diabetes cohort compared with the matched cohort without diabetes (the reference group) during earlier and later time windows

Cancer site	Frequency of physician visits in previous 2 years ( <i>n</i> )	Time window following index date							
		0–3 months				3 months–10 years			
		HR	95% CI	<i>p</i> value	<i>p</i> value for interaction	HR	95% CI	<i>p</i> value	<i>p</i> value for interaction
Lung	0–5 ( <i>n</i> =353)	7.11	4.63, 10.92	<0.0001	<0.0001	1.02	0.78, 1.34	0.88	0.90
	6–12 ( <i>n</i> =466)	2.98	1.92, 4.63	<0.0001		1.15	0.93, 1.42	0.19	
	13–20 ( <i>n</i> =561)	2.10	1.36, 3.24	0.001		1.02	0.84, 1.24	0.84	
	21–34 ( <i>n</i> =784)	2.53	1.77, 3.62	<0.0001		1.01	0.86, 1.19	0.89	
	>34 ( <i>n</i> =1,047)	2.33	1.68, 3.23	<0.0001		1.04	0.90, 1.21	0.56	
Breast	0–5 ( <i>n</i> =168)	1.97	0.71, 5.50	0.19	0.02	1.46	1.02, 2.10	0.04	0.27
	6–12 ( <i>n</i> =360)	2.35	1.41, 3.90	0.001		0.90	0.71, 1.14	0.38	
	13–20 ( <i>n</i> =469)	0.71	0.36, 1.40	0.33		0.97	0.79, 1.19	0.78	
	21–34 ( <i>n</i> =630)	1.12	0.69, 1.80	0.66		0.98	0.82, 1.16	0.78	
	>34 ( <i>n</i> =754)	1.13	0.73, 1.75	0.58		0.98	0.83, 1.16	0.84	
Cervical	0–5 ( <i>n</i> =33)	33.08	9.57, 114.39	<0.0001	<0.0001	0.92	0.34, 2.51	0.88	0.10
	6–12 ( <i>n</i> =33)	2.89	0.29, 28.98	0.37		2.65	1.17, 5.96	0.02	
	13–20 ( <i>n</i> =44)	2.05	0.36, 11.67	0.42		0.85	0.45, 1.62	0.63	
	>20 ( <i>n</i> =94)	0.78	0.14, 4.36	0.78		0.85	0.54, 1.34	0.48	
Prostate	0–5 ( <i>n</i> =420)	4.05	2.57, 6.38	<0.0001	<0.0001	0.81	0.63, 1.04	0.10	0.63
	6–12 ( <i>n</i> =690)	2.99	2.11, 4.24	<0.0001		0.81	0.69, 0.96	0.02	
	13–20 ( <i>n</i> =819)	2.05	1.44, 2.92	<0.0001		0.83	0.71, 0.97	0.02	
	21–34 ( <i>n</i> =935)	1.63	1.15, 2.29	0.006		0.75	0.65, 0.86	<0.0001	
	>34 ( <i>n</i> =979)	1.35	0.94, 1.93	0.10		0.88	0.76, 1.02	0.09	

All analyses adjusted for age, sex, SES, frequency of physician visits, year of diagnosis

The reference group is participants without diabetic people in the subgroup

3 months of each other, it is difficult to accurately determine which disease appeared first. In other words, it is possible to have a detection bias in both directions, which may lead to ambiguity with respect to the association between the two conditions, especially in the situation where there is only a short follow-up period after the first diagnosis. This would warrant further investigation to assess the risk of cancer diagnosis prior to the diabetes index date.

With this in mind, there nonetheless appears to be further evidence for detection bias with two cancers in particular: lung and prostate cancer. Lung cancer has generally not been associated with an increased risk in people with diabetes [3], yet in the early time window of our analysis, we observed a 170% increased risk of lung cancer, which subsequently disappeared after that initial 3-month period. Even more suggestive is the 98% increased risk of prostate cancer in the early time window but a subsequently decreased risk among men with diabetes in the longer term, which is what would have been expected based on physiological plausibility and previous epidemiological evidence [3].

Of particular interest are the observed risk estimates of breast cancer in our cohorts. A previous meta-analysis of cohort and case-control studies found that women with diabetes had a moderate, but significant, 20% increased risk of breast cancer [11]. This risk is similar to that observed during the early period in our analyses, but this was followed by a lack of association in the subsequent time period. This suggests that previous epidemiology studies of breast cancer risk in women with diabetes may not have fully accounted for an initial detection bias.

To our knowledge, none of the previous epidemiology studies of diabetes and cancer risk has considered cancer incidence during different time windows following the diabetes index date. Therefore, while we continued to see significant associations during the latter time period following the diabetes index date for certain subtypes of cancer (colorectal, liver, endometrial, pancreatic and prostate), our risk estimates are likely to be more reliable and accurate, given that we were able to take into account the initial time window of heightened screening. It is possible that previous studies may have overestimated the risk of various types of cancer in patients



with diabetes if they did not properly account for or take into consideration this initial detection bias following a diagnosis of diabetes.

A key strength of this study relates to our use of population-based administrative health data. We were able to assemble very large cohorts, and our diabetes cohort captured the entire population of incident diabetes in the province of British Columbia. As such, this study does not suffer from the potential biases that exist in selective, non-population-based studies. Furthermore, the British Columbia Linked Health databases and British Columbia Cancer Registry have been used in numerous epidemiology studies [12–15]. In our analyses, we matched our cohorts by age, sex and index year, thus balancing the cohorts equally on these potential confounders. Finally, we were able to look at the incidence of several different types of cancer, although the numbers were small for some subtypes, such as thyroid and cervical cancer.

On the other hand, there are some limitations to our study. As we used administrative data, we lacked information on potentially important clinical covariates, such as treatment modality (diet, oral agents and insulin), smoking status, weight or BMI, glycaemic control and physical activity. These are all potential confounders in the relationship between diabetes and the incidence of cancer. However, other studies that have included information on these clinical covariates when looking at the risk of cancer in people with diabetes have observed results similar to ours [1, 2].

In summary, this study provides new evidence on the relationship between type 2 diabetes and cancer incidence by examining the risk of cancer during different time windows following diabetes index date. It appears, from the patterns of cancer incidence, that reverse causality may play a role, but it does not account for all of the elevated risk of cancer in type 2 diabetes. On the other hand, there does appear to be a substantial detection bias for most cancers in people with new-onset type 2 diabetes. Although an elevated risk of some cancers is sustained over time, detection bias following the onset of type 2 diabetes probably leads to an overestimation of cancer risk.

**Acknowledgements** J. A. Johnson is a Senior Scholar with Alberta Innovates–Health Solutions and holds a Canada Research Chair in Diabetes Health Outcomes. C. A. Marra holds a Canada Research Chair in Pharmaceutical Outcomes. This work was supported in part by an Operating Grant from the Canadian Institute for Health Research (CIHR) (MOP-82737) and a CIHR Team Grant to the Alliance for Canadian Health Outcomes Research in Diabetes (ACHORD) (OTG-88588), sponsored by the CIHR Institute of Nutrition, Metabolism and Diabetes. The funding source played no role in the study design, the collection, the analysis or the interpretation of the data.

**Contribution statement** All authors contributed to: (1) conception and design of the study; (2) analysis and interpretation of data; (3) drafting of the manuscript or revising it critically for important intellectual content; and (4) final approval of the version to be published.

**Duality of interest** The authors declare that there is no duality of interest associated with this manuscript.

## References

- Nicolucci A (2010) Epidemiological aspects of neoplasms in diabetes. *Acta Diabetol* 47:87–95
- Vigneri P, Frasca F, Sciacca L, Pandini G, Vigneri R (2009) Diabetes and cancer. *Endocr Relat Cancer* 16:1103–1123
- Giovannucci E, Harlan DM, Archer MC et al (2010) Diabetes and cancer: a consensus report. *Diabetes Care* 33:1674–1685
- Kasper JS, Giovannucci E (2006) A meta-analysis of diabetes mellitus and the risk of prostate cancer. *Cancer Epidemiol Biomarkers Prev* 15:2056–2062
- Chamberlayne R, Green B, Barer ML, Hertzman C, Lawrence WJ, Sheps SB (1998) Creating a population-based linked health database: a new resource for health services research. *Can J Public Health* 89:270–273
- Clotey C, Mo F, LeBrun B, Mickelson P, Niles J, Robbins G (2001) The development of the National Diabetes Surveillance System (NDSS) in Canada. *Chronic Dis Can* 22:67–69
- Hux JE, Ivis F, Flintoft V, Bica A (2002) Diabetes in Ontario: determination of prevalence and incidence using a validated administrative data algorithm. *Diabetes Care* 25:512–516
- Parkin DM, Whelan SL, Ferlay J, Raymond L, Young J (eds). *Cancer incidence in five continents*. Vol. VII. Lyon, France: IARC, 1997
- Moreau T, Oquigley J, Mesbah M (1985) A global goodness-of-fit statistic for the proportional hazards model. *Applied Statistics—Journal of the Royal Statistical Society Series C* 34:212–218
- Huxley R, Ansary-Moghaddam A, Berrington de Gonzalez A, Barz F, Woodward M (2005) Type 2 diabetes and pancreatic cancer: a meta-analysis of 36 studies. *Br J Cancer* 92:2076–2083
- Larsson SC, Mantzoros CS, Wolk A (2007) Diabetes mellitus and risk of breast cancer: a meta-analysis. *Int J Cancer* 121:856–862
- Pilote L, Beck CA, Karp I et al (2004) Secondary prevention after acute myocardial infarction in four Canadian provinces, 1997–2000. *Can J Cardiol* 20:61–67
- Schneeweiss S, Dormuth C, Grootendorst P, Soumerai SB, Maclure M (2004) Net health plan savings from reference pricing for angiotensin-converting enzyme inhibitors in elderly British Columbia residents. *Med Care* 42:653–660
- Schneeweiss S, Maclure M, Carleton B, Glynn RJ, Avorn J (2004) Clinical and economic consequences of a reimbursement restriction of nebulised respiratory therapy in adults: direct comparison of randomised and observational evaluations. *BMJ* 328:560
- Schneeweiss S, Wang PS, Avorn J, Maclure M, Levin R, Glynn RJ (2004) Consistency of performance ranking of comorbidity adjustment scores in Canadian and US utilization data. *J Gen Intern Med* 19:444–450